

УДК 004.89

А.Ю. СОКОЛОВ¹, О.Г. МОЛЧАНОВА¹, В.Г. ИВАНОВ²¹ *Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина*² *Национальный университет «Юридическая академия Украины им. Ярослава Мудрого»*

МЕТОДЫ ПОВЫШЕНИЯ ТОЧНОСТИ ШКАЛИРОВАНИЯ РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ

Работа является продолжением исследований авторов в области шкалирования и интерпретации результатов тестирования, применяемых в современных системах тестирования. В работе рассматриваются методы штрафования и стимулирования, основанные на классической теории тестирования. Предложен метод уточнения результатов, который позволяет улучшить показатели качества оценивания уровней подготовленности испытуемых как функции разнообразия оценочных баллов. В качестве основы уточнения баллов предложено использование сложностей заданий, влияющих на расположение участников тестирования внутри группы с одинаковым баллом. Приведен иллюстративный пример, демонстрирующий преимущества предложенного метода.

Ключевые слова: классическая теория тестов, штрафная функция, сложность задания, способность, латентные параметры.

Введение

Большинство исследований в области электронного обучения посвящено методикам проведения тестирования, анализу результатов и прогнозирования обучения в адаптивных системах. Развитию же методов педагогических измерений уделяется недостаточно внимания, подразумевая, что оценивание результатов базируется на суммарных оценках за выполненные задания и дальнейшем шкалировании результатов в той или иной шкале в зависимости от целей тестирования. Данная тема считается достаточно изученной как в классической теории педагогических измерений, так и в современной теории IRT – item response theory и хорошо представлена в литературе [1 – 3].

Однако всегда оставалась актуальной задача повышения точности оценок не только в определенной шкале, но и первичных оценок за пройденный тест – так называемых тестовых оценок. Под повышением точности понимается увеличение разрядности шкалы измерения. Обычно единицей разрядности тестовых баллов является 1. В работе [4] предложен подход к уточнению оценок в шкале логитов на основании учета сложности заданий в каждой их групп испытуемых, набравших одинаковый балл. Предложенный подход базировался на использовании модели Г. Раша и предполагал дополнительную итерационную процедуру уточнения баллов внутри каждой группы с дальнейшим перешкалированием результатов в общепринятую в системе тестирования в Украине шкалу 100-200 баллов. Поскольку предложенный подход основан на использовании теории IRT, осуществить его перенос в классиче-

скую теорию педагогических измерений непосредственно невозможно.

Целью настоящей статьи является разработка метода уточнения тестовых баллов, не выходя за рамки классической теории тестирования и сравнения результатов, полученных в предыдущей статье, основанной на модели IRT. Таким образом, предполагается создать комплекс методов уточнения результатов тестирования в наиболее распространенных теориях педагогических измерений.

Метод штрафования

Один из распространенных методов оценивания, позволяющих повысить разнообразие баллов, является введение штрафов за неправильный ответ на задания с множественным выбором – MCQ (multiple choice question) [1].

В основе этого метода лежит предположение, что участник тестирования, который дал неправильный ответ на задачу, мог это сделать сознательно или угадать. Поэтому для учета угадывания рекомендуется следующая формула

$$X_{\text{согг}} = R - \frac{W}{C-1}, \quad (1)$$

где $X_{\text{согг}}$ – скорректированный балл; R – количество правильных ответов; W – количество неправильных ответов; C – количество вариантов ответов на задания.

Например, пусть количество заданий – 50, количество правильных ответов – 25, количество неправильных ответов – 12, количество пропущенных заданий – 13, количество предлагаемых вариантов – 5. Тогда $X_{\text{согг}} = 25 - 12/(5 - 1) = 22$, а не 25, как было бы в исходном оценивании.

Рассмотрим еще один пример из результатов внешнего независимого тестирования по математике 2010 года. Из 36 заданий теста первые 25 заданий являются заданиями MCQ с одним правильным ответом из 5 дистракторов, остальные задания являются многобальными. Для задач с ответами MCQ можно применить метод штрафования, при этом будем считать, что все неправильные ответы являются результатом угадывания. Это, безусловно, преувеличение реального положения, но несущественное, чтобы продемонстрировать важный недостаток данного метода – перестановку положения участников тестирования при определенных исходных тестовых баллах после уточнения по формуле (1). Рассмотрим двух участников тестирования со следующими результатами, приведенными в таб. 1.

Первый участник набрал 14 тестовых баллов, а второй – 15. После применения формулы (1) к первым 25 заданиям, оказалось, что у первого участника скорректированный тестовый балл равен 11, а у второго – 9,75, т.е. влияние штрафных баллов наряду с уточнением шкалы привело не только к уменьшению исходного тестового балла, но и к изменению первичного ранжирования участников тестирования.

Таким образом, данный метод коррекции нарушает важный принцип в тестировании – он не должен изменять исходный порядок участников. Кроме того, метод штрафования является субъективно негативным для восприятия, поскольку уменьшает исходные тестовые баллы. Тем не менее, он на практике применяется в некоторых системах тестирования (например, в США в системе SAT [2]).

