УДК 37.016:51

ИССЛЕДОВАНИЕ ТРУДНОСТИ И ДИФФЕРЕНЦИРУЮЩЕЙ СПОСОБНОСТИ КОНКУРСНЫХ ЗАДАНИЙ

THE STUDY OF DIFFICULTY AND DIFFERENTIAL ABILITY OF COMPETITIVE ITEMS

Лазарев В. А., Хайбуллин Р. Я.

Уфимский государственный нефтяной технический университет, г. Уфа, Российская Федерация

V. A. Lazarev, R. Ya. Khaybullin

Ufa State Petroleum Technological University, Ufa, Russian Federation

e-mail: lazva@mail.ru, khayrya@mail.ru

Аннотация. Данная статья посвящена методу оценки относительной трудности и дифференцирующей способности заданий, предлагаемых в рамках одного теста, конкурса, предметной олимпиады. Основным объектом исследования, на основе которого строятся все выводы, является эмпирическая функция успеха решения задания. Степень согласованности ее с теоретической функцией, принятой в Item Response Theory (IRT), служит количественным признаком дифференцирующей способности конкретного задания. Входящие в модель параметры характеризуют трудность задания и указывают, относительно какого уровня подготовки производится разделение студентов по их подготовке. Численной характеристикой соответствия теоретической и эмпирической функций служит сумма квадратов разностей их значений на равномерной сетке. Проверка гипотезы адекватности проводится с помощью критерия согласия, критические значения которого определяются в численных экспериментах методом Монте-Карло. Уровень значимости интерпретируется как показатель валидности задания, его соответствия уровню подготовке участников олимпиады. Исследование мотивировано необходимостью точного оценивания характеристик олимпиадных задач с целью более обоснованного определения победителей. В заключении приведен пример практического анализа задач организованной УГНТУ олимпиады по математике среди школьников.

Abstract. This article is devoted to the method of assessing both the relative difficulty and differentiating ability of any items offered within the one test, competition or thematic Olympiad. The main object of research, which all the conclusions are based on, is the empirical function of the success of the solution. The adequacy with the theoretical function accepted in the Item Response Theory (IRT) gives as a quantative measure of the differentiating ability of the particular problem. The parameter in the model characterizes the difficulty of the task and indicates the level of preparation relatively which the division is made. A numerical measure of the likelihood theoretical and empirical functions is the sum of squares of deviations of their values on a uniform grid. The adequacy test is carried out with the help of some consent criterion. For obtaining the criteria distributions we apply the numerical simulations according to the Monte Carlo approach. The level of significance is interpreted as an indicator of the reliability of the obtained results. The study is motivated by the need for accurately assessing the difficulty of the Olympiad problems in order to determine the winners more reasonably. Finally, an example of practical analysis of the math problems from the school Olympiad organized by USPTU is given as concluding remarks.

Ключевые слова: дифференцирующая способность, функция успеха, IRT, метод Монте-Карло, статистические гипотезы, закон распределения критерия.

Key words: discrimination power, item discrimination, IRT, Monte Carlo method, statistical hypothesis, criteria distributions.

Тематические олимпиады, конкурсы прочно вошли в учебный процесс не только средней, но и высшей школы, выполняя целый ряд функций, самая важная из которых – выявление наиболее одаренных школьников, студентов по соответствующим дисциплинам. Как и все соревнования по олимпийской системе, предметные олимпиады (и школьные, и студенческие) проходят в несколько этапов при достаточно жестком отборе между этапами: в следующий тур выходит один из примерно 5-10 участников. В силу этого особенно актуален вопрос о качестве задач, судействе, о правилах отбора победителей и призеров. Тем более что к предметным олимпиадам, как инструменту конкурсного отбора, все активней в последнее время прибегают высокотехнологичные компании и корпорации, такие как Росатом, Газпром, Транснефть, чтобы выявить и привлечь потенциально перспективных специалистов для обучения в свои опорные вузы. Сами вузы также проводят свои олимпиады (как правило, одноэтапные) с целью дать преференции наиболее успешным ученикам. Участие во многих конкурсах, смотрах, олимпиадах помогает молодым людям лучше осознать свои способности, раскрыть свой потенциал, найти наиболее перспективный путь дальнейшего саморазвития и карьеры. Такая ситуация удобна для всех, поэтому есть основания считать, что эта тенденция сохранится.

