УДК 378.146, 004.94
DOI: http://dx.doi.org/10.21686/2500-3925-2019-2-89-98

Е.Б. Белов, М.В. Алексеев, Н.П. Китаев, А.И. Кучумов

Федеральное учебно-методическое объединение в системе высшего образования по укрупненной группе специальностей и направлений подготовки «Информационная безопасность», Москва, Россия

Анализ влияния первичных баллов заданий на уровни подготовленности респондентов в дихотомической модели Раша на основе взвешенного метода максимального правдоподобия

Цель исследования. *При обработке результатов тестирования часто используется модель Раша. Однако при использовании этой модели и метода максимального правдоподобия (ММП), оценки уровней подготовленности респондентов зависят только от числа правильно выполненных заданий теста и не зависят от трудности выполненных заданий. Цель исследования – анализ влияния трудностей заданий на уровни подготовленности респондентов на основе взвешенного метода максимального правдоподобия (ВММП). Для получения весовых коэффициентов ВММП в работе используются первичные баллы заданий.*

Материалы и методы. *Анализ влияния трудностей заданий на уровни подготовленности респондентов проведен при использовании дихотомической таблицы, полученной при тестировании знаний 19 респондентов по курсу основ электроники. Использовались индикаторные переменные 16 тестовых заданий. Для заданий рассчитывались первичные баллы, определяющие их трудности. Весовые коэффициенты используемого ВММП зависят от первичных баллов заданий и от коэффициента влияния К. При К = 0 ВМПП превращается в МПП. С увеличением К от 0 до 2 весовые коэффициенты увеличиваются и появляется возможность подробного анализа влияния трудностей заданий на уровни подготовленности респондентов. Для расчета параметров модели Раша на основе ВММП используются программы (М-файлы) для среды MATLAB и программа Ministep (Winsteps).*

Результаты. *Использование ВММП с весовыми коэффициентами, полученными на основе первичных баллов трудностей заданий, позволяет дополнительно дифференцировать уровни подготовленности респондентов в классической дихотомической модели Раша. Результаты анализа, проведенного с использованием данных теста по электронике, показывают, что при прочих равных условиях новые уровни подготовленности увеличиваются, если респонденты выполняют трудные задания и, наоборот, уровни подготовленности респондентов уменьшаются, если респонденты выполняют легкие задания. При этом уровни трудностей самих заданий практически не изменяются. Как правило, чем больше коэффициент влияния К, тем больше отличается оценка подготовленностей респондентов, полученная на основе ВММП от оценки на основе ММП. Однако имеются респонденты, уровень подготовленности которых не изменяется или изменяется незначительно при увеличении коэффициента К от 0 до 2. Для теста по электронике при малом коэффициенте $K \leq 1$ исходный порядок расположения респондентов по уровням подготовленности, рассчитанных на основе ММП, сохраняется. При увеличенном коэффициенте влияния $K \geq 1,5$ новые уровни подготовленности, рассчитанные с помощью ВММП, обуславливают изменение порядка распределения респондентов по уровням подготовленности. Расчеты, проведенные на основе полученных формул с помощью пакета MATLAB, подтверждаются данными полученными с помощью программы Winsteps. Отличия без учета экстремальных респондентов не превышают 0,01 логит при максимальном значении коэффициента К, равном 2.*

Заключение. *На основе взвешенного метода максимального правдоподобия предложена методика учета трудностей заданий на уровни подготовленности респондентов в дихотомической модели Раша при использовании первичных баллов заданий. Результаты анализа, проведенного с использованием данных теста по электронике, показывают, что в этом случае получим дифференцированные уровни подготовленности респондентов набравших одинаковые баллы по сравнению с оценками метода максимального правдоподобия. Заметим, что результаты, полученные с помощью ВМПП при использовании данных теста по электронике, не противоречат данным, полученным на основе классической дихотомической модели Раша и МПП. Результаты, полученные на основе ВММП, позволяют уточнить уровни подготовленности респондентов, полученные на основе ММП.*

Ключевые слова: *модель Раша, взвешенный метод максимального правдоподобия, весовые коэффициенты, программа Winsteps*

Evgeniy B. Belov, Mikhail V. Alekseev, Nikolay P. Kitaev, Aleksander I. Kuchumov

Federal Educational and Methodical Association in the System of Higher Education on the Enlarged Group of Specialties and Directions of Training «Information Security», Moscow, Russia

Analysis of influence of the item total scores on the levels of ability of respondents in the dichotomous Rasch model based on the weighted maximum likelihood method

Purpose of the study. *Rasch model is often used in processing test results. However, when using this model and the maximum likelihood method (ML), the estimates the levels of ability of respondents depend only on the number of correctly performed test items and do not depend on the difficulties of the items. The purpose of the research is to analyze the influence of the difficulties of the items on the levels of abilities of the respondents based on the weighted maximum likelihood method (WML). To obtain the weights of the WML, the item total scores are used.*

