

УДК 004.62:519.2

Раков С.А., Мазорчук М.С., Бондаренко Е.О.

Национальный аэрокосмический университет им. Н.Е. Жуковского
«Харьковский авиационный институт»

АЛГОРИТМ КОРРЕКТИРОВКИ ТЕСТОВЫХ БАЛЛОВ НА ОСНОВЕ АНАЛИЗА СЛОЖНОСТИ ЗАДАНИЙ

Целью работы является разработка алгоритма корректировки тестовых баллов на основе анализа сложности заданий тестов. В работе рассмотрены тесты с закрытой формой вопроса. Корректировка баллов теста необходима для дифференциации участников тестирования, набравших одинаковые тестовые баллы. В работе представлены основные этапы алгоритма корректировки результатов тестирования и приведены расчеты, показывающие эффективность применения данного алгоритма. Для проведения расчетов используются пакеты для анализа данных, такие как SPSS и MS Excel.

Ключевые слова: *тесты, сложность заданий теста, корректировка тестовых баллов, линейное преобразование, процентильное преобразование.*

Введение

Методы анализа результатов тестирования базируются как на классической теории тестов [1,2], так и на современной теории IRT (однопараметрическая модель, модель Раша, двух- и трехпараметрические модели Бирнбаума) [3,4]. Тестовые баллы, полученные участниками тестирования в рамках модели IRT являются статистически достаточными характеристиками, что дает возможность качественно интерпретировать результаты тестирования, т.к. модель Раша устанавливает зависимость между уровнем подготовленности испытуемого и сложностью заданий¹.

Традиционно при шкалировании результатов тестирования в рамках классической теории тестов сложность заданий тестов не используется. Одинаковые результаты участников тестирования никак не дифференцируются. Однако, в нормативно-ориентированных тестах, в основе которого лежит сравнение результатов, результирующий балл играет важную роль. Испытуемых, которые решили более сложные задания, целесообразно поощрить за счет корректировки тестового балла, но при этом общий рейтинг участников не должен нарушаться, т.е. участники тестирования, набравшие больше баллов, не должны после корректировки оказаться ниже исходного рейтинга. Корректировка должна осуществляться в пределах одного результирующего рейтингового балла. Поэтому, актуальным является разработка методов, моделей и алгоритмов, позволяющих корректировать исходные баллы на основании трудности заданий в пределах одного тестового балла.

¹ Следует, однако, отметить, что модели IRT относятся к классу жестких моделей, для корректного применения которых как задания теста, так и участники тестирования должны пройти проверку на применимость методов IRT, вследствие чего эти методы непосредственно не применяются для шкалирования результатов тестов высоких ставок, к которым относятся национальные вступительные тестирования. Методы IRT широко используются на этапе подготовки тестов, ведения банков тестовых заданий, для оценки результатов мониторинговых исследований, в частности международных исследований TIMSS, PISA, PIRLS и т.д.

1. Анализ существующих методов повышения дифференциации результатов тестирования

Существуют различные методы, которые позволяют учитывать сложность заданий. Один из них – это метод штрафных баллов¹. Корректировка баллов по данному методу предполагает уточнение баллов на основе анализа вероятности угадывания. Результирующий балл будет рассчитываться по формуле[5]:

$$X_{res} = R - \frac{N}{C-1},$$

где R – количество правильных ответов; N – количество неправильных ответов; C – количество вариантов ответов на задание.

Оценка за выполнение задания с выбором правильного ответа на вопрос выбором правильного ответа из C предложенных вариантов:

$$X_{res} = \begin{cases} 1, & \text{в случае неправильного ответа} \\ -\frac{1}{C-1}, & \text{в случае правильного ответа} \end{cases}$$

Данный метод, к сожалению, нарушает основной принцип шкалирования, поскольку после перевода «сырых» тестовых баллов может измениться рейтинг участников тестирования.

Также известным методом корректировки тестовых баллов является метод, который используется в системе государственного тестирования Белоруссии [6]. Результирующий тестовый балл определяется по формуле:

$$B_i = \frac{\sum_{j=1}^k a_{ij}(1-s_j)}{\sum_{j=1}^k (1-s_j)} \cdot 100, \quad j = 1, \dots, n$$

где $s_j = \frac{c_j}{n}$ – сложность задания; $c_j = \sum_{i=1}^n a_{ij}$ – сумма баллов за j -е задание.

