УДК 378.14

МОДЕЛИ БИРНБАУМА ДЛЯ ОЦЕНКИ КАЧЕСТВА ТЕСТА «ЛИНЕЙНАЯ АЛГЕБРА, АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ»

Л.А. Муратова[1]

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

Исследуется качество теста «Линейная алгебра, аналитическая геометрия» с помощью современной теории тестирования, а именно строятся две модели Бирнбаума: двухпараметрическая и трехпараметрическая. Двухпараметрическая модель в отличие от однопараметрической модели Раша содержит параметр, характеризующий дифференцирующую способность задания. В трехпараметрической модели появляется еще и третий параметр, учитывающий угадывание правильного ответа. На этом этапе выполняется анализ полученных по тесту результатов и в соответствии с рекомендациями делается вывод об использовании тех или иных заданий. Далее строятся информационные функции отдельных заданий и всего теста. При этом используется тот факт, что количество информации (полученное с помощью данного задания) об уровне знаний обратно пропорционально стандартной ошибке измерения этого уровня. Результат по информационной функции говорит о качественности тестовых заданий.

Ключевые слова: *модель Раша, двухпараметрическая и трехпараметрическая модели Бирнбаума, дифференцирующая способность задания, информационная функция.*

В рамках современной теории тестирования (IRT), как правило, рассматриваются три модели: однопараметрическая (модель Раша), двух- и трехпараметрические модели Бирнбаума [1–5].

В статье [6] для теста «Линейная алгебра, аналитическая геометрия» была построена однопараметрическая модель Раша. При этом были использованы формулы, согласно которым вероятность $P_j(\theta)$ правильного ответа испытуемых с разным уровнем подготовки θ на j задание теста равна

$$P_j(\theta) = (1 + \exp\left(-1{,}7\left(\theta - \beta_j\right)\right))^{-1},$$

[1] *Лидия Александровна Муратова*, к.т.н., доцент кафедры «Высшая математика и прикладная информатика».

а вероятность $P_i(\beta)$ правильного выполнения различных по трудности β заданий i испытуемым равна

$$P_i(\beta) = (1 + \exp(-1{,}7(\theta_i - \beta)))^{-1}.$$

Выполним еще два построения: построим двухпараметрическую и трехпараметрическую модели Бирнбаума для того же теста.

Двухпараметрическая модель отличается от модели Раша тем, что содержит параметр a_j, характеризующий дифференцирующую способность j задания. В этом случае вероятность правильного ответа испытуемых с разным уровнем подготовки θ на j задание теста определяется формулой [1–5]

$$P_j(\theta) = (1 + \exp\left(-1{,}7a_j(\theta - \beta_j)\right))^{-1}.$$

Формула для a_j имеет вид

$$a_j = \frac{(r_{bis})_j}{\sqrt{1-((r_{bis})_j)^2}},$$

где $(r_{bis})_j$ – коэффициент бисериальной корреляции j задания. По ряду причин [5] вместо этого коэффициента в силу схожести используют точечный бисериальный коэффициент r_{pb}^j – коэффициент корреляции каждого задания с тестовым баллом студента (индивидуальным баллом)

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

Здесь n_1 – число студентов, выполнивших данное задание;

n_0 – число студентов, не выполнивших его; $n = n_1 + n_0$ – общее количество студентов;

$\overline{X}_1$ – средний индивидуальный балл студентов, справившихся с данным заданием (отношение суммы индивидуальных баллов студентов, справившихся с данным заданием, к n_1);

$\overline{X}_0$ – средний индивидуальный балл студентов, не справившихся с данным заданием (отношение суммы индивидуальных баллов студентов, не справившихся с данным заданием, к n_0);

s_x – стандартное отклонение для индивидуальных баллов всех студентов.

Итак, будем считать, что

$$a_j \approx \frac{r_{pb}^j}{\sqrt{1-(r_{pb}^j)^2}}.$$

Параметр a_j прямо пропорционален тангенсу угла наклона характеристической кривой $P_j(\theta)$ в точке перегиба. Чем больше значение этого параметра, тем больше крутизна характеристической кривой и, значит, больше дифференцирующая способность задания.