Materials and methods. *The analysis of the influence of the difficulties of the items on the levels of abilities of the respondents is investigated using the dichotomous table obtained when testing the knowledge of*

19 respondents in the course "Fundamentals of Electronics". Indicator variables of 16 test items were used. For items, we calculate the item total scores that determine their items difficulties. The weighting coefficients of the used WML depend on the item total scores and on the coefficient of influence K. When K = 0, WML convert into ML. As K increases from 0 to 2, the weighting coefficients increase and it becomes possible to analyze in detail the influence of the difficulties of the items on the respondents' ability levels. To calculate the parameters of the Rasch model based on WML, programs (M-files) for the MATLAB environment and Ministep (Winsteps) are used.

***Results.** The use of WML with weighting coefficient obtained on the basis of the item total scores of the difficulties of the items allowed us to further differentiate the levels of respondents' abilities in the dichotomous Rasch model. The results of the analysis performed using the data of the test on electronics show that, ceteris paribus, new levels of person's abilities increase if respondents perform difficult items and, conversely, the respondent's ability levels decrease if respondents perform light items. At the same time, the difficulty levels of the items practically do not change. As a rule, the greater the coefficient of influence K, the more different the estimation of abilities of respondents, obtained on the basis of WML, from the estimation on the basis of ML. However, there are respondents whose ability level does not change or change slightly when the coefficient K is increased from 0 to 2. For the data of the test on electronics with a coefficient K ≤ 1, the original order of respondents in their ability levels calculated on the basis of ML is preserved. With an increased coefficient of influence K ≥ 1,5, new levels of ability, calculated using WML, cause a change in the order of distribution of respondents according to ability levels. Calculations performed using the MATLAB package are confirmed by data obtained using the Winsteps program. Differences without extreme respondents do not exceed 0.01 logit with the maximum value of the coefficient K equal to 2.*

***Conclusion.** On the basis of WML, a method is proposed for taking into account the influence of the difficulties of items on the levels of respondents' abilities in the Rasch dichotomous model when using the item total scores. The results of the analysis performed using the data of the test on electronics show that in this case we will obtain a differentiation of the levels of abilities of the respondents who score the same points. Note that the results obtained using WML and using the data of the test on electronics do not reject the data obtained on the basis of the classical dichotomous Rasch model and ML. The results obtained on the basis of WML, allow to refine the levels of abilities of the respondents, obtained on the basis of ML.*

***Keywords:** Rasch model, weighted likelihood, weight coefficients, program Winsteps*

Введение

В современной теории тестирования для определения латентных показателей качества образовательного процесса часто используется модель Раша [1–5]. В связи с этим актуальными становятся теоретические и практические вопросы оценивания параметров модели Раша и анализ влияния на них различных факторов.

Параметры модели Раша, как правило, рассчитываются при использовании метода максимального правдоподобия (ММП). Как известно, оценки максимального правдоподобия являются состоятельными и асимптотически эффективными. Методы максимального правдоподобия используются для определения параметров модели Раша в коммерческих программах RUMM 2030 и Winsteps.

Однако, при использовании ММП, когда объем выборки конечен, оценки уровней подготовленности респондентов зависят только от числа правильно выполненных заданий теста и не зависят от трудности выполненных заданий. Это дает повод некоторым исследователям отказываться от использования ММП и модели Раша [6–9].

Одним из известных методов, позволяющим устранить отмеченный недостаток, является взвешенный метод максимального правдоподобия (ВММП) [10–11]. Взвешивание часто используется в статистике. Весовые коэффициенты позволяют управлять свойствами математической модели, обеспечивая более точное описание реальных явлений. Как известно, оценки ММП могут быть смещёнными. Поэтому уточненные уровни подготовленности респондентов, полученные с помощью ВММП, могут лучше соответствовать модели Раша.

Для получения весовых коэффициентов используются различные методы на основе различных критериев. Например, Warm [12], ввел весовую функцию в уравнение логарифмического правдоподобия, которая обеспечивает коррекцию смещения оценок параметров модели Раша. Эффективность предложенного взвешивания рассмотрена в работе [13]. Linacre описал простой метод взвешивания, который реализован в компьютерной программе Winsteps [14]. Метод взвешенного правдоподобия, используемый для совместного анализа дихотомических и политомических данных, описан в статье [15].

В данной работе предлагается использовать весовые коэффициенты, зависящие от первичных баллов заданий и от коэффициента влияния *K*. При *K* = 0 весовые коэффициенты равны 1 и ВММП превращается в ММП. С ростом *K* весовые коэффициенты увеличиваются, и появляется возможность анализа различий в оценках уровней подготовленности респондентов при использовании ММП и ВММП. Выбор первичных баллов заданий для определения весовых коэффициентов, объясняется простотой их расчета, а также тем, что они являются достаточными статистиками.