Данный метод также нарушает принцип шкалирования, т.к. участники с более высоким баллом могут оказаться по рейтингу выше участников, набравших меньше тестовых баллов.

Чтобы искусственно повысить по сравнению с классической теорией Раша степень дифференциации участников тестирования, можно учитывать не только количество верно выполненных заданий теста, но и характеристики трудности этих заданий. Для этой же цели

¹ Метод штрафных баллов – устоявшийся в русскоязычной литературе термин, которому в англоязычной литературе соответствует термин *guessing correction formula* – формула поправки на угадывание, который отражает важный дидактический принцип ответов на задания тестов в условиях неполного знания. При использовании этой формулы математическое ожидание оценки участников тестирования, которые просто выбирают ответ случайным образом будет равно 0 (правильный ответ оценивается 1), если же участник тестирования обладает частичным знанием и может отбросить какие-то заведомо неправильные варианты ответов и будет выбирать только из оставшихся правдоподобных вариантов, то математическое ожидание его ответа будет больше нуля и в оценке участников будет учитываться также и неполные знания, участники с частичными знаниями будут получать более высокую оценку, чем те, которые их не имеют. Анализ ответов участников ЗНО показывает, что украинские абитуриенты придерживаются именно такой стратегии, поэтому использование формулы поправки на угадывание было бы оправданным, поскольку в используемой в данный момент схеме преимущество получает полный невежда, выбирающий вариант ответа случайным образом. Введению формулы поправки на угадывание мешает именно ее трактовка как метода штрафных санкций. Таким образом, этот метод по недоразумению называется методом штрафных баллов и его смысл прямо противоположный – поощрение за частичное знание.

нужны и коэффициенты дискриминации заданий теста. В [7] приводится метод для корректировки тестовых баллов, который не нарушает основной принцип шкалирования, т.е. корректировка осуществляется в пределах одного «сырого» тестового балла.

Процедура проведения расчетов по данному методу является довольно трудоемкой, так как требуется пересчитывать тестовые баллы для каждой группы испытуемых, имеющих одинаковый результат. Поэтому, целесообразным является разработка алгоритмического и программного обеспечения для решения задачи корректировки тестовых баллов.

Целью данной работы является разработка алгоритма корректировки тестовых баллов с учетом сложности заданий.

2. Исходные данные

Предлагаемый в статье алгоритм можно использовать для нормативно -ориентированных тестов с заданиями произвольных форматов, как закрытого, так и открытого типов, но для простоты изложения ниже будут рассматриваться только задания закрытого типа.

Исходными данными для алгоритма корректировки тестовых баллов являются следующие характеристики теста:

t_i - суммарный тестовый балл, набранный участником тестирования с номером i ;

m - суммарный тестовый балл, набранный всеми участниками тестирования;

$A = \{a_{ij}\}_{i,j}^{n,k}$ - матрица ответов, где a_{ij} - баллы за каждое задание (баллы, набранные

участником i за задание j);

k - количество заданий в тесте;

n - количество участников тестирования;

p - количество участников, которые набрали m баллов;

c_j - сумма баллов за j -е задание, набранных всеми участниками тестирования;

I_j - максимальный балл, которым оценивается задание j (вес задания)¹;

s_j - сложность задания j ;

$N_m = \{n_1^m, n_2^m, \dots, n_p^m\}$ - множество номеров участников, которые набрали m баллов;

Q_j^m - множество номеров заданий, которые выполнил участник n_j^m ;

d_j^m - сумма сложностей заданий. Рассчитывается по формуле: $d_j^m = \sum_{v \in Q_j^m} s_v$.

3. Алгоритм корректировки результатов тестирования

Основные этапы реализации данного алгоритма следующие:

Этап 1. Рассмотрим матрицу ответов A . Она будет иметь вид, представленный в таблице 1.

Таблица №1

Исходная матрица ответов по тесту

Задания	Задание 1	Задание 2	Задание j	...	Задание k	Тестовый балл i-го участника
Номер участника							
1	2	3	4	5	6	7	8
1	a_{11}	a_{12}	...	a_{1j}	...	a_{1k}	t_1
2	a_{21}	a_{22}	...	a_{2j}	...	a_{2k}	t_2
...