Поэтому в процессе создания нормативно-ориентированного теста, предполагающего сравнение уровней знаний тестируемых между собой [1, 5], принципиальным является отбор заданий в зависимости от значений параметра a_j.

В таблицу помещены значения r_{pb}^j (взяты из статьи [7]), вычисленные по приведенной выше формуле значения a_j, а также трудности заданий β_j (формулы для вычисления взяты из статьи [6]).

Расчетные параметры для построения характеристических кривых

Номер задания	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
r_{pb}^j	0,29	0,55	0,36	0,45	0,51	0,29	0,44	0,54	0,42	0,52	0,35	0,39	0,55	0,64	0,59	0,60	0,59	0,58	0,37
a_j	0,30	0,66	0,39	0,50	0,59	0,30	0,49	0,64	0,46	0,61	0,37	0,42	0,66	0,83	0,73	0,75	0,73	0,71	0,40
β_j	-1,35	-0,67	-0,98	-1,41	0,17	0,24	-0,72	0,01	0,17	-0,31	0,57	-0,10	-0,57	0,83	1,39	0,35	0,46	0,35	1,12

Рассмотрим рекомендации по отбору заданий, следуя которым можно получить тест с хорошей различающей способностью, что, как правило, приводит к повышению его надежности и валидности [5].

Необходимо совсем отказаться от заданий с отрицательными значениями параметра a_j (в данном тесте таких заданий нет), считающихся бесполезными. Это связано с тем, что на них хорошо отвечают испытуемые с низким уровнем знаний и плохо – с высоким уровнем знаний, что противоречит здравому смыслу.

Кроме того, нужно отбирать задания с достаточно большими значениями a_j – из интервала $(0{,}5; 2{,}5)$. В тесте с этой точки зрения самыми плохими будут задания 1, 3, 6, 11.

Дальнейший анализ предполагает отбор заданий с наибольшей дифференцирующей способностью при равной трудности.

Рассмотрим задания 5 и 9, имеющие одинаковую трудность $\beta_5 = \beta_9$ и отличающиеся параметром a_j: $a_5 = 0{,}59$, $a_9 = 0{,}46$. Согласно однопараметрической модели Раша оба задания имеют одну и ту же кривую вероятностей правильного ответа испытуемых (кривая 1, рис. 1), то есть с точки зрения дифференцирующей способности задания неразличимы. В случае двухпараметрической модели получим две различные характеристические кривые: более крутую (2) для задания 5 и менее крутую (3) для задания 9. Таким образом, при минимизации длины теста предпочтительно задание 5.

Одинаковую трудность имеют еще два задания: 16 и 18. В этом случае можно оставить задание 16, имеющее чуть большее значение параметра a_j.