1. ВММП с весовыми коэффициентами, рассчитанными на основе первичных баллов заданий

Логистическая функция, лежащая в основе дихотомической модели Раша, имеет вид

$$p_{ij} = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)},$$

где p_{ij} – вероятность правильного ответа i-го респондента на j-е задание теста, θ_i – переменный уровень подготовленности i-го респондента, β_j – переменный уровень трудности j-го задания.

Вероятности появления данных, приводимых в дихотомической таблице, при использовании классической модели Раша равны

$$p_{ij} = \frac{\exp[a_{ij}(\theta_i - \beta_j)]}{1 + \exp(\theta_i - \beta_j)},$$

где a_{ij} – элемент дихотомической матрицы, равный 1 при правильном ответе и 0 при неправильном ответе. Функция правдоподобия L определяется в виде произведения

$$L = \prod_{i=1}^{N} \prod_{j=1}^{M} p_{ij},$$

где N – число респондентов, M – число заданий. Учитывая, что при использовании логарифма свойство монотонности сохраняется, при расчетах удобно использовать логарифмическую функцию правдоподобия

$$\ln L = \\ = \sum_{i=1}^{N} \sum_{j=1}^{M} [a_{ij}(\theta_i - \beta_j) - \ln(1 + \exp(\theta_i - \beta_j))].$$

Метод максимального правдоподобия позволяет найти оценки уровней трудностей заданий β^*_j и уровней подготовленности респондентов θ^*_i, при которых функция правдоподобия максимальна. Уровни подготовленности респондентов определяются первичными баллами, полученными респондентами. Эти первичные баллы являются достаточными статистиками [1]. Респонденты, набравшие одинаковые первичные баллы, получают одинаковые оценки уровней подготовленности. Действительно, в приведенной выше функции правдоподобия используется произведение вероятностей, где сомножители можно переставить местами. При транспонировании дихотомической матрицы, при перестановке ее строк или столбцов, при перемене местами двух любых элементов дихотомической матрицы и т.п. ММП будет давать один и тот же результат, т.к. информация о сложности j-го задания, которая определяется нулями или единицами в j-м столбце дихотомической матрицы, при этом утрачивается.

Число различных значений уровней подготовленности будет определяться числом заданий в тесте. Например, для теста, включающего16 заданий, число различных значений уровней подготовленностей равно 17 (от 0 до 16). При этом минимальная цена деления измерительной шкалы в логитах примерно равна $d = 4/M$ логит [1]. При числе заданий $M = 16$ цена деления d равна примерно 0,25 логит. С большей точностью расчеты уровней подготовленности на основе ММП невозможны.

Чтобы учесть трудность заданий и повысить точность оценки латентных параметров, функцию правдоподобия запишем с использованием весовых коэффициентов w_j следующим образом

$$L = \prod_{i=1}^{N} \prod_{j=1}^{M} p_{ij}^{w_j}.$$

Отметим, что в приведенной формуле записывается особая вероятность, которая уже не является обычной из-за возведения в степень. Логарифмируя приведенное соотношение, получим

$$\ln L = \\ = \sum_{i=1}^{N} \sum_{j=1}^{M} [w_j a_{ij}(\theta_i - \beta_j) - w_j \ln(1 + \exp(\theta_i - \beta_j))]. \quad (1)$$

Используя дихотомическую таблицу результатов тестирования, найдем c_j – первичный балл j-го задания, равный числу правильных ответов респондентов на задание. Разность $q_j = N - c_j$ характеризует трудность задания. Назовем эту разность q_j первичным баллом трудности j-го задания. Целое число q_j равно числу неправильных ответов респондентов на j-е задание. Чем их больше, тем труднее задание.

Первичные баллы, которые далее будем использовать для определения весовых коэффициентов, выбраны с учетом их достоинств: объективности, первичности и простоты.

Весовые коэффициенты на основе известных первичных баллов трудностей заданий рассчитываем по формуле

$$w_j = \left(q_j \frac{M}{\sum_{j=1}^{M} q_j} - 1 \right) K + 1, \quad (2)$$

где K – коэффициент влияния трудностей заданий на весовые коэффициенты и соответственно на уровни подготовленности респондентов. При $K = 0$ все весовые коэффициенты равны единице и ВММП превращается в ММП. Если $K = 1$, то весовые коэффициенты равны нормированным первичным баллам трудностей заданий.

Разница между большими и малыми весовыми коэффициентами не должна быть слишком большой. Рекомендуется для больших весовых коэффициентов выбирать значения K в диапазоне от 1 до 2, а для малых весовых коэффициентов выбирать значения K в диапазоне от 0 до 1 [15].

Дробь $M / \sum_{j=1}^{M} q_j$ в формуле (2) обеспечивает нормирование весовых коэффициентов так, чтобы общий объем статистической информации в данных при использовании весовых коэффициентов не изменялся. Поэтому сумма весовых коэффициентов независимо от указанных выше значений коэффициентов K равна числу заданий M.

