УДК: 51-7
Doi: 10.17223/16095944/66/10

В.М. Карнаухов
Российский государственный аграрный университет,
Москва, Россия

КОРРЕКЦИЯ ПЕРВИЧНЫХ БАЛЛОВ ПРИ ПОМОЩИ НЕЧЕТКИХ МНОЖЕСТВ

В последние 20 лет активно развивается теория нечетких множеств, результаты которой широко используются и в тестировании. В статье исследуется точность метода оценки уровня подготовленности учащегося, основанного на использовании нечетких множеств. Для сравнения в качестве контрольных методов были выбраны два метода, используемые в ЕГЭ: метод шкалирования и метод логарифма Раша. В частности, автором была выявлена решающая роль выбора функции шкалирования, влияющего на точность выставляемых оценок при помощи вышеперечисленных методов.

Ключевые слова: модель Раша, метод Монте-Карло, функция шкалирования, метод первичных баллов, латентные параметры, уровень подготовленности, нечеткие множества.

В настоящее время в образовании распространены различные системы автоматизированного (компьютерного) контроля знаний. К таким системам можно отнести системы, функционирующие на Едином государственном экзамене (ЕГЭ), при аттестации студентов вузов, при проведении текущих и итоговых контрольных работ в различных учебных заведениях (школы, колледжи, вузы). В этих системах используются различные методики получения итоговых оценок знаний (уровней подготовленности) учащихся.

К наиболее часто используемым методикам можно отнести следующие: методику шкалирования первичных баллов (КМШ – классический метод шкалирования) [5], методики перевода первичных баллов в латентные параметры тестирования уровней подготовленности учащихся либо при помощи метода логарифмирования первичных баллов (МЛР – метод логарифма Раша, ранее назывался методом первичных баллов) [4, 5], либо при помощи метода моментов (ММР – метод моментов Раша) [5]. Краткий обзор некоторых из вышеперечисленных методов дан ниже.

За последние 20 лет в тестировании широко используется теория нечетких множеств, которая также может быть использована при оценке знаний учащихся [3]. В данной статье приведены результаты исследования точности метода нечетких множеств (МНМ), основанного на использовании теории нечетких множеств. В качестве контрольных методов были выбраны методы КМШ, МЛР, а также метод модифицированного шкалирования (ММШ), который изложен ниже.

Описание современных методов оценки знаний

В классической методике шкалирования результатов ЕГЭ, используемой в 2011–2014 гг., реализуется поэтапное установление соответствия тестовых и первичных баллов для каждого общеобразовательного предмета, по которому проводится ЕГЭ.

I этап.

Сначала в диапазоне первичных баллов от нуля до максимального первичного балла ПБmax для каждого общеобразовательного предмета ЕГЭ выбираются два значения первичных баллов: ПБ1 и ПБ2, разделяющих группы участников с различным уровнем подготовки по данному предмету.

Величина ПБ1 выбирается как наименьший первичный балл, получение которого свидетельствует об усвоении участником экзамена основных понятий и методов по соответствующему общеобразовательному предмету. Он определяется на основе экспертизы демонстрационного варианта по данному общеобразовательному предмету специалистами общего образования, ссузов и вузов различного профиля из разных субъектов РФ. Экспертиза осуществляется с учетом уровня сложности каждого задания и значимости проверяемого им содержания, умения, навыка, способа деятельности в контексте общеобразовательного предмета. При этом требования к значению ПБ1 соответствуют требованиям, которые использовались при определении ПБ1 прошлого года (для обеспечения эквивалентности шкал двух лет).

Величина ПБ2 определяется профессиональным сообществом как наименьший первичный балл, получение которого свидетельствует о высоком уровне подготовки участника экзамена, а именно о наличии системных знаний, владении комплексными умениями, способности выполнять творческие задания по соответствующему общеобразовательному предмету.

II этап.

Первичным баллам ПБ1 и ПБ2 ставятся в соответствие тестовые баллы ТБ1 и ТБ2 (уровни подготовленности учащихся, выраженные в процентах) по каждому общеобразовательному предмету согласно табл. 1.

Таблица 1

Значения граничных первичных и тестовых баллов в 2013 г.

Предмет	ПБ1	ТБ1	ПБ2	ТБ2
Русский язык	17	36	54	73
Математика	5	24	15	63
Обществознание	15	39	48	72
История	13	32	47	72
Физика	12	39	33	62
Химия	14	36	58	80
Биология	17	36	60	79
География	14	37	43	69
Информатика	8	40	35	84
Иностранные языки	16	20	65	82
Литература	8	32	36	73

III этап.

По каждому общеобразовательному предмету определяется соответствие между первичным и тестовым баллом на основе следующей процедуры. Первичному баллу 0 ставится в соответствие тестовый балл 0, а максимальному первичному баллу ПБmax ставится в соответствие тестовый балл 100.

Все промежуточные первичные баллы между 0, ПБ1, ПБ2 и ПБmax переводятся в тестовые, пропорционально распределенные между соответствующими значениями тестовых баллов: 0, ТБ1, ТБ2 и 100. На рис. 1 представлена получаемая зависимость сплошной ломаной линией.