Несмотря на широкое внедрение методов педагогических измерений в практику учителя и преподавателя, подбору задач к олимпиадам и их анализу в литературе практически не уделяется внимания. По инерции считается, что учебные программы достаточно однородны и постоянны во времени, как и уровень знаний обучающихся. Нередко организаторы вообще не видят особой проблемы в подборе задач, считая, что более сильный участник любой набор задач решит лучше, чем более слабый. А так как ставится именно задача ранжирования участников, то может показаться, что набор задач не столь существенен. Это всё очевидные для практикующих преподавателей заблуждения. Система образования, несмотря на приписываемую ей инертность, достаточно чутко реагирует на изменения в структуре мирового производства, экономике, технологиях, социальном устройстве общества. Меняются учебные программы, образовательные стандарты, глубина изучения тех или иных разделов, уровень логической обоснованности, смещаются акценты с одних компетенций на другие и так далее. Вследствие этого необходимо корректировать тематику задач, их уровень, номинальные баллы. Причем делать это исходя не из прошлого опыта, а из реальных результатов, показанных участниками «здесь и сейчас». Более того, зачастую участники олимпиады не проходят никакого предварительного конкурсного отбора, к участию допускаются все желающие. Это накладывает на конкурсное задание дополнительное требование дифференцирования участников по способностям в очень широком диапазоне. Тем более что в некоторых олимпиадах призеры и дипломанты могут составлять около 50 % от общего числа участников. А так как время, отводимое на решение задач, ограничено, следовательно, ограничено число задач и возникает закономерный вопрос о надежности получаемых результатов.

Эти аргументы находят подтверждение и в практической работе. Например, в 2017 году коэффициент ранговой корреляции Спирмена между результатами вузовского тура и 1-го тура Интернет-олимпиады оказался равным 0,2. Конечно, корреляция в результатах олимпиад заметно меньше, чем для промежуточного тестирования, тем не менее, этот факт заставил задуматься о надежности результатов проводимых олимпиад и мотивировал данное исследование.

Методика исследования. Основным объектом исследования, на основе которого строятся все выводы, является эмпирическая функция успеха решения задания. Согласование ее с теоретической функцией, принятой в IRT, служит качественным признаком дифференцирующей способности конкретного задания.

Согласно общему определению, дифференцирующая способность (ДС) задания – это способность заданий теста выявлять сильных и слабых учащихся, дифференцировать испытуемых по их подготовке. При анализе дифференцирующей способности выделяются два подхода, традиционный и современный. Первый основывается на статистическом подходе [1], второй – на положениях IRT [2, 3].

Ключевое место в IRT занимает функция успеха – функциональная зависимость, указывающая вероятность правильного выполнения дихотомического задания в зависимости от уровня трудности этого задания с одной стороны и уровня подготовленности испытуемого – с другой. Начиная с работы Г. Раша [2], функция успеха задается формулой

$$P(s) = \frac{1}{1+e^{-\delta(s-s_0)}}, \quad (1)$$

где $\delta -$ показатель ДС, $s, s_0 -$ сила участника и трудность задания, выраженные в одной шкале (графики представлены на рисунке 1).