Метод, учитывающий сложность заданий

Рассмотрим один из известных методов шкалирования с использованием сложности заданий используется в системе государственного тестирования в Беларуси.

Пусть матрица $A = \{a_{ij}\}_{i,j=1,1}^{n,k}$ - матрица ответов

на k заданий типа MCQ, где n участники тестирования. Идея метода базируется на вычислении сложности заданий s_j

$$s_j = \frac{c_j}{n}, \quad (2)$$

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

Далее вычисляется тестовый балл как взвешенная сумма баллов с весом сложности для каждого участника с нормированием по следующей формуле

$$B_i = \frac{\sum_{j=1}^k a_{ij}(1-s_j)}{\sum_{j=1}^k (1-s_j)} \cdot 100, i = 1, \dots, n. \quad (3)$$

Рассмотрим на простом примере, что использование формулы (3), к сожалению, также может изменять первоначальный рейтинг.

Пример 1.

Пусть в тесте MCQ задание 1 имеет сложность 0,4, задание 2 – 0,8, задание 3 – 0,1.

Пусть участник 1 выполнил задания 1 и 2, а участник 2 – задание 3.

Тестовые баллы у них соответственно составляют 2 и 1.

Однако, используя процедуру взвешивания, получим для участника1 – $0,8 / (0,6 + 0,2 + 0,9) = 0,4706$ баллов, а для участника2 – $0,9 / (0,6 + 0,2 + 0,9) = 0,5294$ балла. То есть исходный рейтинг изменен (табл. 2).

Метод, учитывающий сложность заданий в группах

Метод базируется на следующих положениях:

- 1) не изменяет порядок расположения участников в шкале тестовых баллов;
- 2) расширяет диапазон значений в пределах интервала между соседними тестовыми баллами.

Таблица 1

Результаты тестирования

№ Задания № Участника	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Участник1	1	0	1	1	0	0	1	0	1	0	1	1	1	1	0	1	1	0
Участник2	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
№ Задания № Участника	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Участник1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Участник2	0	0	0	0	0	0	0	3	0	4	0	2	2	0	0	0	0	0

Таблица 2

Результаты тестирования с коррекцией

№ Участника \ № Задания	Задание1	Задание2	Задание3	Тестовый балл	Скорректированный тестовый балл
Участник1	1	1	0	2	0,4706
Участник2	0	0	1	1	0,5294
...
Сложность задания	0,4	0,8	0,1		

Алгоритм состоит в следующем.

Рассматриваются все группы участников тестирования, набравшие одинаковые баллы – от 0 до максимального в данной выборке. Рассмотрим группу участников, имеющую m баллов.

Если в тесте есть только задания типа MCQ, то m баллов означает, что эта группа выполнила ровно m заданий. Если в тесте есть задания с большим количеством баллов, то количество выполненных заданий может быть меньше для некоторых участников из этой группы. Рассмотрим этот более общий случай. Сложности всех задач также считаются известными.

Пусть матрица $A = \{a_{ij}\}_{i,j=1,1}^{n,k}$ – матрица ответов с многобалльными заданиями, где n участники тестирования. Сложность задания определяется по формуле

$$s_j = \frac{c_j}{n \cdot I_j}, \quad (4)$$

где $c_j = \sum_{i=1}^n a_{ij}$ – сумма баллов по j – му заданию, а

I_j – максимальный балл, предоставляемой за правильно выполненное-то задачи.

Пусть $N_m = \{n_1^m, \dots, n_p^m\}$ – группа участников, набравших m баллов.

Поскольку на этапе корректировки тестовых баллов учитывается только сложности заданий, будем считать, что влияние каждого задания на корректировку одинаков, т.е. для многобалльного задания уже сами по себе баллы внесли всей вклад в формирование тестового балла, и нет необходимости учитывать это еще и на корректировку. То есть каждое задание со своей сложностью, которое выполнил участник тестирования, имеет одинаковое влияние на корректировку. Тогда для множества N_m имеем множество из величин, составляющих соответствующие суммы сложностей

$$D_m = \{d_1^m, \dots, d_p^m\},$$

где $d_j^m = \sum_{v \in Q_j^m} s_v$; Q_j – множество номеров зада-

ний, которые выполнены участником тестирования n_j^m . Тогда имеем табл. 3 для группы N_m .

Таблица 3

Сложность задания

Участник	Сумма сложностей выполненных заданий
n_1^m	d_1^m
n_2^m	d_2^m
...	...
n_p^m	d_p^m

Далее процедура корректировки состоит в формировании скорректированного балла в интервале $[0, 1)$ значений в группе N_m в соответствии с суммами d_j^m .

Пример 2.