Модифицированный метод шкалирования отличается от классического метода шкалирования

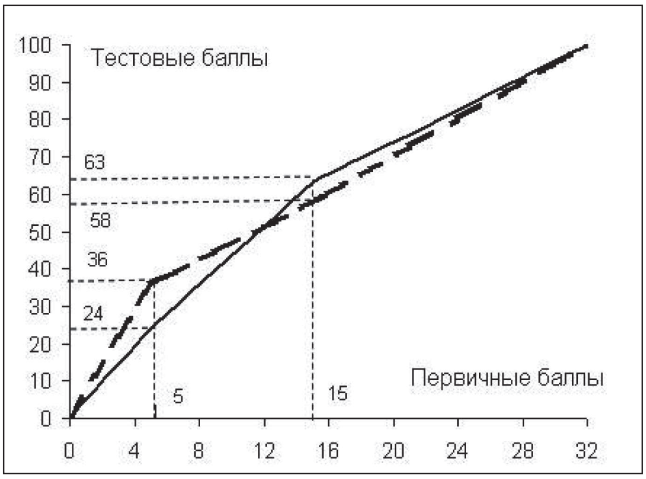

Рис. 1. Соответствие между первичными и тестовыми баллами для методов КМШ и ММШ

тем, что на этапе II подбираются другие значения ТБ1 и ТБ2, позволяющие повысить точность выставляемых оценок. Результаты исследований, связанных с подбором оптимальных значений ТБ1 и ТБ2, приведены в работе [5] и представлены штрихованной ломаной линией на рис. 1. Используя на практике метод ММШ, можно добиться выигрыша в точности примерно в 2,3 %.

В качестве дополнительного контрольного метода был выбран метод логарифма Раша (МЛР) (см. [4]), который по точности, простоте и устойчивости является самым привлекательным методом из всех вышеперечисленных. Согласно этому методу вначале вычисляются оценки $\bar{\theta}_k$, $\bar{\delta}_j$ латентных параметров тестирования по формулам

$$\bar{\theta}_k = \ln\left(\frac{k}{K-k} K_1\right),$$

причем $K_1 = \dfrac{\sum\limits_{k=0}^{K} (K-k) \cdot N_k}{\sum\limits_{k=0}^{K} k \cdot N_k}$,

$k = 1, \ldots, K-1$,

K = ПБmax – максимальный первичный балл;

N_k – число учащихся, набравших ПБ = k первичных баллов;

$\bar{\theta}_0 = -\theta_{max}$, $\bar{\theta}_K = \theta_{max}$, $(\theta_{max} = 5)$;

$$\overline{\delta}_j = \ln\left(\frac{N \cdot m_j - c_j}{c_j} \cdot K_2\right),$$

причем $K_2 = \dfrac{\sum_{j=1}^{M} c_j}{\sum_{j=1}^{M} N \cdot m_j - c_j};$

$j = 1, \ \ldots, M$,

M – число заданий теста;

N – число участников тестирования: $N = \sum_{k=0}^{K} N_k;$

m_j – максимальный балл, получаемый за решение j-го задания,

c_j – первичный балл для j-го задания, равный количеству всех баллов, набранных всеми N участниками тестирования.

А затем оценки латентных параметров переводятся в тестовые баллы по формуле

$$T = \frac{\theta + \theta_{\max}}{2 \cdot \theta_{\max}} \cdot 100\%.$$

Описание метода нечетких множеств (МНМ)

Очевидно, что выставляемые баллы участникам тестирования за решение задач теста не отражают действительную картину уровней знаний учащихся. Например, полученный нулевой балл за решение задачи совершенно не означает, что учащийся, решавший эту задачу, имеет «нулевые знания» по данной теме. Конечно, в этом случае необходимо заменить нулевую оценку на положительный (в смысле числа, большего нуля) балл. Но какой? Для корректировки выставляемых баллов можно использовать теорию нечетких множеств.

Согласно этой теории необходимо рассмотреть лингвистическую переменную B =«балл, выставляемый учащемуся за решение задачи», с заданным терм-множеством:

B_0=«B=0» – учащийся набрал за решение данной задачи 0 баллов;

B_1=«B=1» – учащийся набрал за решение данной задачи 1 балл;

.....

B_m=«B=max» – учащийся набрал за решение данной задачи максимальное число баллов, которое устанавливается экспертами.

