

АВТОМАТИЗИРОВАННЫЕ ИНФОРМАТИЗАЦИОННЫЕ СИСТЕМЫ В ОБРАЗОВАНИИ И НАУКЕ

УДК 51-7

В.М. Карнаухов

Московский государственный университет природообустройства, Москва, Россия

МЕТОДЫ ОЦЕНКИ ВЛИЯНИЯ КОЛИЧЕСТВА И ТРУДНОСТИ ЗАДАНИЙ ГРУППЫ В НА ТОЧНОСТЬ РЕЗУЛЬТАТОВ ЕГЭ

Представлены три метода получения оценок уровней подготовленности абитуриентов на ЕГЭ. Первый из них используется в настоящее время на ЕГЭ, второй был предложен автором в своих предыдущих работах, третий – метод первичных баллов – достаточно известен в теории тестирования и был обоснован автором в своих работах. Автор исследует зависимость точности вышеупомянутых методов от количества и трудности заданий теста ЕГЭ группы В и на основании полученных графиков и таблиц дает конкретные рекомендации с целью повышения точности оценок, выставляемых на ЕГЭ.

Ключевые слова: модель Раппа, метод Монте-Карло, функция шкалирования, метод шкалирования, метод первичных баллов, латентные параметры, уровень подготовленности, уровень трудности.

В настоящее время тест, используемый на ЕГЭ, включает задания групп В и С. Задания группы А с выбором правильного варианта ответа, как известно, были исключены из теста по причине низкого уровня трудности несколько лет назад. Задания группы В и С обладают числовым ответом, который угадать практически невозможно. Отличие заданий этих групп состоит в том, что уровень трудности заданий группы В ниже уровня трудности заданий группы С. Поэтому проверка решений заданий группы В осуществляется по конечным числовым ответам автоматизированным образом (на компьютере). Для решения заданий группы С конечного числового ответа недостаточно, для них требуется подробное текстовое описание хода решения. Такие задания проверяются обученными для этих целей экспертами.

За последние два года произошли изменения в структуре теста ЕГЭ. В 2014 г. увеличилось количество заданий группы В с 14 до 15. В 2015 г. увеличилось количество заданий группы С с 6 до 7. В связи с этим возникает естественный вопрос: как повлияет изменение структуры теста на точность выставляемых оценок? В этой статье автор приводит результаты исследования зависимости точности оценок от количества заданий группы В.

Описаны три основных метода получения оценок уровней подготовленности абитуриентов на едином государственном экзамене. Среди них два прямых метода: метод шкалирования, используемый в настоящее время на ЕГЭ, и модифицированный метод шкалирования, предложенный

автором в работе [4]. Эти методы позволяют переводить первичные баллы, набранные абитуриентами, на шкалу процентных логитов, характеризующих уровни подготовленности абитуриентов. Третий метод, изложенный в данной статье, обоснован автором в работе [3] – метод первичных баллов. Этот метод является косвенным методом, позволяющим за два шага получать оценки уровней подготовленности абитуриентов. Первый шаг состоит в получении оценок латентных параметров уровней подготовленности, измеряемых в логитах. На втором шаге логиты переводятся в процентные логиты.

После описания методов получения оценок уровней подготовленности абитуриентов следует изложение основных этапов имитационного моделирования процесса тестирования. Моделирование использует известную в теории математическую модель тестирования датского математика Г. Раппа [1, 2]. Обсуждаются некоторые элементы компьютерной программы, осуществляющей имитационное моделирование тестирования.

Результатом работы вышеупомянутой программы являются графики и таблицы, которые выявляют характер зависимости точности трех методов от числа и трудности заданий группы В теста ЕГЭ. На основании полученного материала автор делает конкретные выводы и дает рекомендации, направленные на повышение точности оценок ЕГЭ.

Метод 1 – метод шкалирования. В 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. Этот метод реализуется в 3 этапа.

I этап.

Сначала в диапазоне первичных баллов от нуля до максимального первичного балла ПБ_{max} (ПБ – первичный балл) 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, овладении комплексными умениями, способности выполнять творческие задания по соответствующему общеобразовательному предмету.

