

АНАЛИЗ И ОЦЕНКА ВАЛИДНОСТИ ТЕСТА «ТЕОРИЯ ФУНКЦИЙ КОМПЛЕКСНОЙ ПЕРЕМЕННОЙ»

*Л.А. Муратова*¹

Самарский государственный технический университет
443100, г. Самара, ул. Молодогвардейская, 244
E-mail: muratova-la@mail.ru

Проводится анализ качества теста «Теория функций комплексной переменной». Оценивается валидность теста по нормальному распределению, по содержанию, по различающей способности. Вид и пропорции кривой распределения набранных баллов свидетельствуют в пользу нормального распределения. Тест достаточно полно и глубоко отражает изучаемый материал, что позволяет сделать вывод о его содержательной валидности. Для проверки качества теста по различающей способности найдены две числовые характеристики для каждого задания: точечный би-серийный коэффициент и коэффициент дискриминативности. Последующий анализ показал, что только два задания не удовлетворяют одному из условий качества. При этом критической точкой считалось среднее значение среди тех, что предлагают разные авторы. После анализа содержания заданий предлагается одно из них переработать, а другое сохранить.

Ключевые слова: валидность, дискриминативность, нормальное распределение.

Тестирование является основным средством контроля качества обучения. Его преимущества: научная обоснованность теста по количеству и качеству предлагаемых заданий, технологичность, точность измерений, объективность при проведении и оценивании результатов [1].

Как бы хорошо ни был составлен тест, как правило, есть возможность его улучшить. С этой целью материалы, полученные по результатам тестирования, подвергаются математико-статистической обработке. Найденные характеристики анализируются на предмет качества самого теста с помощью существующих теорий, после чего тест подвергается переработке с учетом выявленных недостатков. При этом рекомендуется использовать несколько методик подбора и анализа критериев качества теста и относиться к полученным оценкам как к правдоподобным утверждениям, имеющим ту или иную степень достоверности [2].

¹ *Лидия Александровна Муратова*, кандидат технических наук, доцент кафедры высшей математики и прикладной информатики.

Научное обоснование качества теста предполагает оценку таких показателей, как надежность, трудность, валидность, дискриминативность.

Валидность относится к важнейшим критериям оценки качества теста, занимая второе место после надежности. Рассмотрим этот критерий – валидность – более подробно. Сам по себе вопрос о валидности теста считается одним из самых сложных. Валидность определяется как понятие, указывающее, что тест измеряет и насколько хорошо он это делает [3]. Иначе, валидность должна показывать, насколько хорошо тест делает то, для чего он был создан [4]. Некоторая неопределенность формулировки этого понятия приводит к разнообразным способам проверки валидности. Понятие валидности включает в себя разные ее виды, со своим особым смыслом. Так, рассматривают валидность по содержанию, конструктивную валидность, критериальную (текущую и прогностическую), валидность по распределению, по различающей способности (дискриминативности) [5, 6].

В качестве предмета рассмотрения возьмем тест «Теория функций комплексной переменной» (табл. 1) курса высшей математики СамГТУ. Тест используется при текущем контроле знаний у студентов второго года обучения/ Он включает в себя все темы раздела в соответствии с программой курса, состоит из различных по сложности заданий. Это однородный тест закрытого типа, к каждому из двенадцати заданий предлагается пять ответов, среди которых лишь один правильный. В данном случае отобраны работы студентов нефтетехнологического факультета. После исключения работ студентов, ответивших на все вопросы правильно либо ответивших на все вопросы неправильно, в выборке осталось 158 работ.

Оценим валидность теста по трем критериям: по нормальному распределению, по содержанию и по различающей способности [1, 7].

Различают тесты критериально-ориентированные и нормативно-ориентированные. Первые дают возможность определить, в какой степени испытуемые владеют учебным материалом. Вторые позволяют сравнить уровни знаний тестируемых. Существует мнение, что можно получить качественный тест, дающий целостную картину полученных знаний, совмещая оба подхода. Опираясь на эту точку зрения, проанализируем полученные в результате обработки теста результаты.