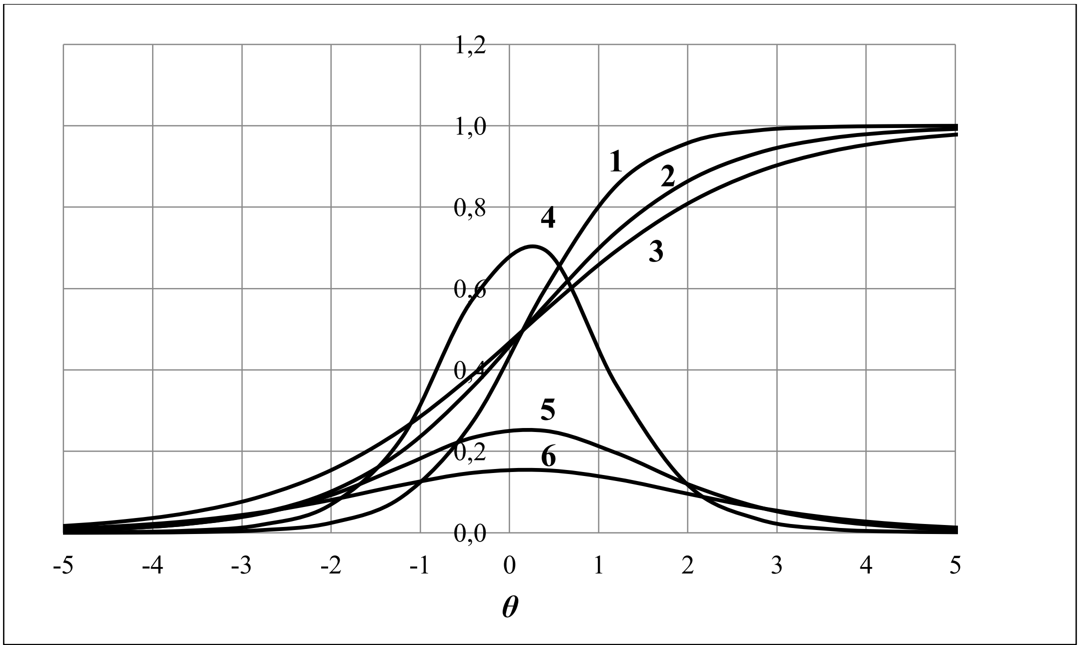

Рис. 1. Модель Раша и двухпараметрическая модель Бирнбаума для заданий 5 и 9

Рассматриваемый тест относится к тестам закрытого типа с выбором единственного правильного ответа из пяти предложенных для каждого задания. В таких случаях, чтобы снизить эффект угадывания, предлагается использовать трехпараметрическую модель Бирнбаума [1–5].

Перейдем к ее построению.

Эта модель Бирнбаума содержит еще один параметр c_j, характеризующий вероятность правильного ответа на задание j в том случае, если этот ответ угадан, а не основан на знаниях. При этом вероятность правильного ответа испытуемых на задание j теста выражается формулой

$$P_j(\theta) = c_j + (1 - c_j)(1 + \exp(-1{,}7a_j(\theta - \beta_j)))^{-1},$$

где $c_j = \frac{1}{k_j}$, k_j – число ответов на задание j. В рассматриваемом тесте $k_j = 5$, $c_j = 0{,}2$.

Нижние асимптоты характеристических кривых заданий этой модели проходят через точки $c_j > 0$, поэтому сами характеристические кривые становятся более пологими. Дифференцирующая способность теста в этом случае снизится.

На рис. 2 изображены кривые вероятностей правильного ответа испытуемых на задание 5 в зависимости от уровня подготовки θ, построенные согласно двухпараметрической модели (кривая 1) и трехпараметрической модели (кривая 2).