Если спецификация экзаменационного варианта не изменилась по сравнению с прошлым годом, то ПБ1 и ПБ2 также остаются неизменными. Если же структура экзаменационной работы или сложность заданий контрольных измерительных материалов поменялись, то устанавливаются новые значения ПБ1 и ПБ2 с учетом имеющихся изменений.

II этап.

Первичным баллам ПБ1 и ПБ2 ставятся в соответствие тестовые баллы ТБ1 и ТБ2 по каждому общеобразовательному предмету.

Для всех предметов в качестве величин ТБ1 выбираются минимальные тестовые баллы ЕГЭ 2013 г., установленные распоряжениями Рособнадзора. Данные значения совпадают с минимальными баллами ЕГЭ 2012 г.

Тестовые баллы ТБ2 по всем предметам, кроме географии и истории, устанавливаются равными аналогичным баллам 2012 г. По сравнению с 2012 г. на 1 балл уменьшился ПБ2 по географии и на 1 балл увеличился ПБ2 по истории. Это связано с изменением структуры экзаменационных работ по этим предметам. В табл. 1 представлены значения ПБ1 и ПБ2, ТБ1 и ТБ2 на 2013 г.

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 баллом и тестовым баллом на основе следующей процедуры. Первичному баллу 0 ставится в соответствие тестовый балл 0, а максимальному первичному баллу ПБ_{max} ставится в соответствие тестовый балл 100. Все промежуточные первичные баллы между 0, ПБ1, ПБ2 и ПБ_{max} переводятся в тестовые, пропорционально распределенные между соответствующими значениями тестовых баллов: 0, ТБ1, ТБ2 и 100. На рис. 1 представлена получаемая зависимость.

Если промежуточные первичные баллы соответствуют дробным значениям тестовых, то производится округление тестового балла до ближайшего большего целого числа. Указанная процедура позволяет согласовывать тестовые баллы одинаково подготовленных участников

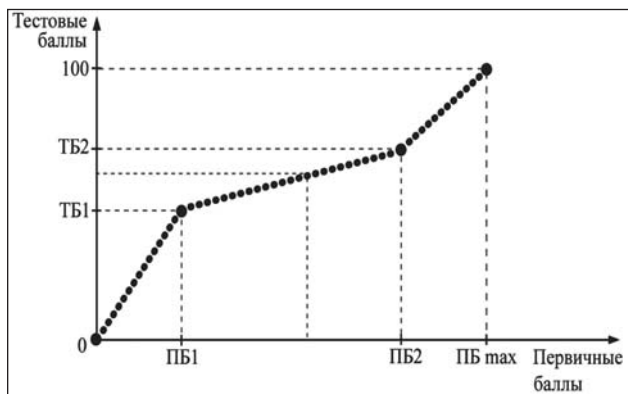


Рис. 1. Соответствие между тестовыми и первичными баллами

2011–2013 гг. и обеспечивает сравнительную сопоставимость результатов экзамена по годам.

Метод 2 – метод модифицированного шкалирования. Описанный выше метод шкалирования можно усовершенствовать [4]. Например, можно рассмотреть семейство функций перевода первичных баллов в тестовые, которые отличаются между собой только значениями в точках ПБ1 = 5 и ПБ2 = 15. Исследование этого семейства приводит нас к наиболее эффективной функции зависимости уровня подготовленности от первичного балла. Ломаная линия зависимости изображена на рис. 2.

Для полученной линии ТБ1 = 36, ТБ2 = 58, тогда как для прежней, используемой на ЕГЭ, ТБ1 = 24, ТБ2 = 63. Используя на практике полученные значения, можно добиться выигрыша в точности примерно в 2,3 %.