¹Максимальный балл, набранный всеми участниками тестирования за выполнения задания j может быть меньше I_j (в том случае, когда ни один участник не решил задание j полностью).

1	2	3	4	5	6	7	8
i	a_{i1}	a_{i2}	...	a_{ij}	...	a_{ik}	t_i
...
n	a_{n1}	a_{n2}	...	a_{nj}	...	a_{nk}	t_n
Максимальный балл за j-е задание	I_1	I_2	...	I_j	...	I_k	
Сумма баллов за j-е задание	c_1	c_2	...	c_j	...	c_k	
Сложность j-го задания	s_1	s_2	...	s_j	...	s_k	

Этап 2. Определяем максимальный балл за каждое задание: $I_j = \max(a_{ij})$.

Этап 3. Рассчитываем сумму баллов по каждому j-му заданию по формуле:

$$c_j = \sum_{i=1}^n a_{ij} \text{ (таблица 1).}$$

Этап 4. Определяем сложность каждого задания: по формуле: $s_j = \frac{c_j}{n \cdot I_j}$ (таблица 1).

Этап 5. Определяем, для какой группы участников будем выполнять корректировку, т.е. задаем m.

Этап 6. Формируем множество участников, которые набрали m баллов, т.е. для всех $t_i = m$:

$$N_m = \{n_1^m, n_2^m, \dots, n_p^m\}$$

Этап 7. Определяем для всех $t_i = m$ множество выполненных заданий Q_j^m по следующему правилу:

$$\text{Если } a_{ij} \neq 0, \text{ то } Q_j^m := 1, \text{ иначе } Q_j^m := 0.$$

Этап 8. Для всех $t_i = m$ рассчитываем сумму сложностей заданий по следующему правилу:

$$d_j^m = 0;$$

Для $i = 1$ до p

Для $j = 1$ до k

$$\text{Если } Q_j^m := 1, \text{ то } d_j^m = d_j^m + s_j,$$

$$\text{Иначе } d_j^m = d_j^m + 0.$$

Этап 9. Вычисляем поправочные значения P_i^m для каждого участника из группы N_m по следующему алгоритму:

1-й способ: Находим максимальное значение суммы сложностей задания: $d_{\max}^m = \max\{d_i^m\}$.

Находим минимальное значение суммы сложностей задания: $d_{\min}^m = \min\{d_i^m\}$.

Для каждого i-го участника ($i = \overline{1..m}$) рассчитываем поправочное значение P_i^m по следующей формуле:

$$P_i^m = \frac{0,99(d_i^m - d_{\max}^m)}{d_{\min}^m - d_{\max}^m}$$

2-й способ: Поправочные значения рассчитываются как нижние процентильные ранги для сумм сложностей выполненных заданий каждым участником. Как показали вычислительные эксперименты на реальных данных, оба способа дают приблизительно одинаковые результаты. Поэтому предпочтительным следует считать 1-й предложенный способ, который и будет обсуждаться в дальнейшем. Следует также иметь в виду и доступность интерпретации результатов, полученных по первому и второму способу: для первого способа достаточно опубликовать таблицу сложностей заданий и тогда каждый самостоятельно сможет пересчитать свой поправочный балл, для второго случая необходимо знать всю статистику и выполнять достаточно громоздкие вычисления.

Этап 10. Рассчитываем скорректированный тестовый балл t'_i : прибавляем к тестовому баллу $t_i = m$ величину поправочного значения Pr_i^m . Результат представлен в таблице 2.

Таблица № 2

Таблица скорректированных тестовых баллов для группы участников с одинаковыми первичными результатами

Задания	Тестовый балл i -го участника	Сумма сложностей задания	Поправочное значение	Скорректированные (уточненные) баллы
Номер участника				
n_1^m	$t_1 = m$	d_1^m	Pr_1^m	$t'_1 = t_1 + Pr_1^m$
n_2^m	$t_2 = m$	d_2^m	Pr_2^m	$t'_2 = t_2 + Pr_2^m$
n_3^m	$t_3 = m$	d_3^m	Pr_3^m	$t'_2 = t_2 + Pr_2^m$
...
n_p^m	$t_p = m$	d_p^m	Pr_p^m	$t'_p = t_p + Pr_p^m$

Далее выполняется перевод «сырых» баллов в требуемую шкалу, например, для перевода в шкалу 100-200 перевод осуществляется по методу эквипроцентильной нормализации [3]. Таблица перевода будет довольно большой по сравнению с исходной, так как дифференциация участников тестирования приводит к увеличению размерности, т.е. шаг в таблице перевода будет равен 0,1, а не 0,5 баллов. В данном алгоритме этот этап не рассматривается.