Элементам этого множества соответствуют нечеткие множества, определенные на отрезке U=[0,1], с функциями принадлежности $\mu_i(x)$,

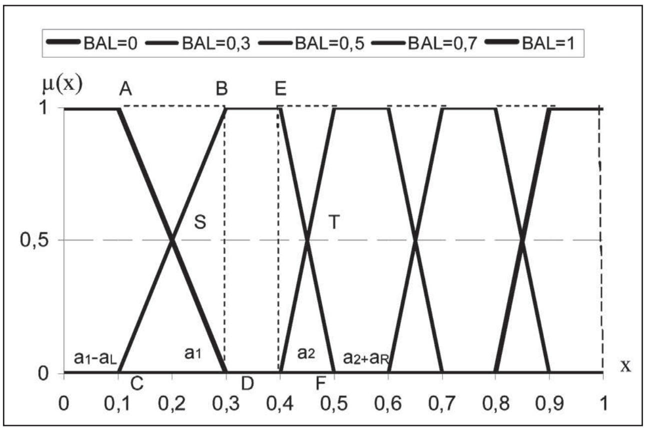

Рис. 2. Функции принадлежности для баллов ЕГЭ

i=0,..., m, примерные графики которых изображены на рис. 2.

Каждому элементу терм-множества ставится в соответствие нечеткое множество, определенное на отрезке [0, 1], так как любой набранный балл B можно перевести в относительный балл по формуле u=B/max. Таким образом, введенные нечеткие множества на U (см. рис. 2) можно использовать для заданий теста с различными установленными максимальными баллами. Согласно теории нечетких множеств [3] вышеупомянутая лингвистическая переменная должна принадлежать семейству полных ортогональных семантических пространств (ПОСП). А именно, функции принадлежности, соответствующие элементам терм-множества лингвистической переменной, должны удовлетворять следующим свойствам:

1) для каждого B_i, i=0,..., m существует непустое множество («неоспоримая зона») U_i={x∈U: $\mu_i(x)$=1}, которое является либо точкой, либо отрезком;

2) любая функция $\mu_i(x)$, i=0,..., m не убывает слева от множества U_i и не возрастает справа от этого множества;

3) функции $\mu_i(x)$, i=0,..., m имеют не более двух точек разрыва первого рода;

4) для любого значения $x \in U$ существует хотя бы одна функция $\mu_i(x)$, i=0,..., m, для которой $\mu_i(x) \neq 0$;

5) для любого значения $x \in U$ $\sum_{i=0}^{m} \mu_i(x) = 1$.

Значение функции принадлежности $\mu_i(x)$, которое в теории нечетких множеств называется степенью принадлежности значения x нечеткому множеству B_i, можно понимать как вероятность того события B_i, что значение x принадлежит множеству B_i. Напомним, что степень принадлежности равна доле тех экспертов, которые причисляют данное значение x к множеству B_i, поэтому она равна относительной частоте, а значит, вероятности вышесформулированного события.

В силу вероятностного понимания степени принадлежности можно прокомментировать сформулированные 5 свойств следующим образом:

1) для каждого балла B_i существуют «неоспоримые зоны» относительного балла, при появлении которого любой эксперт выставляет балл B_i;

2) двигаясь влево от «неоспоримой зоны» или вправо от нее, эксперты с меньшей уверенностью выставляют соответствующий балл;

3) баллы могут выставляться экспертами по заранее четко сформулированным правилам;

4) за любой набранный относительный балл хотя бы один из экспертов должен начислить определенное количество баллов;

5) за любой набранный относительный балл каждый из экспертов должен начислить определенное количество баллов.

Заметим, что свойство 4 следует из свойства 5.

В работах О.М. Полещук [3] рассчитаны формулы для функций принадлежности (см. рис. 2) при помощи T-чисел, которые приведены ниже в алгоритме. Напомним, что толерантным $(L–R)$-числом называется нечеткое множество с функцией принадлежности вида

$$\mu(x)=\begin{cases} L\left(\dfrac{a_1-x}{a_L}\right), & 0<\dfrac{a_1-x}{a_L}\le 1,\ a_L>0, \\ R\left(\dfrac{x-a_2}{a_R}\right), & 0<\dfrac{x-a_2}{a_R}\le 1,\ a_R>0, \\ 0, & x<a_1-a_L, \\ 0, & x>a_2+a_R \end{cases}$$

и символически записывается в виде $\mu(x)=(a_1, a_2, a_L, a_R)$. При этом отрезок $[a_1, a_2]$ называется интервалом толерантности, а a_L и a_R – соответственно левым и правым коэффициентами нечеткости $(L–R)$-числа. Функция $L\left(\dfrac{a_1-x}{a_L}\right),\ \ 0<\dfrac{a_1-x}{a_L}\le 1,$ называется левой границей числа, а функция $R\left(\dfrac{x-a_2}{a_R}\right),\ \ 0<\dfrac{x-a_2}{a_R}\le 1,$ – правой границей. При $a_L=0$ левая граница равна 0, а при $a_R=0$ правая граница обращается в 0. При $a_1=a_2$ толерантное число превращается в унимодальное и обозначается как $\mu(x)=(a_1, a_L, a_R)$. Если $L(x)=R(x)=1-x$, то $(L–R)$-число называется T-числом, а унимодальное число называется нормальным треугольным числом.

Отметим также, что алгоритм нечетких множеств является адаптивным алгоритмом в том смысле, что для построения функций принадлежности используются результаты тестирования в виде набранных первичных баллов. А именно, предварительно подсчитываются относительные частоты p_{ij} появления балла $B=j$ при решении i-го задания, $i=1, \ldots, M$ (M – число заданий теста), $j=0,\ldots,$ max. Затем функции принадлежности формируются так, чтобы площади криволинейных трапеций, образуемых этими функциями, равнялись p_{ij}.