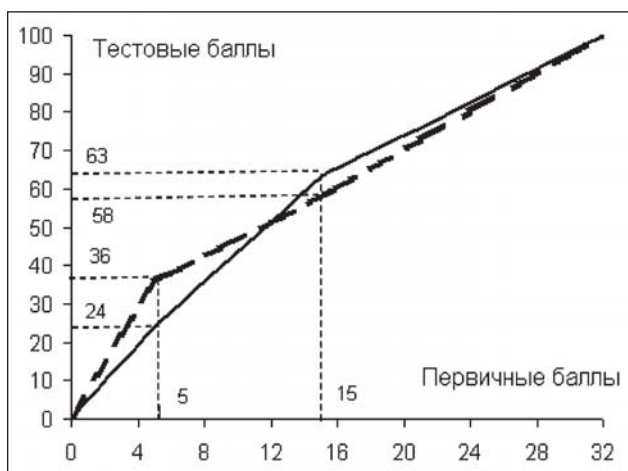


Рис. 2. Изменение соответствия между первичными и тестовыми баллами

Метод 3 – метод первичных баллов. Третьим методом получения оценок уровней подготовленности абитуриентов является метод первичных баллов. Этот метод описан и обоснован в авторской работе [3]. Согласно этому методу оценки $\theta_i, i=0, \dots, K$, уровней подготовленности участников тестирования, где I – число набранных тестовых баллов на экзамене, K – максимально возможное количество набранных баллов, вычисляются по формулам

$$\theta_i \approx \bar{\theta}_i + \bar{\theta}_{cp}, i = 1, \dots, K-1.$$

В этих формулах используются следующие величины:

$$1) \bar{\theta}_i = \ln \left(\frac{i}{K-i} K_1 \right), \text{ причем } K_1 = \frac{\sum_{r=1}^{K-1} (K-r) \cdot N_r}{\sum_{r=1}^{K-1} r \cdot N_r},$$

где N_r – число участников тестирования, набравших r тестовых баллов.

$$2) \bar{\theta}_{cp} = \frac{\sum_{i=1}^{K-1} \bar{\theta}_i \cdot N_i}{N - N_0 - N_K}.$$

Для крайних значений $i: i=0$ и $i=K$ используются следующие оценки:

$$\theta_0 = \theta_{\max}, \theta_K = \theta_{\max}, \text{ где } \theta_{\max} = 5.$$

Точность вышеперечисленных трех методов представлена в табл. 2 (см. [4]).

Таблица 2

Точность различных методов оценки латентных параметров тестирования

№	Метод	Средняя погрешность, %	Максимальная погрешность, %
1	Прямой метод шкалирования	6,9	27
2	Прямой метод модифицированного шкалирования	4,6	22
3	Косвенный метод первичных баллов	4,5	19

Как видно из табл. 2, метод первичных баллов является наиболее точным, что будет подтверждено ниже.

Имитационное моделирование. Для исследования зависимости точности вышеперечисленных методов от количества и трудности заданий группы В была разработана авторская программа, моделирующая при помощи метода Монте-Карло процесс проведения ЕГЭ для абитуриентов в коли-

честве $N = 500$ и теста, состоящего из n_B заданий группы В и n_C заданий группы С. При этом число n_C фиксировано и равно 6, что обычно и было на последних ЕГЭ, а число n_B меняется в пределах от 10 до 20. Благодаря вариативности количества заданий группы В и их трудности можно было провести запланированное исследование.

Процесс ЕГЭ моделировался достаточно большое количество раз (число итераций равно 20). Для каждого моделирования вычислялись две характеристики:

1) среднее отклонение σ_{cp} оценки уровня подготовленности абитуриента от истинного значения этого латентного параметра;

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

Далее вычисленные характеристики усреднялись по всем итерациям.

Для моделирования процесса тестирования использовался метод Монте-Карло. Опишем процесс компьютерной имитации процесса тестирования.