2. Дихотомическая таблица и весовые коэффициенты ВММП

В табл. 1 представлены индикаторные переменные 16 тестовых заданий по вводной теме курса основ электроники [16] для 19 респондентов.

Тестовые задания по основам электроники разработаны для проведения критериально-ори-

Таблица 1

Номер респондента i	Дихотомические результаты выполнения заданий	MatLab $K = 0$ θ_i^*	MatLab $K = 1$ θ_i^*	MatLab $K = 2$ θ_i^*	Winsteps $K = 2$ θ_i^*
1	1 0 1 1 1 1 0 0 0 1 0 1 1 0 0 0	-0,01	-0,13	-0,26	-0,26
2	1 1 0 0 1 0 1 0 1 0 0 1 1 1 1 1	0,55	0,76	0,99	0,98
3	0 1 1 0 1 1 1 1 0 0 1 1 1 1 0 1	0,85	0,85	0,85	0,85
4	1 1 0 1 1 0 0 1 1 0 0 1 1 0 1 0	0,26	0,44	0,63	0,63
5	1 1 0 1 1 1 1 0 1 1 1 1 0 1 0 1	1,18	1,16	1,14	1,14
6	1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1	2,09	2,14	2,19	2,19
7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	-20,68	-19,06	-19,15	-3,80
8	1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1	2,09	1,93	1,81	1,81
9	0 1 1 0 1 1 1 0 0 1 1 0 1 1 0 0	0,26	0,21	0,17	0,17
10	1 1 0 1 1 0 1 0 0 0 0 1 0 0 0 0	-0,56	-0,79	-1,07	-1,07
11	1 1 1 1 0 1 1 0 0 1 1 1 1 0 1 1	1,18	1,20	1,23	1,22
12	1 0 1 1 0 0 0 1 1 0 0 1 0 1 0 0	-0,28	-0,21	-0,13	-0,13
13	1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1	2,88	3,10	3,30	3,29
14	1 0 1 1 1 1 1 1 1 1 1 0 1 1 0 0	1,18	1,20	1,23	1,22
15	0 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0	0,55	0,48	0,43	0,43
16	1 1 1 1 0 1 1 1 0 1 0 1 0 1 1 0	0,85	0,81	0,77	0,77
17	1 1 0 1 1 1 1 1 1 1 1 0 1 0 0 1	1,18	1,25	1,32	1,32
18	0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	-2,85	-2,91	-2,98	-2,98
19	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	20,43	19,87	20,80	4,49

ентированного тестирования с выборкой из большого банка заданий. Для расчета параметров модели Раша использовалась демонстрационная версия программы Winsteps [14] – программа Ministep, вер. 4.3.0 (далее Winsteps). Уровни трудностей разработанных заданий распределены в относительно малом диапазоне, равном примерно 2,2 логит. Экстремальные задания были исключены на этапе разработки теста. С помощью программы Winsteps задания теста, плохо соответствующие модели Раша, были отредактированы или исключены.

Для анализируемого теста с помощью программы Winsteps получена вероятность хи-квадрат, равная примерно 0,58. Это значение существенно выше критического значения. В данном случае совместимость исходных данных табл. 1 с моделью Раша, по установившейся практике, считается удовлетворительной.

Показатель альфа Кронбаха (KR-20) равен с учетом экстремальных респондентов 0,84. Значение этого показателя, большее 0,6, подтверждает, что тест для измерения параметров модели Раша является надежным и воспроизводимым.

Используя дихотомические данные табл. 1 рассчитываем первичные баллы заданий c_j, а затем первичные баллы трудностей заданий q_j (табл. 2). Зная первичные баллы q_j и используя формулу (2), рассчитываем весовые коэффициенты w_j. Весовые коэффициенты при K, равном 1 и 2, приведены в табл. 2.

Таблица 2

j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
q_j	5	5	6	7	7	6	5	8	9	7	8	5	7	8	12	10
w_j при $K = 1$	0,6957	0,6957	0,8348	0,9739	0,9739	0,8348	0,6957	1,1130	1,2522	0,9739	1,1130	0,6957	0,9739	1,1130	1,6696	1,3913
w_j при $K = 2$	0,3913	0,3913	0,6696	0,9478	0,9478	0,6696	0,3913	1,2261	1,5043	0,9478	1,2261	0,3913	0,9478	1,2261	2,3391	1,7826

3. Расчет уровней подготовленности респондентов при использовании ВММП и пакета MATLAB

Для расчета уровней подготовленности респондентов и уровней трудности заданий написана программа (М-файлы) в среде MATLAB.

После старта MATLAB-программа с помощью функции dlmread читает дихотомические данные и определяет две глобальные переменные – число строк N_st и число столбцов N_item дихотомической матрицы. Эти числа используются для задания вектора-строки tb, содержащего нулевые начальные значения уровней подготовленности респондентов и уровней трудности заданий теста.