Для иллюстрации работы алгоритма рассмотрим основные результаты расчетов.

Пример расчета:

Рассмотрим результаты тестирования учеников общеобразовательных школ по иностранному языку. Тест проходило 1048 человек. Всего было 42 задания закрытой формы, по которым можно было набрать максимум 62 балла.

Рассчитаем поправочные коэффициенты для участников, набравших 10 «сырых» баллов, что соответствует 124 баллам по шкале 100-200. Всего 32 участника (3,1% от общего числа), которые набрали 10 баллов.

Результаты расчета поправочных коэффициентов и результирующих баллов представлены на рис. 1. Расчеты проводились в среде MS Excel.

	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	E
1								Ball	BallA	BallB	Ball100				dim	Pr	tnew		
2	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		5,557252	0,757953	10,75795		
3	0	0	0	0	1	1	1	10	10	0	124	129,05	115,36		6,439885	0,304163	10,30416		
4	0	1	0	0	1	0	1	10	10	0	124	129,05	115,36		6,862595	0,086833	10,08683		
5	0	0	0	1	0	0	1	10	10	0	124	129,05	115,36		6,110687	0,473414	10,47341		
6	0	0	0	0	1	0	0	10	10	0	124	129,05	115,36		7,031489	0	10		
7	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		7,010496	0,010793	10,01079		
8	0	0	1	1	0	0	0	10	10	0	124	129,05	115,36		6,414122	0,317408	10,31741		
9	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		6,179389	0,438092	10,43809		
10	1	0	0	0	1	0	0	10	10	0	124	129,05	115,36		5,599237	0,736368	10,73637		
11	0	0	0	0	0	0	1	10	10	0	124	129,05	115,36		6,829198	0,104004	10,104		
12	0	0	0	1	0	0	0	10	10	0	124	129,05	115,36		6,23187	0,411111	10,41111		
13	0	0	0	0	1	0	0	10	10	0	124	129,05	115,36		6,579198	0,232537	10,23254		
14	0	0	0	0	0	1	1	10	10	0	124	129,05	115,36		6,540076	0,252651	10,25265		
15	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		6,796756	0,120684	10,12068		
16	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		6,41126	0,31888	10,31888		
17	0	0	0	0	0	1	1	10	10	0	124	129,05	115,36		6,668885	0,187403	10,1874		
18	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		6,29771	0,37726	10,37726		
19	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		5,105916	0,99	10,99		
20	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		6,564885	0,239896	10,2399		
21	0	0	0	0	1	1	1	10	10	0	124	129,05	115,36		6,265267	0,39394	10,39394		
22	0	0	0	0	1	0	0	10	10	0	124	129,05	115,36		6,750954	0,144232	10,14423		
23	0	0	0	0	0	1	0	10	10	0	124	129,05	115,36		6,725191	0,157478	10,15748		
24	0	0	0	0	0	1	0	10	10	0	124	129,05	115,36		6,621183	0,210951	10,21095		
25	1	1	0	0	1	0	0	10	10	0	124	129,05	115,36		5,956107	0,552889	10,55289		
26	0	0	0	0	0	0	1	10	10	0	124	129,05	115,36		6,375	0,337522	10,33752		
27	0	1	1	0	1	0	0	10	10	0	124	129,05	115,36		6,248092	0,40277	10,40277		
28	1	0	0	0	1	0	0	10	10	0	124	129,05	115,36		5,900763	0,581343	10,58134		
29	0	0	0	0	1	0	0	10	10	0	124	129,05	115,36		5,162214	0,961056	10,96106		
30	0	0	1	0	1	0	0	10	10	0	124	129,05	115,36		6,115458	0,470961	10,47096		
31	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		5,412214	0,832522	10,83252		
32	0	0	0	0	0	0	0	10	10	0	124	129,05	115,36		6,46374	0,291898	10,2919		
33	1	0	0	0	0	0	0	10	10	0	124	129,05	115,36		6,73187	0,154044	10,15404		
34	0,52958	0,461832	0,542939	0,687977	0,658397	0,657443	0,674618								max	7,031489			
35	0	0	0	0	0	0	0								min	5,105916			

Рис. 1 Расчет поправочных значений и уточненных тестовых баллов

В процессе расчетов следует помнить, что сложность заданий рассчитывается для всего теста, а поправочные значения и скорректированные баллы считаются для группы участников, набравших одинаковый тестовый балл. Для каждой группы расчеты выполняются на отдельном листе рабочей книги MS Excel.