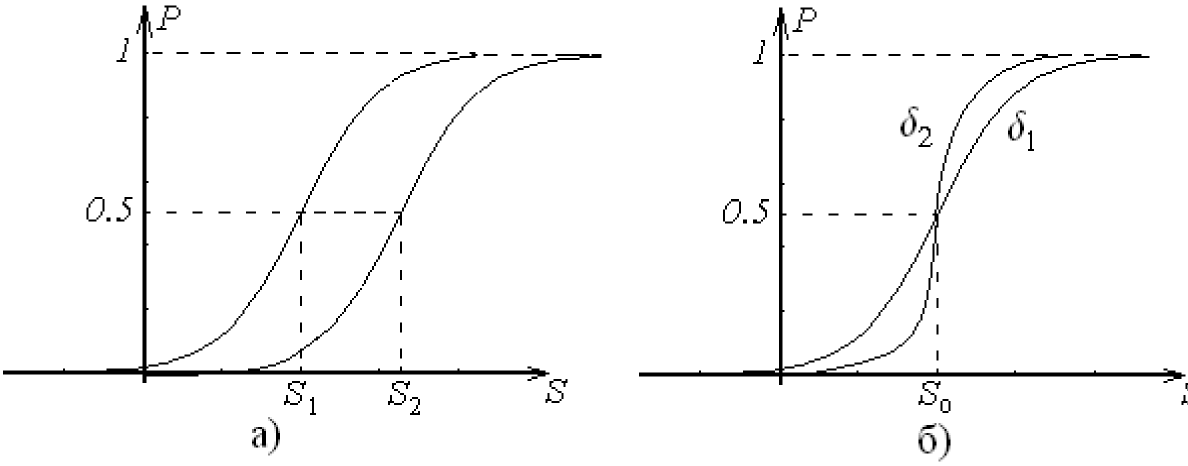


Рисунок 1. Графики функций успеха: а) $s_2 > s_1$, б) $\delta_2 > \delta_1$

Трудность задания s_0 означает, что участник с уровнем подготовки s_0 решит это задание с вероятностью 0,5. При увеличении s_0 график функции успеха смещается вправо (рисунок 1, *а*). Дифференцирующая способность δ характеризует, насколько быстро изменяется вероятность в окрестности $s = s_0$. Если δ велико, то при увеличении s $(s > s_0)$ вероятность успеха очень быстро стремится к 1. Если s убывает и $s < s_0$, то вероятность решения быстро стремится к нулю. При небольших значениях δ вероятность возрастает (убывает) медленно, что делает задачу мало эффективной для диагностики уровня подготовки участника (рисунок 1, *б*).

Зачастую в прикладных исследованиях в области педагогических измерений функция успеха принимается в качестве аксиомы, исходя из которой проверяются те или иные гипотезы, положения и т.д. [4-7]. В рамках этого подхода и выполнено настоящее исследование. Мы полагаем, что результаты качественного теста должны соответствовать теоретической модели, а не наоборот, как свойственно математическому моделированию. Так если результаты тестирования не согласуются с IRT, то это означает, что подбор задач в целом или его отдельные задания имеют некоторые недостатки. Дискуссию об обоснованности, корректности этого подхода мы оставляем в стороне.

Анализ разбивается на два этапа. На первом этапе методом наименьших квадратов из статистики работ определяются параметры функции успеха: δ и s_0. Эти значения будут характеризовать ДС и уровень сложности задачи соответственно. На втором этапе мы проверяем гипотезу о соответствии эмпирической функции успеха теоретической.

Итак, будем считать, что на олимпиаде K участникам предложено N задач, которым присвоен директивный балл B_j, сумма всех директивных баллов $B_\Sigma = \sum_{j=1}^{N} B_j$. Результаты участников олимпиады, полученные после проверки работ, будем представлять в следующем виде (таблица 1).