Для поиска экстремума логарифмической функции максимального правдоподобия использовалась функция fminunc [17]. Все расчеты с помощью этой функции проводились при указании следующих опций: 'LargeScale', 'Off, 'TolFun', 1e-6, 'TolX', 1e-6, 'MaxFunEvals', 1e6.

После определения экстремума из вектора tb выделяется подстрока th, содержащая уровни подготовленности респондентов, и подстрока bt, содержащая уровни трудностей заданий.

Для удобного сравнения полученных данных с результатами расчетов на основе программы Winsteps определяется среднее арифметическое значение mean переменных в массиве уровней трудностей заданий bt, которое затем вычитается из всех значений уровней подготовленности и уровней трудности.

Расчет логарифмической функции максимального правдоподобия был выделен в отдельную подпрограмму-функцию, включающую два вложенных цикла по числу респондентов N_st и по числу заданий N_item:

```
for i = 1:N_st;
 for j = 1:N_item;
   En=En-w(j)*a(i,j)*(th(i)-bt(j))+w(j)*log(1+exp(th(i)-bt(j)));
 end
end
```

Так как функция fminunc предназначена для поиска минимума, то в приведенном выше фрагменте программы знаки перед весовыми коэффициентами изменены на противоположные по сравнению со знаками в формуле (1).

Весовые коэффициенты w_j задаются в виде вектора строки вида

w=[0.3913 0.3913 ... 1.7826];

Рассчитанные с помощью пакета MATLAB уровни подготовленности респондентов при коэффициентах *K*, равных 0, 1 и 2, приведены в табл. 1.

На рис. 1 приведены уровни подготовленности респондентов, рассчитанные с помощью программы (М-файлов) для MATLAB при коэффициенте *K*, равном 0 (ММП) и при коэффициенте *K*, равном 2 (ВММП). Отметим, что большие по модулю уровни подготовленности экстремальных респондентов (7-й и 19-й респонденты), как правило, уточняются по специальным методикам и в дальнейшем не рассматриваются [18, 19].

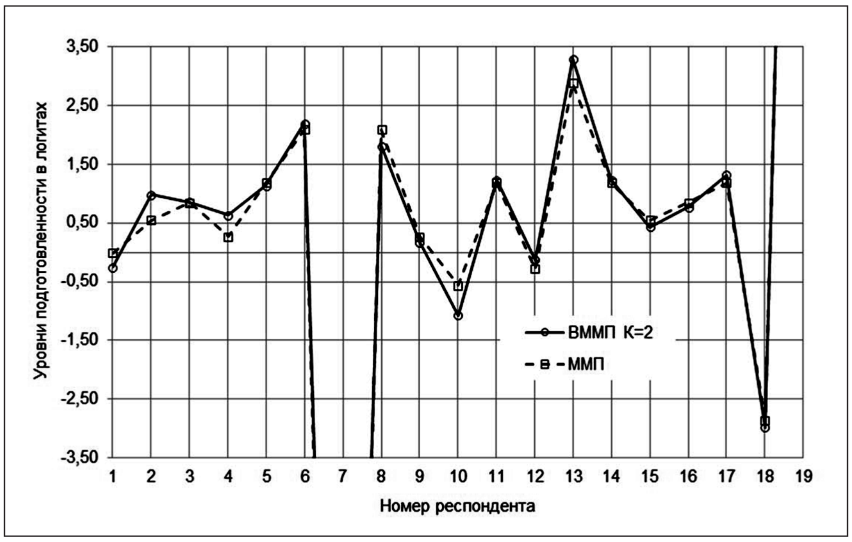

Рис. 1. Уровни подготовленности респондентов, рассчитанные с помощью ММП и ВММП

Как видим, результаты, полученные с помощью ВММП, близки к данным ММП. Как и следовало ожидать, они не противоречат описанию классической дихотомической модели Раша, полученной на основе ММП. Новые результаты только уточняют параметры модели Раша.

При этом уровни трудностей заданий при использовании описанного ВММП практически не изменяются по сравнению с ММП. Даже при $K = 2$ новые уровни трудностей заданий отличаются от уровней трудностей, полученных на основе ММП, как правило, максимум на 0,02 логит. Только для самого сложного 15-го задания, уровень трудности которого максимален и равен 1,47 (при $K = 0$), наблюдается изменение на 0,04 логит.

На рис. 2 приведены зависимости уровней подготовленности 4-го и 9-го респондентов при увеличении коэффициента влияния *K*. Эти респонденты имели одинаковые первичные баллы, но разные суммы первичных баллов трудностей заданий: у 4-го респондента эта сумма (65), больше суммы первичных баллов трудностей заданий 9-го респондента (59).