1) Вначале моделируются истинные уровни подготовленности участников $\theta_i, i=1, \dots, N$ и истинные уровни трудностей заданий $\delta_j, j=1, \dots, M$. Уровни подготовленности участников смоделированы как реализации нормальной случайной величины $N(0,1)$ по формуле

$\theta_i = F_N^{-1}(r_i)$, где $F_N(x)$ – функция распределения нормированной нормальной случайной величины, т.е. $N(0,1)$, которая определяется по формуле

$$F_N(x) = 0.5 + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx,$$

$F_N^{-1}(r_i)$ – 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:

$$\left(\Delta = \frac{0.1}{3}\right): \delta_j \in N(\delta_j^{cp}; \Delta), j=1, \dots, M,$$

$$-3 \leq \delta_j^{cp} \leq 0, j=1, \dots, n_B,$$

$$\delta_j^{cp} = 1, j = n_B+1, n_B+2,$$

$$\delta_j^{cp} = 2, j = n_B+3, n_B+4,$$

$$\delta_j^{cp} = 3, j = n_B+5, n_B+6.$$

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

2) Для каждого абитуриента и для каждого задания вычисляются первичные баллы. Для этого по формуле

$$p_{ij} = \frac{1}{1 + e^{-(\theta_i - \delta_j)}}, i=1, \dots, N, j=1, \dots, M$$

вычисляются вероятности p решения i -м абитуриентом j -го задания. Затем абитуриенту начисляется первичный балл В за решение задания по формуле

$$B = \begin{cases} 0, & r \geq p \\ \left\lceil \frac{r \cdot m}{p} \right\rceil + 1, & r < p, \end{cases}$$

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

m – максимальное число баллов за решение задачи, причем

$$m = 1 \text{ для } j = 1, \dots, n_B,$$

$$m = 2 \text{ для } j = n_B+1, n_B+2,$$

$$m = 3 \text{ для } j = n_B+3, n_B+4,$$

$$m = 4 \text{ для } j = n_B+5, n_B+6,$$

$$[x] - \text{целая часть числа } x.$$

Результаты исследования зависимости погрешности методов от количества заданий группы В. Для заданий группы В в вышеописанной программе был установлен уровень трудности $\delta_B = -1$. Это позволило получить зависимость (рис. 3) погрешности трех вышеописанных методов от количества n_B заданий группы В ($n_B \in [10, 20]$).

Комментарии к рис. 3:

1) Погрешность для метода 3 с ростом n_B уменьшается. Этот метод ведет себя естественным образом: с увеличением количества информации точность увеличивается.

2) Методы 1 и 2 ведут себя неестественно в смысле, описанном выше. Погрешность падает до своего минимума при $n_B = 14$. Напомним, что это количество заданий предлагалось до 2014 г. В 2014 г. было предложено 15 заданий группы В.

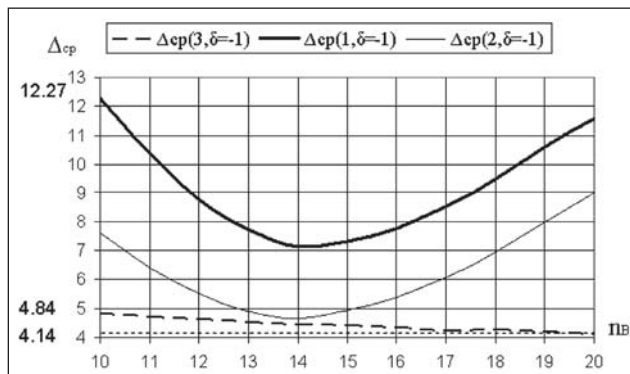


Рис. 3. Зависимость погрешности 3 методов от количества заданий группы В при $\delta_B = -1$

Как видно из графиков, погрешность при этом увеличилась по сравнению со случаем $n_B = 14$.

3) Увеличение n_B приведет к росту погрешности. Поэтому дальнейшее увеличение n_B не рекомендуется.

4) Погрешность метода 1 при $n_B = 14$ практически в 1,5 раза больше погрешностей методов 2 и 3. Поэтому рекомендуется заменить метод 1 на методы 2 и 3 с целью повышения точности выставления оценок.

Результаты исследования зависимости погрешности методов от трудности заданий группы В. Для исследования было установлено количество заданий группы В, равное 15. В результате имитационного моделирования был получен график, изображенный на рис. 4.