С точки зрения правильно сконструированного нормативно-ориентированного теста кривая распределения индивидуальных баллов должна соответствовать нормальному распределению, то есть быть симметричной и унимодальной, при этом типичным считается результат, когда примерно 70 % тестируемых правильно выполняют от 30 до 70 % заданий, а наи-

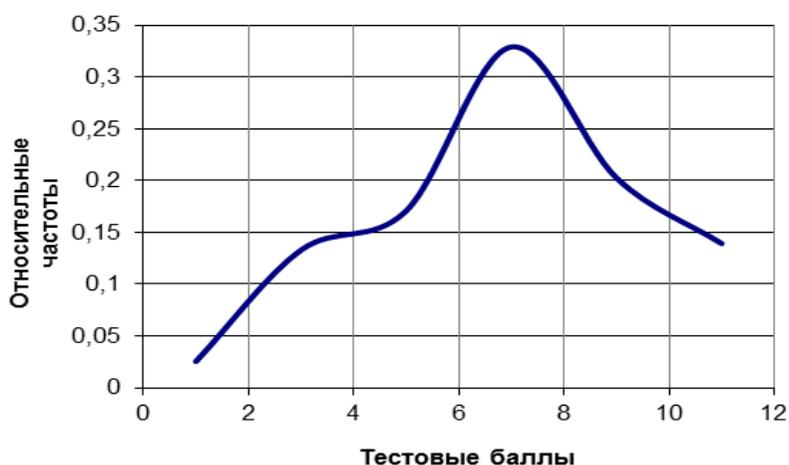
более часто встречается результат 50 % [8]. Для критериально-ориентированного теста нормальное распределение не характерно.

Таблица 1

Теория функций комплексной переменной

1. Представить в показательной форме число $z = -1 + \sqrt{3}i$.
2. Вычислить: $\frac{1-i}{1+i}$.
3. Указать множество точек на комплексной плоскости, удовлетворяющее условию $|z - 2i| \leq 3$.
4. Вычислить: $\cos 2i$.
5. Решить уравнение $z^4 + 1 = 0$ и найти возможные значения $\arg z$.
6. Для аналитической функции $f(z) = u(x, y) + i v(x, y)$ найти $v(x, y)$, если известно, что $u(x, y) = 2x + y$, $f(0) = 0$.
7. Вычислить интеграл $\int_{AB} (1 + 2\bar{z}) dz$, где AB – отрезок прямой между точками $z_1 = -1$, $z_2 = -i$.
8. Вычислить: $\oint_{\Gamma} \frac{z dz}{(z-1)^2(z+1)}$, $\Gamma: |z-5i|=1$.
9. Вычислить: $\oint_{\Gamma} \frac{z dz}{(z-1)^2(z+1)}$, $\Gamma: |z-1|=1$.
10. Вычислить: $\oint_{\Gamma} \frac{z dz}{(z-1)^2(z+1)}$, $\Gamma: |z|=3$.
11. Для функции $f(z) = \frac{e^z - 1}{z^2(z+3)}$ найти особые точки и указать их характер.
12. Найти вычет функции $f(z) = z \cos \frac{1}{z}$ относительно точки $z = 0$ (использовать разложение в ряд по степеням z).

Кривая распределения относительных частот, полученная для теста по ТФКП, представлена на рисунке. Очевидно, что распределение индивидуальных баллов унимодально и не слишком асимметрично. При этом 70,2 % студентов выполнили от 4 до 9 заданий (то есть от 33 до 75 %), причем наиболее часто встречается результат 6 заданий (половина теста). Это соответствует об указанной градации нормального распределения и позволяет сделать вывод о том, что распределение баллов близко к нормальному.



Кривая распределения относительных частот тестовых баллов

Содержательная валидность определяется в качестве показателя «репрезентативности содержания теста по отношению к запланированным для проверки знаниям и умениям» [2].

Что касается рассматриваемого теста, то составители постарались включить в него наиболее важные темы рабочей программы курса. При этом учитывалась значимость поднимаемых вопросов, а также правильность пропорций при подборе заданий [8]. Выполнение данного теста требует достаточно глубоких знаний и серьезной подготовки со стороны студентов. Для нормативно-ориентированного теста содержательная валидность предполагает еще и высокую дифференциацию результатов – различающую способность.

Различающая способность (дискриминативность) – это возможность теста дифференцировать студентов по уровню знаний. Чем выше дифференцирующая способность теста, тем выше его валидность [7].

Один из способов оценки различающей способности состоит в рассмотрении ее числовой характеристики – точечного бисериального коэффициента корреляции r_{pb}^j . Это коэффициент корреляции каждого задания с тес-

товым баллом студента (индивидуальным баллом). Его величину можно рассчитать по одной из формул, дающих близкие значения [8, 9]:

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

Здесь \bar{X}_1 – средний индивидуальный балл студентов, справившихся с данным заданием; \bar{X}_0 – средний индивидуальный балл студентов, не справившихся с данным заданием; \bar{X} – среднее значение баллов по всей выборке; S_x – стандартное отклонение для индивидуальных баллов всех студентов; n_1 – число студентов, выполнивших данное задание; n_0 – число студентов, не выполнивших его; $n = n_1 + n_0$ – общее количество студентов; p – трудность задания (доля правильных ответов).