Итак, приступим к изложению алгоритма метода нечетких множеств, позволяющего «подправлять» первичные баллы.

1) После проведения тестирования проводится статистическая обработка полученной информации (первичных баллов) с целью получения частот p_{ij} (см. выше).

2) Для каждого задания строится лингвистическая переменная со следующими функциями принадлежности (для сокращения обозначим p_{ij} через p_j):

для множества $B_i = 0$ имеем

a) если $p_0 \le p_1$, то $\mu_0(x) = (0, p_0/2, 0, p_0)$;

b) если $p_0 > p_1$, то $\mu_0(x) = (0, p_0-p_1/2, 0, p_1)$;

для множества $B_i = k$, $k = 1, \ldots, m-1$ ($m = \max$) имеем

a) если $p_k \ge \max(p_k-1, p_k+1)$, то

$$\mu_k(x)=(\sum_{r=0}^{k-1}p_r+\frac{p_{k-1}}{2}, 1-1{,}5p_{k+1}, p_{k-1}, p_{k+1});$$

b) если $p_k+1 < p_k < p_k-1$, то

$$\mu_k(x)=(1-p_{k+1}-\frac{p_k}{2}, 1-1{,}5p_{k+1}, p_k, p_{k+1});$$

c) если $p_k-1 < p_k < p_k+1$, то

$$\mu_k(x)=(\sum_{r=0}^{k-1}p_r+\frac{p_{k-1}}{2}, 1-p_{k+1}-0{,}5p_k, p_{k-1}, p_k);$$

d) если $p_k \le \min(p_k-1, p_k+1)$, то

$$\mu_k(x)=(1-p_{k+1}-\frac{p_k}{2}, p_k, p_k);$$

для множества B_i=m имеем

a) если $p_m \le p_{m-1}$, то $\mu_m(x)=(1-p_m/2, 1, p_m, 0)$;

b) если $p_m > p_{m-1}$, то $\mu_0(x)=(1-p_m+p_{m-1}/2, 1, p_{m-1}, 0)$.

3) Для каждой функции принадлежности вычисляется число E_k, $k = 0, \ldots, m$, получающееся деффазификацией нечеткого числа по методу центра тяжести:

$$E_k = \frac{\int_{a_1-a_L}^{a_2+a_R} x \cdot \mu_k(x)dx}{\int_{a_1-a_L}^{a_2+a_R} \mu_k(x)dx} =$$

$$= \frac{a_2^2 - a_1^2 + a_1 a_L + a_2 a_R + \frac{1}{3}(a_R^2 - a_L^2)}{2(a_2 - a_1) + a_L + a_R}$$

4) Производится «корректировка» набранного учащимся числа баллов B за i-е задание по формуле

$$B_{\text{кор}} = \max \cdot \sum_{k=0}^{\max} E_k \cdot \mu_k\left(\frac{B}{\max}\right).$$

5) Вычисляется сумма всех «откорректированных» баллов:

$$ПБ_{кор} = \sum_{i=1}^{M} B_{\text{кор}}^{i}.$$

6) Вычисляется тестовый балл $ТБ_{\text{кор}}$ при помощи шкалирования, используемого в методах КМШ или ММШ (см. выше).

Компьютерное моделирование тестирования

Описанный выше алгоритм является частью компьютерной программы, моделирующей процесс тестирования ЕГЭ при помощи метода Монте-Карло. Обсудим некоторые наиболее важные элементы этой программы.

Программа моделирует процесс тестирования для абитуриентов в количестве $N = 500$ и теста, состоящего из $M = 20$ заданий. Тест ЕГЭ в 2011–2013 гг. состоял из 14 заданий с max=1 (B1–B14), 2 заданий с max=2 (C1 и C2) , 2 заданий с max=3 (C3 и C4) и двух заданий с max=4 (C5 и C6).

Процесс ЕГЭ моделируется достаточно большое количество раз (число итераций $N_{it} = 30$). Для каждого моделирования вычисляются две характеристики:

1) среднее отклонение $\sigma_{\text{ср}}$ оценки уровня подготовленности абитуриента от истинного значения этого латентного параметра;

2) наибольшее отклонение $\sigma_{\max}$ оценки уровня подготовленности от истинного значения этого латентного параметра.

Далее вычисленные характеристики усреднялись по всем итерациям.

Процесс компьютерной имитации тестирования осуществляется следующим образом. Вначале моделируются истинные уровни подготовленности участников θ_i, $i=1,\ldots,N$ и истинные уровни трудностей заданий δ_j, $j=1,\ldots,M$. Уровни подготовленности участников смоделированы как реализации нормальной случайной величины N(0,1) по формуле $\theta_i = F_N^{-1}(ri)$, где $F_N(x)$ – функция распределения нормированной нормальной случайной величины, т.е. $N(0,1)$, которая определяется по формуле

$$F_N(x) = 0{,}5 + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx,$$

$F_N^{-1}(ri)$ – обозначение функции, обратной к функции $F_N(x)$. Значение обратной функции вычисляется в точке r_i, представляющей собой очередную реализацию датчика случайных чисел на отрезке (0,1).