Из анализа кривых рис. 2 следует, что при прочих равных условиях уровень подготовленности респондента увеличивается, если респондент выполняет трудные задания и, наоборот, уровень подготовленности респондента уменьшается, если он выполняет легкие задания. Причем, чем больше коэффициент *K*, тем больше отличается оценка подготовленностей респондентов, полученная на основе ВММП от оценки на основе ММП. В этом случае оценки, основанные на ВММП, могут лучше соответствовать истинной подготовленности респондентов, по сравнению с оценками ММП. Такие оценки можно использовать для дополнительной дифференциации респондентов по уровню знаний.

Однако из анализа данных табл. 1 следует, что имеются респонденты, уровень подготовленности которых не изменяется (3-й респондент) или изменяется незначительно (11-й и

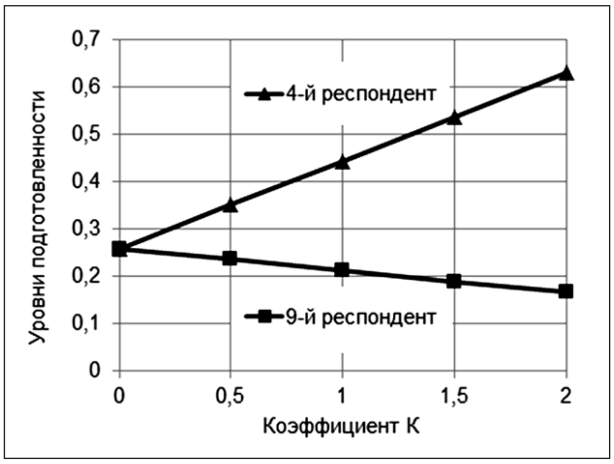

Рис. 2. Уровни подготовленности 4-го и 9-го респондентов

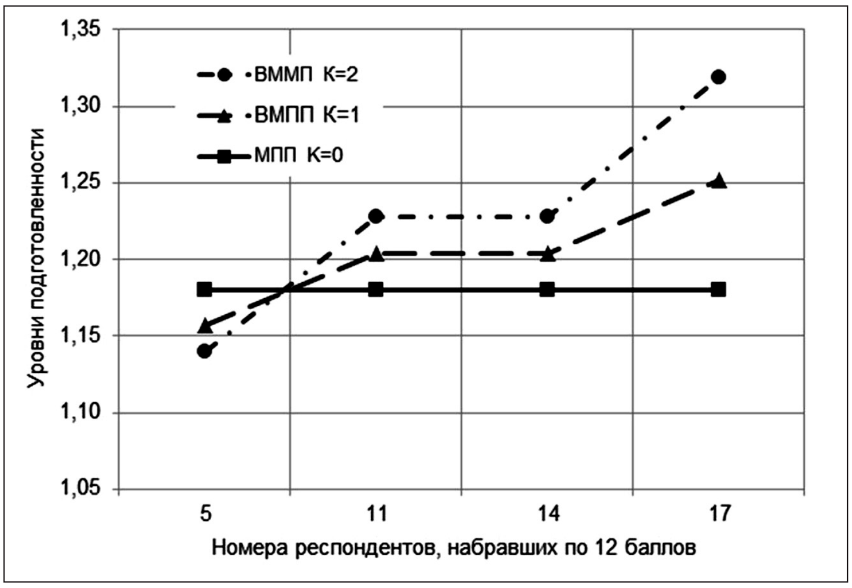

Рис. 3. Уровни подготовленности 5-го, 11-го, 14-го и 17-го респондентов

14-й респонденты) при увеличении коэффициента *K* от 0 до 2.

На рис. 3 показаны уровни подготовленности 4-х респондентов при коэффициентах *K*, равных 0, 1 и 2. Кривая, отмеченная квадратиками, соответствует одинаковым уровням подготовленности, полученным при использовании ММП (при $K = 0$). С ростом *K* уровень подготовленности 5-го респондента снижается. Он набрал только 82 балла за выполненные задания. Уровни подготовленности 11-го, 14-го и 17-го респондентов с ростом *K* увеличиваются, так как они правильно выполнили более трудные задания теста.

Отметим, что 11-й и 14-й респонденты получили не только одинаковые баллы за выполненные задания, но и набрали одинаковые суммы (по 83 балла) первичных баллов трудностей заданий. Поэтому при использовании ВММП одинаковый уровень подготовленности этих респондентов сохраняется.

На рис. 4 приведены уровни подготовленности респондентов, первичные баллы которых за правильно выполненные задания отличаются на единицу, а уровни подготовленности, полученные на основе ММП, отличаются примерно на 0,25 логит. Как видим, при $K \leq 1$ исходный порядок расположения респондентов по уровням подготовленности сохраняется (этот вывод после анализа данных табл. 1 можно распространить на всех респондентов). Если $K > 1{,}5$, то новые уровни подготовленности, рассчитанные с помощью ВММП, приводят к изменению порядка распределения респондентов по уровням подготовленности. Уровень подготовленности 12-го респондента, набравшего 7 баллов, становится больше чем уровень подготовленности 1-го респондента, набравшего 8 баллов. Анализ показал, что 12-й респондент правильно выполнил три из шести самых сложных заданий в тесте, а 1-й респондент не ответил ни на одно из этих сложных заданий теста.