Проанализируем полученные результаты. Как видно из рисунка 2, с помощью предлагаемого алгоритма были дифференцированы участники тестирования, набравшие одинаковые «сырые» баллы. Скорректированные результаты распределены в пределах одного тестового балла. На рис. 3 представлены те же значения, но отсортированные по возрастанию. Результаты монотонно возрастают по закону, близкому к линейному, для данной группы участников скупенности на концах и в середине интервала не наблюдается.

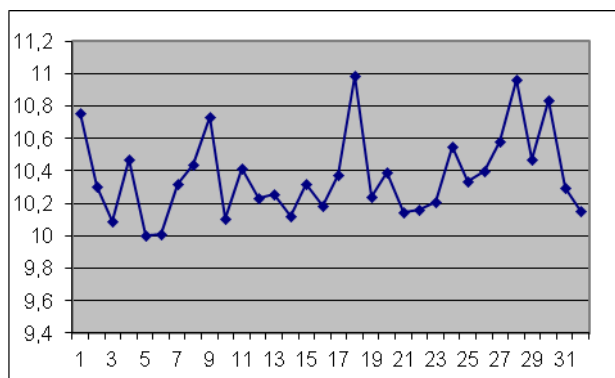


Рис. 2 Результаты уточненных «сырых» баллов

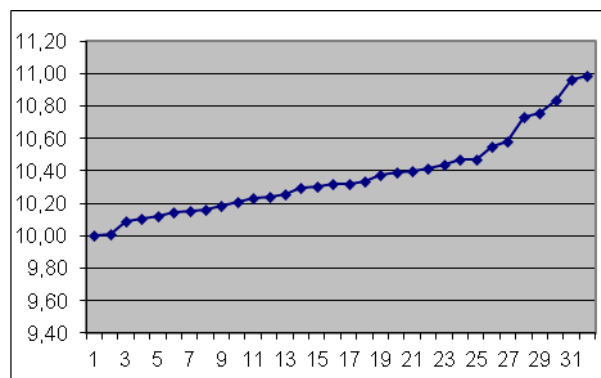


Рис. 3 График результирующих баллов, отсортированный по возрастанию

На рисунках 4 и 5 приведены результаты распределения участников тестирования до и после корректировки тестовых баллов в шкале 100-200. Диаграммы построены с помощью статистического пакета SPSS. Как видно, гистограмма скорректированных баллов является более сглаженной за счет уменьшения дискретности результатов.

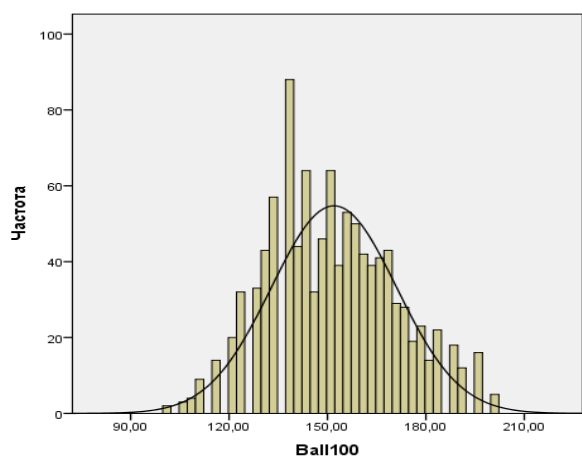


Рис. 4 Гистограмма распределения тестовых баллов до корректировки в шкале 100-200