Таблица 1. Первичные результаты проверки работ

	Задание 1, B_1	Задание 2, B_2		Задание N, B_N	Результат B_Σ
Участник 1	α_{11}	α_{12}	…	α_{1N}	S_1
Участник 2	α_{21}	α_{22}		α_{2N}	S_2
⋮	⋮	⋮		⋮	⋮
Участник K	α_{K1}		…	α_{KN}	S_K

Величины α_{ij} $(i = 1..K,\ j = 1..N)$ равны доле от максимального балла, полученного i-м участником за решение j-й задачи. Они выставляются проверяющими при ручной проверке. Для тестовых заданий с выбором ответа α_{ij} могут принимать только два значения – ноль и один. Подготовленность участника характеризуется суммарным баллом за все решенные задачи:

$$S_i = \sum_{j=1}^{N} \alpha_{ij} B_j. \tag{2}$$

Эти величины мы возьмем в качестве оценки способности студентов.

Для каждой конкретной задачи строится гистограмма вероятностей. Интервал $[0, B_\Sigma]$ разбивается на n равных частей

$$\Delta_q = [h(q-1), hq], \quad h = \frac{B_\Sigma}{n}, \quad q = 1, \dots, n.$$

Пусть рассматривается -я задача. Для каждого диапазона Δ_q вычисляется средний балл, полученный только теми участниками, чей суммарный балл попал в Δ_q. Этот балл, деленный на B_j, трактуется как вероятность решения задачи студентами, чей суммарный балл попал в Δ_q. Обозначим эту вероятность через $\bar{\alpha}_{jq}$. На основе гистограммы строится эмпирическая функция успеха $\hat{P}_j(x)$ в виде (1), дающая наилучшее приближение в смысле метода наименьших квадратов (МНК).

Отметим, что нам приходится несколько отойти от классического изложения теории в смысле работ [2, 8]. Это связано с желанием остаться в естественной шкале способностей, определяемой первичным баллом после проверки. Переходить к логитам мы посчитали нецелесообразно (хотя это и не представляет какой-либо трудности), так как решается локальная проблема качества задач на конкретной олимпиаде среди конкретного контингента. Задачи сравнения результатов, стандартизации в каком-либо смысле (см. обсуждение, например, в [7]) в данной работе не рассматриваются. Таким образом, способности участников изменяются в диапазоне $[0, B_\Sigma]$. Выбрать эту шкалу позволяет наличие в модели параметра δ, который к тому же имеет важную интерпретацию. Трудность заданий также принадлежит этому отрезку $[0, B_\Sigma]$, хотя теоретически может выходить за эти границы.

После определения параметров эмпирической функции успеха возникает вопрос об её адекватности теоретической функции успеха. Ответ на этот вопрос мы получаем традиционным способом с помощью критерия согласия

$$\theta_j = \sum_{j=1}^{n} \frac{(\overline{\alpha}_{jq} - p_{jq})^2}{p_{jq}}, \quad p_{jq} = \frac{1}{|\Delta_q|} \int_{\Delta_q} \hat{P}_j(x) dx .$$

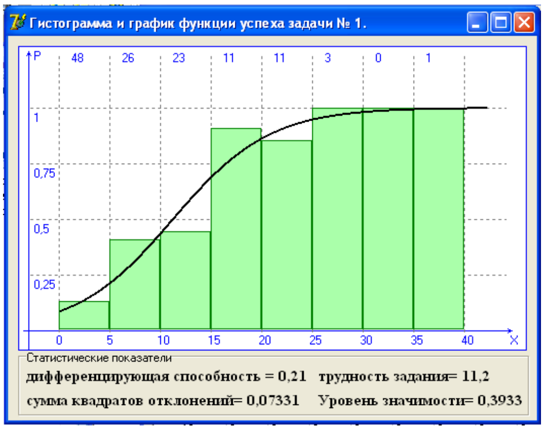
Функция распределения критерия согласия, необходимая для принятия гипотезы, вычисляется специально для каждой задачи методом Монте-Карло. Исходя из предположения истинности гипотезы, проводится серия численных экспериментов, по результатам которых для конкретного значения N восстанавливается с достаточной точностью закон распределения критерия. Сравнивая наблюдаемое по результатам олимпиады значение критерия с полученными в численных экспериментах, мы можем получить уровень значимости при котором принимается гипотеза о соответствии эмпирической функции успеха кривой (1). В российской статистической традиции в качестве значения уровня значимости выбирается число 0,05.