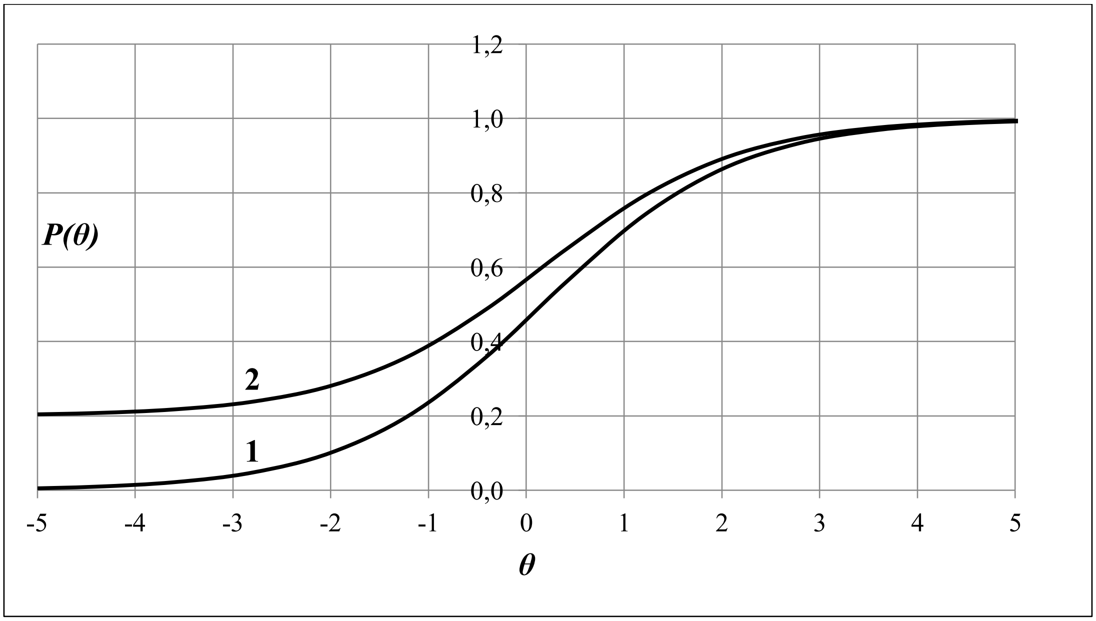

Рис. 2. Характеристические кривые задания 5 согласно моделям Бирнбаума

Каждое задание j дает некоторое количество информации об уровне знаний θ, и эта информация обратно пропорциональна стандартной ошибке измерения θ с помощью данного задания [3]. Чтобы описать информацию, соответствующую заданию j, вводится в рассмотрение информационная функция $I_j(\theta)$ [3, 5]:

$$I_j(\theta) = \frac{(P_j'(\theta))^2}{P_j(\theta)Q_j(\theta)},$$

где $Q_j(\theta) = 1 - P_j(\theta)$.

Информационная функция выступает в качестве характеристики степени пригодности задания и может быть использована для повышения эффективности тестовых измерений, позволяя вычислить дифференцированную стандартную ошибку измерения. Это выгодно отличает современную теорию тестовых измерений от классической [5].

Для модели Раша информационная функция определяется соотношением

$$I_j(\theta) = 2{,}89P_j(\theta)Q_j(\theta).$$

Для двухпараматрической модели Бирнбаума

$$I_j(\theta) = 2{,}89a_j^2 P_j(\theta)Q_j(\theta).$$

Информационная функция $I_j(\theta)$ для трехпараметрической модели Бирнбаума имеет вид

$$I_j(\theta) = \frac{2{,}89a_j^2(1-c_j)}{(c_j+\exp(1{,}7a_j(\theta-\beta_j)))(1+\exp(-1{,}7a_j(\theta-\beta_j))^2}.$$

Максимальное значение информационной функции для модели Раша и двухпараметрической модели Бирнбаума достигается в точке перегиба характеристической кривой, то есть когда трудность (в логитах) равна уровню знаний (в логитах). Таким образом, для θ_i наиболее информативны задания со значениями трудности β из окрестности точки θ_i.

На рис. 1 для заданий 5 и 9 теста построены информационные функции: по модели Раша (кривая 4 – общая для двух заданий), по двухпараметрической модели для задания 5 (кривая 5) и для задания 9 (кривая 6). Трудность заданий $\beta_5 = \beta_9 = 0{,}13$, поэтому эти задания наиболее информативны для θ, близких к значению 0,13.

В случае трехпараметрической модели Бирнбаума максимум информационной функции достигается в точке [3, 5]

$$\theta_{max} = \beta_j + \frac{1}{1{,}7a_j}\left(0{,}5 + 0{,}5\sqrt{1+8c_j}\right).$$

Так, для задания 5 с трудностью $\beta = 0{,}17$ максимум информационной функции достигается в точке $\theta_{max} = 1{,}46$, для задания 9 с той же трудностью – в точке $\theta_{max} = 1{,}83$.

Информационная функция всего теста получается в результате суммирования всех $I_j(\theta)$:

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta).$$

Информационная функция всего теста должна иметь один четко выраженный максимум, иначе тест нуждается в доработке, в него нужно добавлять задания с трудностями, соответствующими областям провала информационной функции [3, 5].

На рис. 3 представлены информационные функции $I(\theta)$ всего теста, построенные согласно модели Раша (кривая 1), двухпараметрической модели (кривая 2), трехпараметрической модели (кривая 3). В рассматриваемом тесте это условие выполняется – каждая кривая имеет одну точку максимума.