Таким образом, уровни подготовленности при использовании ВММП уточняются сложным образом. На изменение уровня подготовленности влияют не только суммы первичных баллов трудностей заданий, но и больший вклад отдельных сложных заданий теста. Например, 5-й респондент набрал сумму первичных баллов трудностей заданий, равную 82, а 3-й респондент набрал только 75 таких баллов. Однако с ростом *K* уровень подготовленности 5-го респондента снижается, а 3-го остается неизменным. С ростом весовых коэффициентов при $K > 1{,}5$ уровень подготовленности 2-го респондента становится больше уровней подготовленности 3-го и 16-го респондентов. Аналогично меняются местами 4-й и 15-й респонденты.

4. Расчет уровней подготовленности респондентов с использованием весовых коэффициентов и программы Winsteps

В программе Winsteps с помощью параметров IWEIGHT и PWEIGHT можно указать

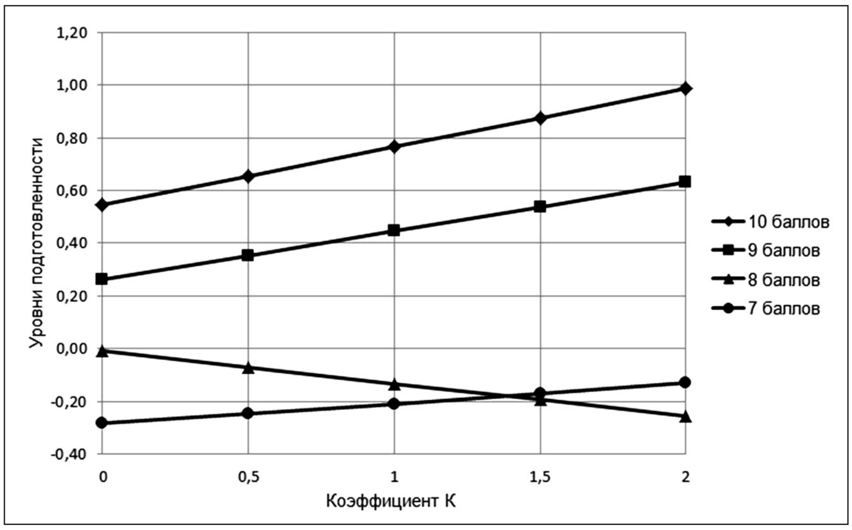

Рис. 4. Уровни подготовленности, респондентов, первичные баллы которых равны 7, 8, 9 и 10 баллов

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	WEIGH	item
1	14	19	-1.1046	.6604	1.12	.4	.87	.0	.56	.73	76.5	82.0	.39	i01
2	14	19	-1.1046	.6604	.98	.1	.85	.0	.59	.73	76.5	82.0	.39	i02
3	13	19	-.7047	.6096	1.62	1.8	2.03	1.6	.29	.63	70.6	77.6	.67	i03
4	12	19	-.3549	.5772	1.36	1.4	1.45	1.0	.40	.56	52.9	73.8	.95	i04
5	12	19	-.3549	.5772	1.54	1.9	2.83	2.8	.26	.56	64.7	73.8	.95	i05
6	13	19	-.7047	.6096	.85	-.4	.71	-.4	.63	.63	82.4	77.6	.67	i06
7	14	19	-1.1046	.6604	.91	-.1	.65	-.4	.63	.73	76.5	82.0	.39	i07
8	11	19	-.0350	.5572	.96	-.1	.83	-.3	.56	.50	76.5	71.0	1.23	i08
9	10	19	.2681	.5466	1.07	.4	1.02	.2	.50	.45	64.7	68.7	1.50	i09
10	12	19	-.3549	.5772	.92	-.2	.76	-.4	.59	.56	76.5	73.8	.95	i10
11	11	19	-.0350	.5572	.78	-1.0	.65	-.7	.64	.50	76.5	71.0	1.23	i11
12	14	19	-1.1046	.6604	1.26	.8	1.30	.6	.47	.73	76.5	82.0	.39	i12
13	12	19	-.3549	.5772	.89	-.3	.75	-.4	.60	.56	76.5	73.8	.95	i13
14	11	19	-.0350	.5572	.97	-.1	.84	-.2	.56	.50	64.7	71.0	1.23	i14
15	7	19	1.1704	.5635	.94	-.2	.79	-.1	.54	.36	70.6	71.7	2.34	i15
16	9	19	.5642	.5441	.68	-1.9	.58	-.9	.66	.41	88.2	67.7	1.78	i16
MEAN	10.8	19.0	.0000	.5749	1.01	.0	1.00	.0			72.8	72.8		
P.SD	2.1	.0	.6584	.0330	.25	1.0	.56	.9			8.6	4.0		

Рис. 5. Таблица 14.1 с уровнями трудностей заданий, полученная с помощью Winsteps при $K = 2$

весовые коэффициенты для заданий и респондентов соответственно [9]. По умолчанию эти весовые коэффициенты в программе равны 1. Как указывается в описании программы, весовому коэффициенту, равному 2, соответствует вероятность, которая дважды появляется в файле данных. Из этого указания следует, что, по-видимому, в программе Winsteps также используется возведение вероятностей в степень.