В результате расчетов r_{pb}^j по всем трем формулам получены идентичные результаты, представленные в табл. 2.

Таблица 2

Коэффициенты валидности и дискриминативности заданий

№	1	2	3	4	5	6	7	8	9	10	11	12
r_{pb}^j	0,396	0,44	0,58	0,355	0,547	0,581	0,569	0,52	0,447	0,599	0,365	0,283
$r_{диск}^j$	0,302	0,256	0,628	0,419	0,628	0,767	0,721	0,628	0,581	0,791	0,349	0,349

При оценке различающей способности теста абсолютно непригодными считают задания, для которых $r_{pb}^j \leq 0$, а кандидатами на удаление – те, для которых $r_{pb}^j < 0,2$ [2]. Таких заданий в тесте нет. В соответствии с классической теорией тестов [6, 7] приемлемым показателем качества считается значение $r_{pb}^j \geq 0,3$. В тесте это условие не выполняется для задания № 12 ($r_{pb}^{12} = 0,283$).

Согласно другой точке зрения [2], валидными считают задания с коэффициентами дискриминативности $r_{pb}^j \geq 0,5$. Половина заданий исследуемого теста удовлетворяют этому условию. В то же время среднее значение точечного бисериального коэффициента корреляции $(r_{pb}^j)_{cp.} = 0,473$, то есть незначительно меньше 0,5.

Рассмотрим подробнее задание № 12 – нахождение вычета в существенно особой точке. По мнению составителей, оно достаточно уникально и его нежелательно удалять из теста, что соответствует мнению некоторых авторов [7], считающих, что в этом случае можно сохранить задание, если даже $r_{pb}^j = 0,1$.

Другой способ оценки различающей способности связан с анализом коэффициента дискриминативности $r_{диск}^j$ задания j . Эта величина определяется как разность между долей правильных ответов p_1^j среди 27 % «лучших» и долей неправильных ответов p_0^j среди 27 % «худших» студентов [1, 2, 6, 7, 8]. Таким образом,

$$r_{диск}^j = p_1^j - p_0^j.$$

Согласно градации в классической тестовой теории [6] задание считается вполне эффективным при $r_{диск}^j \geq 0,4$; удовлетворительным, когда $0,3 \leq r_{диск}^j \leq 0,39$; если $0,2 \leq r_{диск}^j \leq 0,29$, то задание следует анализировать на пригодность; наконец, если $r_{диск}^j < 0,19$, то задание нужно изъять либо переработать.

Коэффициенты дискриминативности $r_{диск}^j$, представленные в последней строке табл. 2, позволяют сделать вывод о том, что только задание № 2 требует анализа на пригодность, восемь заданий попадают в разряд вполне эффективных и три задания – в разряд удовлетворительных. Скорее всего, задание № 2 (деление комплексных чисел) оказалось слишком простым и следует подумать о его переработке. Задание № 12, имеющее низкий точечный бисериальный коэффициент корреляции, на этот раз попадает в разряд удовлетворительных. Среднее значение коэффициента дискриминативности $(r_{диск}^j)_{ср.} = 0,535$.

Таким образом, оценивая валидность теста «Теория функций комплексной переменной» по нормальному распределению, по содержанию и по различающей способности, приходим к следующим выводам.

Проведенный анализ показывает, что кривая распределения баллов близка к нормальной кривой – распределению Гаусса.

Оценивая валидность по содержанию, можно отметить, что в тесте представлены основные темы раздела, требующие достаточно глубоких знаний по изучаемой дисциплине.