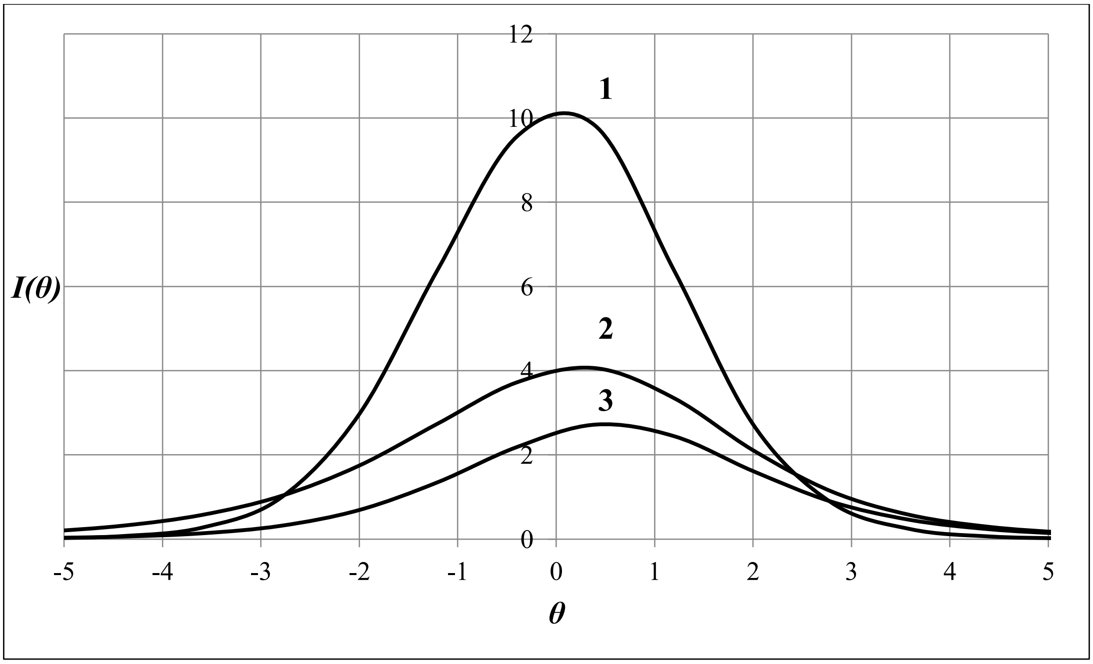

Рис. 3. Информационные функции теста

Итак, для теста «Линейная алгебра, аналитическая геометрия» построены две модели Бирнбаума: двух- и трехпараметрическая. Если считать, что тест должен соответствовать этим моделям, необходимо избавиться от некоторых заданий, а другие изменить. В частности, с точки зрения нормативно-ориентированного теста он должен обладать достаточно высокой дифференцирующей способностью. Поэтому следует убрать задания 9 и 18, одинаковые по трудности с заданиями 5 и 16 соответственно и отличающиеся более низкой дифференцирующей способностью. Кроме того, необходимо заменить или изменить задания 1, 3, 6, 11 так, чтобы их дифференцирующая способность возросла. Что касается информационных функций, графики, построенные по трем моделям, включая модель Раша, не выявили противоречий между теорией и экспериментом.

Отметим, что в результате предлагаемых преобразований может пострадать содержательная сторона теста, может нарушиться его целостность. Дело в том, что рассматриваемый тест совмещает в себе черты нормативно- и критериально-ориентированного теста, а используемые модели предназначены для нормативно-ориентированных тестов. При составлении теста ставилась задача не только ранжировать студентов по степени их подготовленности, но и выяснить, как усвоен и усвоен ли вообще каждый раздел учебной программы. Конечно, при этом желательно выяснить, не слишком ли сложен или слишком прост тест, какова трудность заданий.

Как оценивать эффективность такого теста? Есть ли компромиссное решение? На этот счет существуют различные точки зрения [1, 3, 5].Опираясь на точку зрения авторов, считающих, что компромисс возможен [1, 5], будем искать его, ни в коем случае не отказываясь от полученных рекомендаций при корректировке теста с учетом всех факторов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Жилина Е.В.* Анализ применяемых моделей и методов тестирования для оценки знаний специалиста // Zprávy vědecké ideje – 2011: materiály VII mezinárodní vědecko-praktická konference. 27 října – 05 listopadu 2011 roku. Díl 4. Ekonomické vědy. Praha: Publishing House «Education and Science» s.r.o. 2011. С. 53–62.
2. *Звонников В.И., Челышкова М.Б.* Современные средства оценивания результатов обучения. – М.: Академия, 2007. – 224 с.
3. *Ким В.С.* Тестирование учебных достижений. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.
4. *Олейник Н.М.* Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения: Учеб. пособие. – Донецк: ДонГУ, 1991. – 168 с.
5. *Челышкова М.Б.* Теория и практика конструирования педагогических тестов: Учеб. пособие. – М.: Логос, 2002. – 432 с.
6. *Муратова Л.А.* Построение однопараметрической модели Раша для теста «Линейная алгебра, аналитическая геометрия» // Вестник СамГТУ. Сер. Психолого-педагогические науки. – 2016. – № 4(32). – С. 46–54.
7. *Муратова Л.А.* Качественные изменения педагогического теста «Линейная алгебра, аналитическая геометрия» // Вестник СамГТУ. Сер. Психолого-педагогические науки. – 2016. – № 2(30). – С. 82–88.