В табл. 4 приведены результаты теста и расчет суммарных сложностей. Как видно из таб.4, Участник2 имеет меньшее значение по сумме сложностей, чем Участник1, несмотря на то, что его задание оценено в 2 балла. Таким образом, предлагается два раза не поощрять за одно и тоже (он получил свои два балла уже как основной тестовый балл).

Множество значений скорректированный баллов для группы m с тестовыми баллами может быть не только в пределах $[0, 1)$, когда следующая группа тестовый балл $m+1$, но и большим $-[0, r-m)$, если следующая группа имеет тестовый балл $g > m+1$.

Таким образом, предложенная процедура может выполнять не только функцию более детальной корректировки, но и сглаживания, что актуально, к примеру, для небольших выборок.

Для этого можно воспользоваться разными методами, например, линейным или процентильным преобразованием. Может быть использована более упрощенная процедура, когда с учетом возможных значений сложности $s_j \in [0, 1]$ диапазон возможных дополнительных баллов может считать $[0, m]$. А каждому участнику, получившему m баллов, дополнительно добавляется величина $\frac{d_{jm}}{m}$.

Таблица 4

Окончательный результат

№ Задания / № Участника	Задание1	Задание2	Задание3	Тестовый балл	Сумма сложностей выполненных заданий	Скорректированный тестовый балл
Участник1	1	1	0	2	0,8	2,4
Участник2	0	0	2	2	0,7	2,35
...
Сложность задания	0,6	0,2	0,7			

Очевидно, что максимальное количество добавленных баллов также не превышает 1. Для данного метода в примере 4 скорректированный тестовый балл представлен в последнем столбце.

Заключение

В работе рассмотрена проблема шкалирования и некоторые пути уточнения (корректировки) тестовых баллов. Показано, что использование штрафных функций не является уместным по разным причинам. Другой путь – использование такой характеристики как сложность задач, можно использовать, причем желательно выбрать метод, который не нарушает исходного порядка ранжирования участников. Именно такой метод в рамках классической теории тестирования и предложен в работе.

Литература

1. Звонников, В.И. *Современные средства оценивания результатов обучения [Текст]: учеб. пособие для вузов / В.И. Звонников, М.Б. Чельщикова.* – М.: Издательский центр «Академия», 2007. – 224 с.
2. Самылкина, Н.Н. *Современные средства оценивания результатов обучения [Текст] / Н.Н. Самылкина.* – М.: Бином, Лаборатория знаний, 2007. – 172 с.
3. Чельщикова, М.Б. *Теория и практика конструирования педагогических тестов [Текст]: учебное пособие / М.Б. Чельщикова.* – М.: Логос, 2002. – 432 с.
4. Соколов, А.Ю. *Методы оценивания результатов тестирования в автоматизированных системах обучения [Текст] / А.Ю. Соколов, О.Г. Молчанова // Радиоэлектронні і комп'ютерні системи.* – 2011. – № 1 (49). – С. 117 - 123.

Поступила в редакцию 6.04.2011

Рецензент: д-р техн. наук, проф., заведующий кафедрой авиационных приборов Н.Д. Кошевой, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

МЕТОДИ ПІДВИЩЕННЯ ТОЧНОСТІ ШКАЛЮВАННЯ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

О.Ю. Соколов, О.Г. Молчанова, В.Г. Иванов

Робота є продовженням досліджень авторів в області шкалювання та інтерпретації результатів тестування, що застосовуються в сучасних системах тестування. У роботі розглядаються методи штрафування і стимулювання, засновані на класичній теорії тестування. Запропоновано метод уточнення результатів, що дозволяє поліпшити показники якості оцінювання рівнів підготовленості учасників тестування як функції різноманітності оціночних балів. В якості основи уточнення балів запропоновано використання складностей завдань, які впливають на розташування учасників тестування в межах групи з однаковим балом. Наведено ілюстративний приклад, який показує переваги запропонованого методу.

Ключові слова: класична теорія тестів, штрафна функція, складність завдання, здатність, латентні параметри.

METHODS TO IMPROVE THE ACCURACY OF SCALING TEST RESULTS

O.Y. Sokolov, O.G. Molchanova, V.G. Ivanov

Proposed paper is a continuation of research in the area of scaling and interpretation of test results used in modern test systems. The paper deals with methods penalty and motivation based on classical test theory. Method for refinement of the results that can improve performance quality assessment of levels of preparedness of the subjects as a function of the diversity of evaluation points is proposed. The base of refinement is using of difficulty of items that are used in each group with the same raw score. An illustrative example showing the benefits of the proposed method is given.

Keywords: classical test theory, the penalty function, the complexity of the job, ability, latent parameters.

Соколов Александр Юрьевич – д-р техн. наук, проф., зав. каф. информатики Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина.

Молчанова Ольга Георгиевна – аспирант кафедры информатики Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина.

Иванов Владимир Георгиевич – д-р техн. наук, проф., зав. каф. информатики и вычислительной техники Национального университета «Юридическая академия Украины им. Ярослава Мудрого», Харьков, Украина.