Комментарии к рис. 4:

1) Все три графика имеют точки минимума, лежащие в пределах от $-1,2$ до $-0,6$.

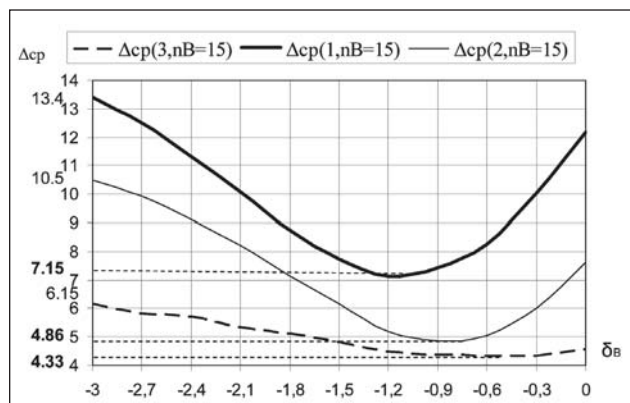


Рис. 4. Зависимости погрешности 3 методов от уровня трудности заданий группы В при $n_B = 15$

2) Предыдущее исследование проводилось практически для точки минимума по уровню трудности заданий группы В.

3) Для составления части В теста рекомендуется подбирать задачи из различных тем математики с уровнем трудности, лежащим в пределах от $-1,2$ до $-0,6$. При этом для $n_B = 15$ будет наименьшая погрешность при выставлении оценок.

4) Из этих графиков видно, что наилучшим методом является метод 3, наихудшим – метод 1. Однако небольшая модификация метода 1, выраженная в методе 2, приведет к существенному (в 1,5 раза) увеличению точности метода.

5) Вариация погрешности метода 1 значительно больше вариации погрешности метода 3. Это обстоятельство обязывает составителей тестов не выходить за некоторые пределы уровня трудности заданий группы В, выраженные интервалом $(-1,5; -0,6)$. В противном случае погрешность выставляемой оценки может увеличиться почти в два раза.

Результаты исследования зависимости погрешности методов от количества и трудности заданий группы В. Все предыдущие исследования были проведены для случаев, когда параметры n_B и δ_B изменялись в пределах прямых вида $n_B = \text{const}$ и $\delta_B = \text{const}$. Ниже предлагаются результаты исследования (табл. 3–5) зависимости погрешности трех методов от параметров n_B и δ_B в случае, когда эти параметры изменяются в пределах плоской области: $10 \leq n_B \leq 20$, $-2 \leq \delta_B \leq 0$.

Комментарии к табл. 3–5:

1) В каждой таблице тенью выделены экстремальные области, т.е. те области в пространстве переменных (n_B, δ_B) , в которых погрешность минимальна.

2) Экстремальная область для метода 3 практически не зависит от δ_B и находится в диапазоне $(-1; -0,2)$, причем с ростом n_B погрешность падает, что подтверждается также графиком рис. 3.

3) Экстремальные области для метода 1 существенно зависят от δ_B , причем наименьшие погрешности получаются для области, соответствующей $n_B = 12$.

4) Экстремальные области для метода 2 существенно зависят от δ_B , причем наименьшие погрешности получаются для области, соответствующей $n_B = 14$.

Таблица 3

Зависимость погрешности метода шкалирования от количества и трудности задач группы В