Поступила в редакцию 22.08.22017
В окончательном варианте 24.09.2017

UDC 378.14

BIRNBAUM'S MODELS FOR THE ASSESSMENT OF QUALITY OF DOUGH «LINEAR ALGEBRA, ANALYTICAL GEOMETRY»

L.A. Muratova[1]

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

Quality of the "Linear Algebra, Analytical Geometry" test by means of the modern theory of testing is investigated, namely, two models of Birnbaum are under construction: two-parametrical and three-parametrical. The two-parametrical model unlike one-parametrical

[1] *Lidia A. Muratova*, Cand. of Tech. Sci., Associate Professor of Higher Mathematics and Applied Informatics Department

model Russia contains the parameter characterizing the differentiating ability of a task. In three-parametrical model appears also the third parameter considering guessing of the correct answer. At this stage the analysis of the results received according to dough is made and according to recommendations the conclusion about use of these or those tasks is drawn. Further information functions of separate tasks and all dough are based. Thus is used that fact that the amount of information (received by means of this task) about the level of knowledge, in inverse proportion to a standard error of measurement of this level. The result on information function speaks about goodness of test tasks.

Keywords: *the Russia model, two-parametrical and three-parametrical models of Birnbaum differentiating ability of a task, information function.*

REFERENCES

1. *Zhilina E.V.* Analiz primenyaemyh modelei i metodov testirovaniya dlya ocenki znanii specialista [The analysis of the applied models and methods of testing for an assessment of knowledge of the expert] Zprávy vědecké ideje – 2011: materiály VII mezinárodní vědecko-praktická konference. 27 října - 05 listopadu 2011 roku. Díl 4. Ekonomické vědy. Praha: Publishing House "Education and Science" 2011, pp. 53-62.
2. *Zvonnikov V.I., Chelyshkova M.B.* Sovremennye sredstva ocenivaniya rezul'tatov obucheniya. [Modern means of estimation of results of training.] M.: Akademiya, 2007. 224 p.
3. *Kim V.S.* Testirovanie uchebnyh dostizhenii. **[**Testing of educational achievements.**]** Ussuriisk: Izd-vo UGPI, 2007. 214 p.
4. *Oleinik N.M.* Test kak instrument izmereniya urovnya znanii i trudnosti zadanii 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]: Doneck, DonGU, 1991. 168 p.
5. *Chelyshkova M.B.* Teoriya i praktika konstruirovaniya pedagogicheskih testov: Uchebnoe posobie. [Theory and practice of designing of pedagogical tests: Manual.] M.: Logos, 2002. 432 p.
6. *Muratova L.A.* Postroenie odnoparametricheskoi modeli Rasha dlya testa «Lineinaya algebra, analiticheskaya geometriya» [Creation of one-parametrical model Russia for the "Linear Algebra, Analytical Geometry" test]. Vestnik SamGTU, Seriya «Psihologo-pedagogicheskie nauki». [Messenger SAMGTU, Psychology and Pedagogical Sciences Series.]. No. 4(32)-2016, pp 46-54.
7. *Muratova L.A.* Kachestvennye izmeneniya pedagogicheskogo testa «Lineinaya algebra, analiticheskaya geometriya» i ih ocenka [High-quality changes of the pedagogical test "Linear Algebra, Analytical Geometry" and their assessment]. Vestnik SamGTU. Seriya «Psihologo-pedagogicheskie nauki». [The Messenger of SAMGTU. Psychology and Pedagogical Sciences series]. No. 2(30)-2016, pp. 82-86.

Original article submitted 22.08.2017
Revision submitted 24.09.2017