В силу правила 3 сигм все реализации выше определенной случайной величины будут находиться в интервале: $\theta_i \in (-3;3)$.

Уровни трудностей заданий смоделированы как реализации нормальных случайных величин $\left(\Delta = \frac{0{,}1}{3}\right)$:

$\delta_j \in N(-2;\Delta)$, $j=1,\ldots,5$,
$\delta_j \in N(-1;\Delta)$, $j=6,\ldots,10$,
$\delta_j \in N(0;\Delta)$, $j=11,\ldots,14$,
$\delta_j \in N(1;\Delta)$, $j=15,\ldots,16$,
$\delta_j \in N(2;\Delta)$, $j=17,\ldots,18$,
$\delta_j \in N(3;\Delta)$, $j=19,\ldots,20$.

В силу правила 3 сигм и малости Δ задания с одним номером в различных вариантах будут мало отличаться друг от друга.

Для каждого абитуриента и для каждого задания вычисляются первичные баллы. Для этого по формуле [1, 2]

$$P_{ij} = \frac{1}{1 + e^{-(\theta_i - \delta_j)}}, \quad i = 1.,\ldots,N, \quad j = 1,\ldots,M$$

вычисляется вероятность p, с которой i-й абитуриент правильно решает j-е задание. Затем абитуриенту начисляется первичный балл S за решение задания по формуле

Таблица 2

Точность трех методов: КМШ, МЛР, МНМ
(для этого метода меняются средние значения для крайних баллов)

Δ_1 / Δ_2 (в %)	0	20	40	60	80	100
0	7,05 4,59 9,55	7,10 4,56 10,81	7,08 4,62 12,49	6,98 4,58 14,09	7,06 4,56 15,79	7,11 4,59 17,59
20	6,96 4,60 7,14	7,07 4,59 7,98	7,04 4,65 9,15	7,05 4,59 10,52	7,07 4,58 12,06	7,05 4,57 13,80
40	7,07 4,62 5,52	7,13 4,63 5,81	6,99 4,56 6,50	7,00 4,62 7,50	7,06 4,60 8,68	7,10 4,62 10,00
60	7,11 4,63 5,49	7,14 4,61 5,02	7,12 4,51 4,97	7,07 4,59 5,37	7,04 4,64 6,04	7,05 4,54 6,96
80	7,07 4,60 7,18	7,01 4,62 6,09	7,06 4,57 5,21	7,04 4,58 4,73	7,06 4,63 4,70	7,12 4,58 5,13
100	7,06 4,61 11,35	6,97 4,60 9,53	7,06 4,64 7,79	7,03 4,58 6,70	7,10 4,62 5,66	7,05 4,53 4,93

$$S = \begin{cases} 0, & r \ge p \\ \left[\dfrac{r \cdot \max}{p}\right] + 1, & r < p, \end{cases}$$

где r – очередная реализация датчика случайных чисел на (0;1);

max – максимальное число баллов за решение задачи;

$[x]$ – целая часть числа x.

Основные результаты

При помощи описанной выше программы проводились исследования зависимости точности метода нечетких множеств от удаленности значений E_0 и E_m от 0 и 1 соответственно. В табл. 1 приведены значения погрешности метода МНМ в зависимости от величин Δ_1 и Δ_2: $0\% \le \Delta_1$, $\Delta_2 \le 100\,\%$, которые характеризуют вышеупомянутые удаленности в соответствии с формулами:

$E_{0,\text{кор}} = E_0(1 - \Delta_1/100)$,

$E_{m,\text{кор}} = E_m + (1 - E_m)\cdot\Delta_2/100$.

Для сравнения в каждой ячейке таблицы приведены значения погрешности методов КМШ, МЛР, МНМ, которые расположены в каждой ячейке в указанном порядке.

В дополнение к этой таблице вычислена погрешность для метода ММШ, в среднем точности для методов ММШ, МЛР и МНМ (Δ_1=80 %, Δ_2=80 %) оказались следующими: 4,68; 4,62; 4,67.

Комментарий к табл. 2:

1) наивысшая точность метода МНМ достигается для значений $\Delta_1 = 80\,\%$ и Δ_2=80 %, при этом $E_{0,кор} = E_0/5$, $E_{m,кор} = 0{,}8 + E_m/5$;

2) сравнивая методы ММШ, МЛР и МНМ, можно утверждать, что все три метода по точности примерно одинаковы.

В представленной статье проведено исследование влияния функции шкалирования (см. КМШ и ММШ) на точность метода МНМ (Δ_1 = 80 % и Δ_2 = 80 %), который был рассмотрен в двух вариантах.

В первом варианте первичный балл $ПБ_{кор}$ шкалируется сразу (см. п. 6) алгоритма МНМ).