δ_B / n_B	10	11	12	13	14	15	16	17	18	19	20
-2	7,54	6,50	6,31	6,79	7,81	9,61	11,71	13,49	15,10	16,60	18,02
-1,9	7,76	6,78	6,25	6,56	7,52	9,12	10,96	13,01	14,63	16,01	17,39
-1,8	7,97	6,93	6,41	6,43	7,24	8,73	10,59	12,29	14,01	15,55	16,80
-1,7	8,28	7,10	6,32	6,53	7,07	8,29	10,08	11,66	13,50	14,81	16,18
-1,6	8,70	7,48	6,61	6,48	6,94	7,94	9,53	11,16	12,76	14,24	15,44
-1,5	8,98	7,74	6,81	6,47	6,80	7,70	9,14	10,65	12,21	13,54	14,96
-1,4	9,41	8,04	7,05	6,77	6,69	7,48	8,77	10,15	11,53	12,90	14,37
-1,3	9,83	8,46	7,30	6,90	6,68	7,33	8,39	9,63	10,00	12,30	13,71
-1,2	10,36	8,78	7,68	7,12	6,82	7,12	8,08	9,07	10,47	11,59	12,83
-1,1	10,89	9,33	8,24	7,45	6,89	7,15	7,85	8,76	9,91	11,02	12,20
-1,0	11,20	9,79	8,65	7,89	7,16	7,34	7,82	8,59	9,56	10,53	11,61
-0,9	11,84	10,36	9,07	8,21	7,63	7,39	7,75	8,37	9,16	10,11	10,94
-0,8	12,43	10,98	9,65	8,65	7,96	7,48	7,82	8,27	8,82	9,61	10,60
-0,7	13,09	11,54	10,20	9,12	8,44	7,95	7,93	8,21	8,60	9,21	10,12
-0,6	13,61	12,29	10,75	9,81	8,91	8,43	8,07	8,16	8,57	8,92	9,69
-0,5	14,19	12,69	11,43	10,39	9,55	8,72	8,56	8,43	8,59	8,89	9,42
-0,4	14,73	13,55	12,18	11,16	10,10	9,32	8,84	8,61	8,69	8,90	9,15
-0,3	15,30	14,28	12,91	11,81	10,81	9,99	9,35	8,95	8,93	8,95	9,28
-0,2	16,22	14,73	13,65	12,58	11,57	10,64	10,10	9,45	9,23	9,19	9,21
-0,1	16,77	15,45	14,29	13,26	12,35	11,50	10,61	9,98	9,66	9,51	9,35
0	17,45	16,25	15,02	14,02	13,11	12,19	11,44	10,65	10,06	9,82	9,85

Таблица 4

Зависимость погрешности метода модифицированного шкалирования от количества и трудности задач группы В

δ_B / n_B	10	11	12	13	14	15	16	17	18	19	20
-2	5,32	5,13	5,37	5,83	6,59	7,85	9,29	10,67	12,09	13,52	14,78
-1,9	5,42	5,16	5,18	5,56	6,31	7,47	8,70	10,27	11,70	13,01	14,21
-1,8	5,43	5,07	5,15	5,40	6,05	7,15	8,33	9,72	11,20	12,67	13,68
-1,7	5,46	5,03	4,89	5,29	5,75	6,78	7,95	9,22	10,78	12,02	13,18
-1,6	5,60	5,10	4,90	5,05	5,54	6,39	7,54	8,84	10,17	11,60	12,60
-1,5	5,67	5,15	4,89	4,89	5,29	6,07	7,18	8,38	9,68	10,98	12,16
-1,4	5,91	5,16	4,85	4,85	5,05	5,83	6,76	7,89	9,10	10,40	11,65
-1,3	6,11	5,36	4,89	4,83	4,89	5,56	6,34	7,30	8,55	9,86	11,09
-1,2	6,43	5,49	5,00	4,84	4,76	5,19	5,96	6,79	8,09	9,20	10,26
-1,1	6,70	5,74	5,23	4,83	4,71	5,06	5,60	6,41	7,48	8,63	9,59
-1,0	6,97	6,06	5,38	4,94	4,62	4,99	5,42	6,12	7,08	8,08	9,04
-0,9	7,25	6,33	5,62	5,15	4,80	4,80	5,12	5,76	6,59	7,60	8,38
-0,8	7,70	6,79	5,92	5,30	4,87	4,72	5,04	5,48	6,09	7,05	7,93
-0,7	8,21	7,19	6,28	5,56	5,15	4,90	4,87	5,29	5,76	6,46	7,28
-0,6	8,64	7,71	6,69	5,95	5,39	5,13	4,93	5,06	5,54	5,99	6,74
-0,5	9,16	8,02	7,17	6,37	5,77	5,27	5,12	5,05	5,33	5,81	6,32
-0,4	9,72	8,71	7,72	6,92	6,11	5,57	5,26	5,12	5,28	5,58	5,96
-0,3	10,04	9,23	8,24	7,36	6,58	6,07	5,57	5,25	5,25	5,48	5,78
-0,2	10,83	9,77	8,85	7,98	7,22	6,47	5,95	5,51	5,44	5,45	5,56
-0,1	11,33	10,27	9,37	8,60	7,80	7,10	6,35	5,86	5,63	5,57	5,48
0	12,02	10,94	10,03	9,16	8,46	7,61	6,97	6,32	5,92	5,73	5,76