При исследовании различающей способности (дискриминативности) по точечному бисериальному коэффициенту r_{pb}^j оказалось, что в разряд при-

емлемых не попадает только задание № 12 (если в качестве критической точки брать 0,3). Если валидность оценивать по коэффициенту дискриминативности $r_{диск}^j$, то лишь задание № 2 требует анализа на пригодность. Предлагается сохранить задание № 12 и переработать задание № 2.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. 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. – 2011. – Вып. 4. – 28 с.
2. Чельщикова М.Б. Теория и практика конструирования педагогических тестов: Учеб. пособие. – М.: Логос, 2002. – 432 с.
3. Анастаси А. Психологическое тестирование. Кн. 1. – М.: Педагогика, 1982. – 320 с.
4. http://kpmmit.wl.dvgu.ru/library/aschepkova_testing17102001/index.shtml#contents
5. Психологическая диагностика: Учеб. пособие / Под ред. К.М. Гуревича, Е.М. Борисовой. – М.: Изд-во УРАО, 1997. – 304 с.
6. Переверзев В.Ю. Критериально-ориентированные педагогические тесты для итоговой аттестации студентов. – М.: НМЦ СПО Минобразования РФ, 1999. – 152 с.
7. Олейник Н.М. Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения: Учеб. пособие. – Донецк: ДонГУ, 1991. – 168 с.
8. Звонников В.И., Чельщикова М.Б. Современные средства оценивания результатов обучения. – М.: Академия, 2007. – 224 с.
9. Ким В.С. Тестирование учебных достижений. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.

Поступила в редакцию 15.06.17

В окончательном варианте 30.06.17

UDC 378.14

ANALYSIS AND ASSESSMENT OF THE VALIDITY DOUGH "THEORY OF FUNCTIONS OF THE COMPLEX VARIABLE"

*L.A. Muratova*¹

Samara State Technical University
244, Molodogvardeyskaya Str., Samara, 443100
E-mail: muratova-la@mail.ru

¹ *Lidiya A. Muratova*, Cand. Tech. Sci., Associate Professor of Higher Mathematics and Applied Informatics.

In article the analysis of quality of the "Theory of Functions of a Complex Variable" test is carried out. The dough validity in normal way to distribution, according to the contents, on the distinguishing ability is estimated. The look and proportions of a curve of distribution of the gained points testifies in favor of normal distribution. The test rather fully and deeply reflects the studied material that allows to draw a conclusion on its substantial validity. For quality check of dough on the distinguishing ability two numerical characteristics for each task are found: dot biserial coefficient and coefficient of a diskriminativnost'. The subsequent analysis showed that only two tasks don't satisfy to one of goodness conditions. Thus average value of that different authors offer was considered as a critical point. Having analyzed the maintenance of tasks, it is offered to process one of them, and to keep another.

Key words: *validity, diskriminativnost', normal distribution.*

BIBLIOGRAPHY

1. *Karpenko A.P., Domnikov A.S., Belous V.V.* Testovyi metod kontrolya kachestva obucheniya i kriterii kachestva obrazovatel'nyh testov [Test method of quality control of training and criteria of quality of educational tests]. Zhurnal. Nauka i obrazovanie: elektronnoe nauchno-tehnicheskoe izdanie. Vypusk 04/2011. 28 s.
2. *Chelyshkova M.B.* Teoriya i praktika konstruirovaniya pedagogicheskikh testov: Uchebnoe posobie [Theory and practice of designing of pedagogical tests: Manual.] M.: Logos, 2002. 432 s.
3. *Anastazi A.* Psihologicheskoe testirovanie. Kn. 1 [Psychological testing. Book 1]. M.: Pedagogika, 1982. 320 s.
4. http://kpmiit.wl.dvgu.ru/library/aschepkova_testing17102001/index.phtml#contents
5. Psihologicheskaya diagnostika. Uchebnoe posobie [Psychological diagnostics. The manual] / Pod red. K.M. Gurevicha, E.M. Borisovoi. M.: Izd-vo URAO, 1997. 304 c.
6. *Pereverzev V.Yu.* Kriterial'no-orientirovannye pedagogicheskie testy dlya itogovoi attestacii studentov [The criteria focused pedagogical tests for total certification of students]. M.: NMC SPO Minobrazovaniya RF, 1999. – 152 s.
7. *Oleinik N.M.* Test kak instrument izmereniya urovnya znaniy i trudnosti zadaniy v sovremennoi tehnologii obucheniya. Uchebnoe posobie [The test as the instrument of measurement of level of knowledge and difficulty of tasks in modern technology of training. Manual]: Donetsk, DonGU, 1991. 168 s.
8. *Zvonnikov V.I., Chelyshkova M.B.* Sovremennye sredstva ocenivaniya rezul'tatov obucheniya [Modern means of estimation of results of training]. M.: Akademiya, 2007. 224 s.
9. *Kim V.S.* Testirovanie uchebnyh dostizhenii [Testing of educational achievements]. Ussuriisk: Izd-vo UGPI, 2007. 214 s.

Original article submitted 15.05.17

Revision submitted 30.05.17