Во втором варианте первичный балл $ПБ_{кор}$ вначале преобразуется в модифицированный первичный балл $ПБ_{кор}^{мод}$ с учетом точного диапазона его изменения по формуле

$$\text{ПБ}_{\text{кор}}^{\text{мод}} = \frac{\text{ПБ}_{\text{кор}} - \text{ПБ}_{\min}}{\text{ПБ}_{\max} - \text{ПБ}_{\min}} \cdot MAX,$$

где $\text{ПБ}_{\min} = \sum_{i=1}^{M} E_{0i} \cdot \max_i$;

$\text{ПБ}_{\max} = \sum_{i=1}^{M} E_{mi} \cdot \max_i$; $MAX = \sum_{i=1}^{M} \max_i$ ($\max_i$ – максимальное количество баллов, начисляемое за правильное решение i-й задачи теста), а затем преобразуется в тестовый балл так же, как в методике ММШ.

Таблица 3

Исследование влияния шкалирования на точность методов ММШ и двух вариантов метода МНМ

ТБ1/ТБ2	53	54	55	56	57	58	59	60	61	62
31	6,10 5,76 7,41	5,94 5,52 7,14	5,66 5,27 6,88	5,47 5,03 6,63	5,36 4,85 6,29	5,27 4,80 6,07	5,34 4,77 5,88	5,39 4,81 5,86	5,45 4,80 5,69	5,60 4,98 5,72
32	5,99 5,67 7,24	5,62 5,34 6,78	5,49 5,18 6,55	5,29 4,94 6,29	5,12 4,73 5,99	5,08 4,69 5,75	5,13 4,71 5,67	5,22 4,78 5,57	5,28 4,88 5,50	5,38 4,99 5,44
33	5,59 5,47 6,82	5,40 5,24 6,55	5,14 4,92 6,11	5,06 4,80 5,90	4,97 4,71 5,72	4,94 4,65 5,57	4,96 4,71 5,46	5,00 4,66 5,29	5,17 4,85 5,25	5,34 4,96 5,29
34	5,49 5,49 6,65	5,17 5,11 6,24	5,03 4,89 5,92	4,87 4,66 5,57	4,79 4,60 5,45	4,82 4,62 5,32	4,88 4,67 5,21	4,94 4,74 5,11	5,09 4,88 5,15	5,33 5,11 5,17
35	5,31 5,35 6,30	5,11 5,14 6,04	4,86 4,88 5,71	4,83 4,77 5,51	4,68 4,64 5,22	4,70 4,57 5,06	4,80 4,69 5,03	5,00 4,80 5,04	5,06 4,93 5,00	5,30 5,15 5,07
36	5,12 5,23 6,10	4,89 4,95 5,67	4,80 4,89 5,52	4,64 4,69 5,20	4,67 4,64 5,10	4,71 4,72 4,99	4,76 4,72 4,91	4,90 4,87 4,87	5,08 5,05 4,87	5,31 5,25 4,97
37	5,03 5,30 5,92	4,86 5,05 5,66	4,62 4,77 5,23	4,64 4,80 5,19	4,60 4,71 4,96	4,63 4,71 4,86	4,76 4,88 4,78	4,95 5,02 4,83	5,11 5,20 4,84	5,40 5,44 4,91
38	4,89 5,14 5,67	4,76 5,03 5,40	4,61 4,87 5,21	4,59 4,81 4,98	4,64 4,83 4,95	4,68 4,89 4,77	4,74 4,97 4,77	4,96 5,15 4,75	5,19 5,28 4,79	5,49 5,59 4,92
39	4,87 5,20 5,55	4,63 4,98 5,26	4,64 4,96 5,09	4,57 4,91 4,97	4,57 4,87 4,80	4,74 4,99 4,81	4,86 5,13 4,74	5,04 5,26 4,76	5,34 5,53 4,85	5,68 5,88 5,12
40	4,86 5,31 5,51	4,76 5,12 5,25	4,61 4,97 5,00	4,61 4,99 4,90	4,71 5,07 4,85	4,87 5,20 4,81	4,92 5,22 4,76	5,24 5,57 4,89	5,43 5,70 4,92	5,72 6,01 5,06

Таблица 4

Экстремальные значения методов

	Метод	Экстремальные значения	Диапазон изменения ТБ1, ТБ2
1	ММШ	4,57 – 4,64	37 ≤ ТБ1 ≤ 40 55 ≤ ТБ1 ≤ 58
2	МНМ (вар. 1)	4,57 – 4,64	34 ≤ ТБ1 ≤ 35 57 ≤ ТБ1 ≤ 58
3	МНМ (вар. 2)	4,74 – 4,78	36 ≤ ТБ1 ≤ 40 58 ≤ ТБ1 ≤ 60
4	Все методы	4,57 – 4,78	ТБ1=35, ТБ2=58

В качестве контрольного метода был выбран метод ММШ. В результате была получена табл. 3 для точностей вышеописанных трех методов.