Таблица 5

Зависимость погрешности метода первичных баллов от количества и трудности задач группы В.

δ_B / n_B	10	11	12	13	14	15	16	17	18	19	20
-2	5,16	5,22	5,33	5,32	5,34	5,26	5,28	5,30	5,21	5,17	5,22
-1,9	5,19	5,16	5,16	5,22	5,25	5,24	5,18	5,06	5,12	5,13	5,05
-1,8	5,15	5,04	5,14	5,16	5,16	5,06	5,08	5,03	5,01	4,96	4,94
-1,7	5,00	5,00	4,97	4,99	4,99	4,99	4,97	4,90	4,90	4,78	4,82
-1,6	4,99	4,96	4,91	4,90	4,95	4,91	4,85	4,83	4,79	4,77	4,77
-1,5	4,88	4,89	4,84	4,81	4,88	4,73	4,74	4,73	4,74	4,67	4,59
-1,4	4,92	4,70	4,73	4,70	4,78	4,74	4,58	4,60	4,63	4,56	4,47
-1,3	4,83	4,77	4,70	4,65	4,61	4,62	4,59	4,50	4,51	4,44	4,39
-1,2	4,83	4,71	4,66	4,70	4,58	4,57	4,48	4,42	4,35	4,32	4,31
-1,1	4,80	4,60	4,65	4,55	4,58	4,50	4,42	4,33	4,29	4,26	4,14
-1,0	4,78	4,69	4,58	4,48	4,41	4,42	4,28	4,29	4,23	4,16	4,09
-0,9	4,71	4,55	4,53	4,54	4,39	4,36	4,26	4,18	4,09	4,11	4,07
-0,8	4,78	4,69	4,52	4,43	4,35	4,25	4,22	4,15	4,16	4,12	3,96
-0,7	4,78	4,71	4,58	4,45	4,41	4,30	4,18	4,19	4,08	3,98	3,95
-0,6	4,82	4,73	4,66	4,49	4,37	4,34	4,28	4,13	4,11	3,96	3,91
-0,5	4,90	4,69	4,59	4,54	4,39	4,41	4,24	4,22	4,07	3,96	3,91
-0,4	5,10	4,81	4,70	4,58	4,43	4,34	4,23	4,19	4,07	3,94	3,96
-0,3	4,90	4,83	4,76	4,58	4,41	4,44	4,33	4,28	4,09	4,04	3,97
-0,2	5,12	5,04	4,80	4,66	4,57	4,47	4,33	4,21	4,20	4,08	4,06
-0,1	5,16	4,98	4,72	4,81	4,63	4,55	4,37	4,36	4,32	4,16	4,11
0	5,17	5,03	4,92	4,80	4,76	4,51	4,47	4,46	4,37	4,26	4,18

Выводы-рекомендации:

1) Для выставления оценок на ЕГЭ рекомендуется метод первичных баллов, точность которого в 1,5–2 раза выше, чем у метода шкалирования, который используется в настоящее время на экзамене.

2) При составлении теста ЕГЭ, на котором применяется метод шкалирования, рекомендуется использовать 12 заданий группы В, трудность которых определяется величиной –2 (в логитах); если применяется метод первичных баллов, то рекомендуется использовать наибольшее допустимое число заданий с уровнем трудности, лежащим в интервале (0,4; 0,8) (в логитах).