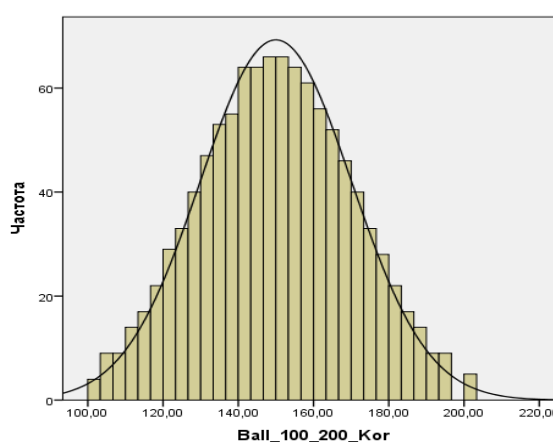


Рис. 5 Гистограмма распределения тестовых баллов после корректировки «сырого» балла в шкале 100-200

Выводы

1. Как показали результаты расчетов, предлагаемый алгоритм позволяет с использованием известных пакетов обработки данных оперативно рассчитать скорректированные баллы участников тестирования на основании учета сложности тестовых заданий;
2. Корректировка тестовых баллов участников на основании учета сложности тестовых заданий позволяет:
 - 2.1. уменьшить дискретность шкалы тестовых баллов на основе валидного увеличения число делений шкалы тестовых баллов, что ведет к более точной дифференциации участников тестирования;
 - 2.2. повысить эффективность использования метода эквипроцентильной нормализации тестовых баллов при переводе тестовых баллов в шкалу 100-200: распределение баллов в шкале 100-200 существенно ближе к эталонному нормальному распределению, а значит более точно определяются пороги 124, 140 и т.д. баллов, а также облегчит проведение ВУЗами конкурсного отбора в силу снятия (по крайней мере ослабления) проблемы полупроходного балла.
3. На основе предложенного алгоритма может быть разработана информационная система для корректировки тестовых баллов, которая позволит не только рассчитывать скорректированные значения тестовых баллов, но и проводить аналитические исследования результатов тестирования в рамках классической теории тестов: как параметров тестовых заданий, так и учебных достижений участников тестирования.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.
2. Ким В.С. Тестирование учебных достижений. Монография. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.
3. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов: Учебник. – М.: Логос, 2010. – 668 с.
4. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – Москва, 2000. – 168 с.
5. Роберт Б. Фрарі. Обчислення балів за формулою для тестів множинного вибору (коригування вгадування) // Вісник. Тестування і моніторинг в освіті. – 2008. – № 9. – С. 32–40.

6. Методика расчета тестового бала [Електрон. ресурс]. – Режим доступу: Республиканский институт контроля знаний. – 2000 – 2011: <http://www.rikz.unibel.by/ru/testing/info/metotdika.rar>
7. Раков С.А., Соколов О.Ю., Горох В.П. Методы «м'якого» уточнення тестового бала з урахуванням складності завдань // Вісник. Тестування і моніторинг в освіті. – 2011. – № 5-6. – С. 65 - 71.

Стаття надійшла до редакції 22.03.2013

Sergey Rakov, Maria Mazorchuk, Elena Bondarenko
National Aerospace University "KhAI"

ALGORITHM CORRECTION SCORE TEST BASED ON ANALYSIS COMPLEXITY OF TASKS

Purpose of work is develop an algorithm corrections test scores based on the analysis of complexity test questions. In this paper the tests with a closed form of the question are described. Corrections of test scores required for the differentiation of test participants who have scored equal test scores. The paper presents the main steps of the algorithm corrections of test results and present calculations that show the effectiveness of the algorithm. For the calculations used for data analysis packages such as SPSS and MS Excel.

Keywords: tests, test questions complexity, corrections test scores, a linear transformation

Раков С.А., Мазорчук М.С., Бондаренко О.О.

Национальный аэрокосмический университет им. Н.Е. Жуковского
«Харьковский авиационный институт»

АЛГОРИТМ КОРИГУВАННЯ тестових балів НА ОСНОВІ АНАЛІЗУ СКЛАДНОСТІ ЗАВДАНЬ

Метою роботи є розробка алгоритму коректування тестових балів на основі аналізу складності завдань тестів. В роботі розглянуті тести із закритою формою питання. Коригування балів тесту необхідна для диференціації учасників тестування, які набрали однакові тестові бали. У роботі представлені основні етапи алгоритму коректування результатів тестування та приведені розрахунки, що показують ефективність застосування даного алгоритму. Для проведення розрахунків використовуються пакети для аналізу даних, такі як SPSS та MS Excel.

Ключові слова: тести, складність завдань тесту, коректування тестових балів, лінійне перетворення, процентильне перетворення.