В каждой клетке три значения точности для контрольного метода и вариантов 1 и 2 метода МНМ соответственно. Функции шкалирования изменялись в соответствии со значениями параметров ТБ1 и ТБ2 (см. выше). Диапазон изменения параметров ТБ1 и ТБ2 был выбран так, чтобы оптимальное значение [4] этих параметров для метода ММШ оказалось в его середине. Выделенные в таблице значения претендуют на экстремальные значения метода (см. табл. 4).

Комментарий к табл. 3, 4:

1) Результаты приведенных выше таблиц соответствуют результатам, полученным в работе [4].

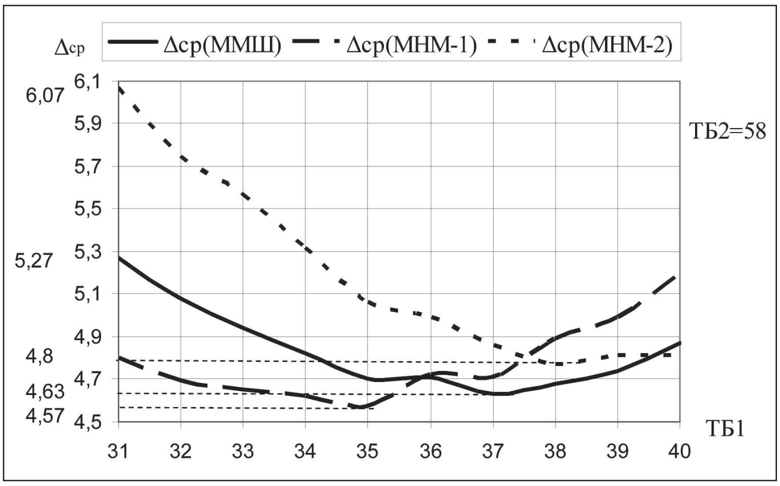

Рис. 3. Зависимость точности основных методов оценки знаний от ТБ1

2) Метод ММШ, который в настоящее время используется при проведении ЕГЭ, и метод МНМ, который исследуется в статье, обладают практически одинаковой точностью.

Для определения степени влияния шкалирования были сформированы три графика (рис. 3), выражающие зависимость точности методов от ТБ1 (ТБ2=58).

Анализируя графики рис. 3, можно заметить, что:

1) выбор функции шкалирования заметно влияет на точность таких методов, как ММШ, МНМ (подбирая параметр ТБ1 должным образом, точность метода ММШ можно увеличить на 0,6 %, метода МНМ (вар.1) – на 0,45 %, метода МНМ (вар.2) – на 1,3 %);

2) все три зависимости имеют точки экстремума (для ММШ – ТБ1=37, для МНМ (вар. 1) – ТБ1=35, для МНМ (вар. 2) – ТБ1=38);

3) для малых значений ТБ1 (ТБ1<36) наиболее эффективным методом является метод МНМ (вар. 1), для больших значений ТБ1 (ТБ1>36) – метод ММШ или метод МНМ (вар. 2) (ТБ1=40).

Аналогичный рисунок можно получить для изучения зависимости точности приведенных выше методов от ТБ2 при ТБ1=37 (рис. 4).

Анализируя графики рис. 4, можно заметить, что:

1) выбор функции шкалирования заметно влияет на точность таких методов, как ММШ, МНМ (подбирая параметр ТБ2 должным образом, точность метода ММШ можно увеличить на 0,8%,

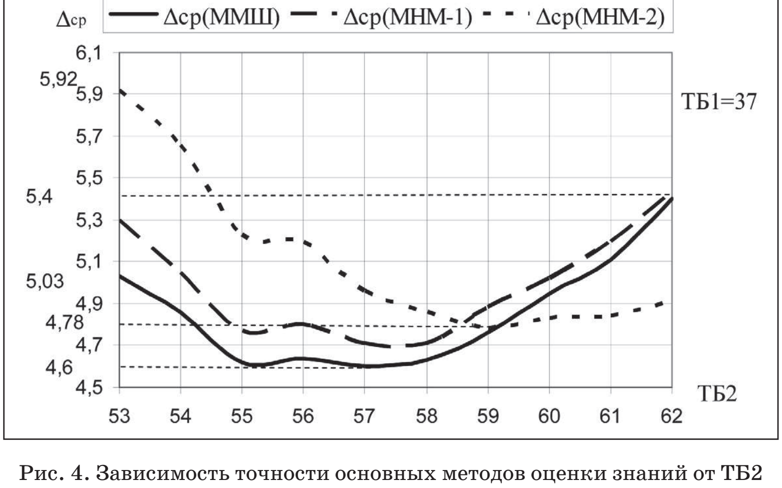

Рис. 4. Зависимость точности основных методов оценки знаний от ТБ2

метода МНМ (вар.1) – на 0,7%, метода МНМ (вар. 2) – на 1,2%);

2) все три зависимости имеют точки экстремума (для ММШ – ТБ2=57, для МНМ (вар.1) – ТБ2=57,58, для МНМ (вар. 2) – ТБ2=59);

3) для меньших значений ТБ2 (ТБ2<59) наиболее эффективным методом является метод ММШ, для больших значений ТБ2 (ТБ2>59) – метод МНМ (вар. 2).