Анализ заданий олимпиады УГНТУ среди школьников. Олимпиада состоялась в апреле 2017 года. Предлагалось 8 задач, каждая из которых была оценена в пять баллов. Анализ проводился только среди работ, набравших не менее одного балла (123 работы). Тексты задач, их показатели и обсуждение результатов мы приводим ниже. Отметим, что в наших исследованиях достаточно хорошими показателями значимости мы считаем значения от 0 до 0.15. Это связано с наличием дополнительных факторов, которые невозможно отразить в модели, например, распределение вероятности промежуточных баллов в политомической модели Раша, влияние ручной проверки разными преподавателями, приближенный характер критических значений. Также уместно напомнить о проблеме директивных оценок для задач [8, 9]. Их выбор существенно влияет на полученные студентами итоговые баллы и, следовательно, на эмпирические функции успеха. Таким образом, неправильная оценка трудности заданий может дополнительно исказить статистические показатели.

Таблица 2. Численные характеристики олимпиадных задач и их интерпретация.

1. Решить систему уравнений

$$\begin{cases} x^3 + xyz = \sqrt{xyz}, \\ y^3 + xyz = \sqrt{xyz}, \\ z^3 + xyz = \sqrt{xyz}. \end{cases}$$



Анализ задания. Относительно простое задание, содержащее только одну «ловушку» – потерю корня, показывает плохое соответствие эмпирической гистограммы теоретической кривой (уровень значимости всего 0,3933). Очевидно, внимательность не всегда коррелирует с прочими компетенциями.

2. Решить уравнение

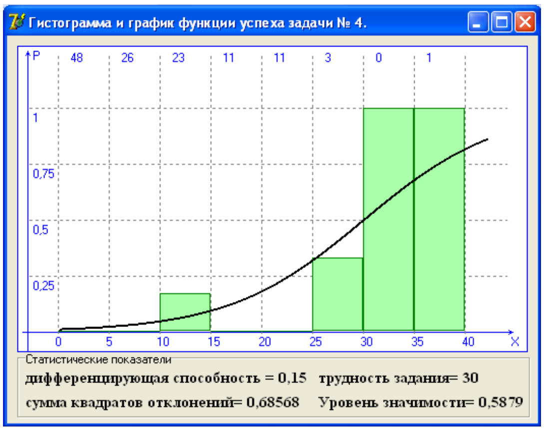
$$(4\cos^2 x + 4\sqrt{3}\cos x + 7)\cdot(4\log_3^2(tg\, y) - \log_3(ctg^4\, y) + 9) = 32.$$

Анализ задания. Задание требовало знаний свойств элементарных функций, знакомства со специфичным методом решения, умения решать модельные логарифмические и тригонометрические уравнения. Вследствие этого трудность оказалась на высоком уровне 28,4. Уровень значимости 0,01 свидетельствует о прекрасной дифференцирующей способности задания среди хорошо подготовленных учеников.

3. Доказать, что выражение $(5x+7y)^3+(7x+5y)^3$ делится без остатка на $12(x+y)$.

Анализ задания. Это «утешительное» задание оказалось рекордным по ДС и уровню значимости. Хотя трудность 7,2 означает дифференцирование только среди слабых участников, это задание в олимпиаде, на наш взгляд, было уместно, так как в призеры помимо прочих вошли и участники, набравшие 6 баллов из 40.

4. Точка O – центр вписанной окружности треугольника ABC, точка D – середина стороны AB. Известно, что угол OAD – прямой. Докажите равенство $AB+BC=3AC$.