3) Если использовать 15 заданий группы В, то рекомендуется установить следующие уровни трудности (в логитах) для заданий группы В:

для метода шкалирования – (–1,5; –0,6);

для модифицированного метода шкалирования – (–1,5; –0,3);

для метода первичных баллов – (–1,8; 0).

ЛИТЕРАТУРА

1. *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. – Copenhagen, 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. *Карнаухов В.М.* Модель Раша как игровая модель // Открытое и дистанционное образование. – Томск, 2014. – № 4 (56). – С. 69–76.

4. *Карнаухов В.М.* Исследование точности оценок ЕГЭ // Информатизация образования и науки. – 2015. – № 1 (25). Янв. – С. 116–127.

Karnaukhov V.M.

Moscow State University of Environmental Engineering, Moscow, Russia

EVALUATION METHODS OF QUANTITY INFLUENCE AND TASK COMPLEXITY OF B-GROUP ON THE RESULT ACCURACY OF UNIFIED STATE EXAMINATION

Keywords: Rasch's model, Monte-Carlo method, scaling method, method of primary points latent parameters, qualification level, level of complexity.

Today the test being used in the unified state examination includes tasks of B and C groups. The tasks of A group involving the choice of a correct answer version were excluded from the test few years ago because of low level of complexity. Tasks

of B and C groups have numerical answers which are almost impossible to guess. The difference between these groups is that the level of complexity of B group is lower than that of C group. Therefore, the check of B group results is carried out by means of the final numerical answers electronically (via the computer). To solve the task of C group the final numerical answer is not enough, it requires detailed text description of the solution. These tasks are checked by experts trained for this purpose.

The article describes the three main methods of estimation of qualification levels of students passing the Unified State Examination. There are two direct methods: a method of scaling which is used in the exam nowadays and a modified scaling method proposed by the author in his previous works. These methods allow us to convert a primary score of students to a scale of percentage logits which characterize a qualification level of the participant. The third method investigated in this work was proved by the author in previous works, the method of initial points. This method is an indirect method that makes it possible to calculate the evaluation of participants' qualification levels in two steps. The first step consists in obtaining assessments of latent parameters of qualification levels measured in logits. The second step consists in converting logits to percentage logits.

After describing the methods of estimates obtaining of qualification levels of testing participants the article highlights the main stages of simulation of testing process. The simulation uses the well-known theory of G. Rasch, a mathematical model of testing. It discusses some elements of a computer program performing a simulation of testing.

The result of the work of this program is diagrams and tables that show the character of dependence of accuracy of the three methods on the number and complexity of tasks of B group. On the basis of the material obtained the author makes specific conclusions and gives recommendations aimed at improving in accuracy of exam estimations.

1) to estimate the results of the unified state exam, it is recommended to use the method of initial points whose accuracy is 1.5 - 2 times higher than that of the scaling method which is used in the exam nowadays.

2) When developing the examination test in which the scaling method is used, it is recommended to use 12 tasks of B group, the difficulty of which is determined by 2 logits; if the method of initial points is used, it is recommended to use the maximum acceptable number of tasks with a complexity level ranging (0,4; 0,8).

3) If 15 tasks of B group are used, it is recommended to set the following levels of difficulty for these tasks: for scaling method - (-1,5; -0,6), for the method of initial points (-1,8; 0), for the modified scaling method - (-1,5; -0,3).

REFERENCES

1. *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. – Copenhagen, Denmark: Danish Institute for Educational Research, 1968.
2. *Nejman Ju.M., Hlebnikov V.A.* Vvedenie v teoriju modelirovaniya i parametrizacii pedagogicheskikh testov. – M., 2000. – 169 s.
3. *Karnauhov V.M.* Model' Rasha kak igrovaja model' // Otkrytoe i distancionnoe obrazovanie. – Tomsk, 2014. – № 4 (56). – S. 69–76.
4. *Karnauhov V.M.* Issledovanie tochnosti ocenok EGJe // Informatizacija obrazovanija i nauki. – 2015. – № 1 (25). Janv. – S. 116–127.