Основные выводы статьи

1. Использование в тестировании метода нечетких множеств позволяет повысить точность классического метода шкалирования, применяемого в настоящее время в ЕГЭ, в среднем на 2,3 %.

2. Метод нечетких множеств в точности уступает методу логарифма Раша (МЛР) в среднем на 0,15 % (см. табл. 1).

3. Выявлена сильная зависимость точности выставляемых оценок от выбора функции шкалирования: удачный выбор функции позволяет повысить точность рассмотренных методов на 2,5–3 %.

4. Наиболее удачным из рассмотренных в этой статье методов является метод логарифма Раша (МЛР), который не зависит от выбора функции шкалирования.

ЛИТЕРАТУРА

1. *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. – Copengagen Denmark: Danish Institute for Educational Research, 1968.

2. *Нейман Ю.М., Хлебников В.А.* Введение в теорию моделирования и параметризации педагогических тестов. – М., 2000. – 169 с.

3. *Полещук О.М.* Методы предварительной обработки нечеткой экспертной информации на этапе ее формализации // Вестник Московского государственного университета леса – Лесной вестник. – 2003. – № 5. – С. 160–167.

4. *Карнаухов В.М.* Модель Раша как игровая модель // Открытое и дистанционное образование. – Томск, 2014. – № 4 (56). – С. 69–76.

5. *Карнаухов В.М.* Точность оценок ЕГЭ для различных методик // Открытое и дистанционное образование. – Томск, 2015. – № 2(58). – С. 20–27.

Karnaukhov V.M.
Russian state agrarian University,
Moscow, Russia

CORRECTION OF PRIMARY POINTS VIA FUZZY SETS USING

Keywords: Rasch's model, Monte-Carlo method, function scaling, the method of primary points, latent parameters, the level of preparedness, fuzzy sets.

In the course of the last 20 years the theory of fuzzy sets is being actively developed. The results of this theory are widely used in testing. This paper investigates the accuracy of the method for estimation of student level of preparedness, based on the use of fuzzy sets. For comparison two methods have been chosen, which are usually used in the exam: scaling method and the method of the Rasch's logarithm. In particular, the article shows the decisive role of the choice of the scaling function, affecting the accuracy of estimates.

The method of fuzzy sets is an adaptive algorithm. Thus, the test results in the form of primary points are used for construction of membership functions of fuzzy sets. We describe the steps of the algorithm, which makes it possible to «tweak» the primary points.

1) The relative frequencies pij appearance points $B=j$ by the decision of the i-th task are calculating,

where $i=1, \ldots, M$ (M is the number of test's tasks), j=0,..., max.

2) The membership functions are forming so that the areas of their a curvilinear trapezoid equals to p_{ij}.

3) The numbers of E_k, $k = 0, \ldots,$ max, are calculating for each membership function. These numbers are the result of diffusivities for fuzzy sets by the method of severity's center.

4) Adjustment of the primary point B for the i-th task is made by the formula:

$$B_{кор} = max \cdot \sum_{k=0}^{max} E_k \cdot \mu_k\left(\frac{B}{max}\right).$$

5) The sum of all «adjusted» points calculates:

$$ПБ_{кор} = \sum_{i=1}^{M} B_{кор}^{i}.$$

6) Test points $ТБ_{кор}$ is calculated by using of scaling.

The main results of the article are as follows:

1) The method of fuzzy sets in testing enables to improve the accuracy of the classical scaling method, which is currently used in the exam, on average by 2.3 %.

2) However, the method of fuzzy sets in accuracy is worse than the method of Rasch's logarithm, on average by 0.15 %.

3) Strong dependence of the accuracy of methods on the selection of scaling function is determined: a good choice of function improves the accuracy of the considered methods in 2.5-3 %.

4) Method of Rasch's logarithm is the most successful method, which does not depend on the choice of the scaling function.

REFERENCES

1. *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. – Copengagen Denmark: Danish Institute for Educational Research, 1968.

2. *Nejman Ju.M., Hlebnikov V.A.* Vvedenie v teoriju modelirovanija i parametrizacii pedagogicheskih testov. – M., 2000. – 169 s.

3. *Poleshhuk O.M.* Metody predvaritel'noj obrabotki nechetkoj jekspertnoj informacii na jetape ee formalizacii // Vestnik Moskovskogo gosudarstvennogo universiteta lesa – Lesnoj vestnik. – 2003. – № 5. – S. 160–167.

4. *Karnauhov V.M.* Model' Rasha kak igrovaja model' // Otkrytoe i distancionnoe obrazovanie. – Tomsk, 2014. – № 4 (56). – S.69–76.

5. *Karnauhov V.M.* Tochnost' ocenok EGJe dlja razlichnyh metodik // Otkrytoe i distancionnoe obrazovanie. – Tomsk, 2015. – № 2(58). – S. 20–27.