Анализ задания. Геометрия всегда была и остается проблемным разделом школьного математического образования. Данная задача, которая решается с помощью дополнительного построения, оказалась непосильной для большинства участников. Фрагментарные решения у различного уровня участников отражены в слабом уровне значимости 0,59. Задача была бесполезна для определения победителей, и мы считаем, что задачу по планиметрии следовало бы предложить более простую.

5. Дано $\sin x, \sin y, \sin z$ – возрастающая арифметическая прогрессия. Докажите, что числа $\cos x, \cos y, \cos z$ не образуют (строго) убывающую арифметическую прогрессию.

Анализ задания. Данная задача решалась нестандартным построением на единичной окружности и также не соответствовала уровню участников. Вероятность решения во всех группах оказалась мала и не имела ярко выраженной тенденции. Задачи 4 и 5, несомненно, были бы полезны при более сильном подборе участников.

6. При каких значениях a прямая $y = ax - 5$ касается кривой $y = 3x^2 - 4x - 2$?

Анализ задания. По гистограмме видно хорошее соответствие эмпирических и теоретических значений вероятностей, однако вследствие ручной проверки, многоэтапности решения, необходимости перехода от геометрии к методам анализу, уровень значимости не мал – 0,1619. Тем не менее, эту задачу на олимпиаде следует признать уместной, как задачу с хорошей дифференцирующей способностью среднего уровня сложности. Ученикам уровня выше среднего (>20) каких-либо затруднений эта задача доставить не может.

7. Если образующая конуса равна $\sqrt{3}$ см, расстояние от вершины конуса до центра вписанного в него шара равно 1 см, то чему равен угол между образующей и плоскостью основания конуса?

Анализ задания. Стереометрические задачи такого сорта, в которых грамотно проведенное сечение сводит решение к относительно простой планиметрической задаче – прекрасная альтернатива сложной планиметрии. Этот тезис частично подтверждается и числовыми характеристиками: хорошая дифференцирующая способность (0,17), немаленькая трудность (15,2).

8. Решите неравенство $\left||x|-|2x-7|\right|<1$ и найдите сумму всех целых решений.

Анализ задания. Алгебраические уравнения и неравенства с модулем – классика различного уровня олимпиад, турниров, тестов. Простота метода, сводящего сложную задачу к серии простых, но требующих известной аккуратности при их решении, отражается в наличии продвижений у слабых участников и не гарантирует отсутствие ошибок у сильных. Все это точно характеризуется относительно слабой дифференцирующей способностью (0,1), медианной трудностью (22,8 из 40) при высоком уровне значимости (0,0715).

Выводы

Авторами разработан метод и компьютерная программа, позволяющие дать численные оценки дифференцирующей способности задания и его трудности. Отслеживать эти величины приходится, чтобы в условиях неопределенности относительно состава участников олимпиады предложить такие задания, которые бы позволили с высокой степени надежности выявить действительно сильнейших участников. Наш подход сочетает в себе теоретическую обоснованность, практическую реализуемость, легкость интерпретации результатов. Полученные в примере результаты, с одной стороны, подтвердили наши представления относительно предложенных на конкретной олимпиаде задач, а с другой – дали некоторую пищу для размышлений об улучшении олимпиадных заданий.

Список используемых источников

1 Челышкова М. Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. 432 с.

2 Rasch G. Probabilistic Model for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Research, 1960. 126 p.

3 Linden W., Hambleton R. Handbook of Modern Item Response Theory. NY: Springer Verlag. 1997. 510 p.

4 Цыганов Ш. И. Математические методы педагогических измерений. // Вестник Башкирского университета. 2009. Т. 14, № 3. С. 1263-1270.

5 Саяпин А. В., Сафонов К. В. Оценка дифференцирующей способности компьютерного теста методами имитационного моделирования // Вестник Красноярского государственного педагогического университета им. В. П. Астафьева. 2012. № 2. С. 138-143.

6 Сафаров Р. Х., Панищев О. Ю. Численное моделирование инвариантности оценки знания относительно трудности тестовых заданий в рамках модели Г. Раша // Образовательные технологии и общество. 2012. Т. 15. № 1. С. 424-435.

7 Белобородов В. Н., Татур А. О. Применение современной теории тестирования IRT в системе контроля измерительных свойств диагностических материалов // Педагогические измерения. 2016. № 2. С. 85-97.

8 Лазарев В. А. Итерационная модель оценки результатов предметных олимпиад // Педагогическая информатика. 2017. № 1. С. 3-9.

9 Лазарев В. А., Хайбуллин Р. Я. Метод статистической оценки относительной сложности олимпиадных и тестовых задач // Нефтегазовое дело: электрон. науч. журн. 2014. № 5. С. 420-430. URL: http://ogbus.ru/issues/5_2014/ogbus_5_2014_p420-430_LazarevVA_ru.pdf.

References

1 Chelyshkova M. B. Teoriya i praktika konstruirovaniya pedagogicheskikh testov. M.: Logos, 2002. 432 s. [in Russian].

2 Rasch G. Probabilistic Model for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Research, 1960. 126 p.

3 Linden W., Hambleton R. Handbook of Modern Item Response Theory. NY: Springer Verlag. 1997. 510 p.

4 Tsyganov Sh. I. Matematicheskie metody pedagogicheskikh izmerenii. // Vestnik Bashkirskogo universiteta. 2009. T. 14, № 3. S. 1263-1270. [in Russian].

5 Sayapin A. V., Safonov K. V. Otsenka differentsiruyushchei sposobnosti komp'yuternogo testa metodami imitatsionnogo modelirovaniya // Vestnik Krasnoyarskogo gosudarstvennogo pedagogicheskogo universiteta im. V. P. Astaf'eva. 2012. № 2. S. 138-143. [in Russian].

6 Safarov R. Kh., Panishchev O. Yu. Chislennoe modelirovanie invariantnosti otsenki znaniya otnositel'no trudnosti testovykh zadanii v ramkakh modeli G. Rasha // Obrazovatel'nye tekhnologii i obshchestvo. 2012. T. 15. № 1. S. 424-435. [in Russian].

7 Beloborodov V. N., Tatur A. O. Primenenie sovremennoi teorii testirovaniya IRT v sisteme kontrolya izmeritel'nykh svoistv diagnosticheskikh materialov // Pedagogicheskie izmereniya. 2016. № 2. S. 85-97. [in Russian].

8 Lazarev V. A. Iteratsionnaya model' otsenki rezul'tatov predmetnykh olimpiad // Pedagogicheskaya informatika. 2017. № 1. S. 3-9. [in Russian].

9 Lazarev V. A., Khaibullin R. Ya. Metod statisticheskoi otsenki otnositel'noi slozhnosti olimpiadnykh i testovykh zadach // Neftegazovoe delo: elektron. nauch. zhurn. 2014. № 5. S. 420-430. URL: http://ogbus.ru/issues/5_2014/ogbus_5_2014_p420-430_LazarevVA_ru.pdf. [in Russian].

Сведения об авторах

About the authors

Лазарев В. А., канд. физ.-мат. наук, доцент кафедры «Математика», ФГБОУ ВО «УГНТУ», г. Уфа, Российская Федерация.

V. A. Lazarev, Candidate of Physics and Mathematics, Associate Professor of the Mathematics Department, FSBEI HE «USPTU», Ufa, Russian Federation.

e-mail: lazva@mail.ru

Хайбуллин Р. Я., канд. техн. наук, доцент кафедры «Математика», ФГБОУ ВО «УГНТУ», г. Уфа, Российская Федерация.

R. Ya. Khaybullin, Candidate of Engineering Sciences, Associate Professor of the Mathematics Department, FSBEI HE «USPTU», Ufa, Russian Federation.

e-mail: khayrya@mail.ru