На рис. 5 приведена таблицы 14.1, полученная с помощью программы Winsteps при указании с помощью параметра IWEIGHT весовых коэффициентов, полученных при K, равном 2 (табл. 2). На этом рисунке выделен параметр MEAN в столбце MEASURE, равный нулю. Простая проверка показала, что среднее арифметическое приведенных в таблице трудностей заданий не равно нулю, а равно $\Delta = -0{,}3344$. Оказалось, что при использовании весовых коэффициентов в столбце MEASURE указывается взвешенное среднее [20]. По этой причине первоначальные данные для уровней подготовленности, полученные с помощью программы Winsteps, существенно отличались от данных полученных с помощью MATLAB.

Однако, после смещения рассчитанных в Winsteps уровней подготовленности респондентов на указанную выше величину Δ разница уровней, полученных с помощью Winsteps и MATLAB, не превышала 0,01 логит для неэкстремальных респондентов (табл. 1). Таким образом, расчеты, проведенные с помощью MATLAB, подтверждаются результатами, полученными с помощью Winsteps.

Заключение

Использование ВММП с весовыми коэффициентами, полученными на основе первичных баллов трудностей заданий, позволяет дополнительно дифференцировать уровни подготовленности респондентов в классической дихотомической модели Раша. При прочих равных условиях новые уровни подготовленности увеличиваются, если респонденты выполняют трудные задания и, наоборот, уровни подготовленности респондентов уменьшаются, если респонденты выполняют легкие задания. При этом уровни трудностей самих заданий практически не изменяются.

Как правило, чем больше коэффициент влияния K, тем больше отличается оценка подготовленностей респондентов, полученная на основе ВММП от оценки на основе ММП. Однако имеются респонденты, уровень подготовленности которых не изменяется или изменяется незначительно при увеличении коэффициента K от 0 до 2.

Для теста по электронике при коэффициенте $K \leq 1$ исходный порядок расположения респондентов по уровням подготовленности, рассчитанных на основе ММП, сохраняется. При коэффициенте влияния $K \geq 1{,}5$ новые уровни подготовленности, рассчитанные с помощью ВММП, обуславливают изменение порядка распределения респондентов по уровням подготовленности.

Расчеты, проведенные на основе полученных формул с помощью пакета MATLAB, подтверждаются данными полученными с помощью программы Winsteps. Отличия без учета экстремальных респондентов не превышают 0,01 логит при максимальном значении коэффициента *K*, равном 2.

Литература

1. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. М.: Прометей, 2000. 169 с.
2. Ким В.С. Тестирование учебных достижений. Монография. Уссурийск: Издательство УГПИ, 2007. 214 с.
3. Аванесов В. С. Item Response Theory: Основные понятия и положения // Педагогические измерения. 2007. № 2. С. 3–28.
4. Челышкова М.Б. Теория и практика конструирования педагогических тестов: Учеб. пособие. М.: Логос, 2002. 432 с.
5. Маслак А.А. Теория и практика измерения латентных переменных в образовании. М.: Юрайт, 2016. 256 с.
6. Попов А.П. Критический анализ параметрических моделей Раша и Бирнбаума // Материалы 4-й НМК «Инновационные методы и средства оценки качества образования». М.: Изд-во МГУП, 2006. С. 231–235.
7. Шрайфель И.С., Елисеев И.Н. Теоретическое обоснование единого итерационного процесса совместной количественной оценки трудностей заданий и уровней подготовки студентов // Сибирский журнал вычислительной математики. 2016. Т. 19. № 1. С. 107–123.
8. Сербин В.И. Управление процессом обучения с помощью марковской модели // Прикаспийский журнал: управление и высокие технологии. 2013. № 4 (24). С. 29–35.
9. Сёмов А.М., Сёмова М.А., Хлебников В.А. Единый итерационный процесс совместной количественной оценки трудностей заданий и уровней подготовленности участников тестирования // Труды центра тестирования, выпуск 2. М.: Издательство «Прометей», 1999. С. 54–60.
10. Hu F., Zidek J.V. The weighted likelihood // The Canadian Journal of Statistics. 2002. Vol. 30. № 3. P. 347-371. DOI: 10.2307/3316141.
11. Симахин В.А., Черепанов О.С. Исследование оценок взвешенного метода максимального правдоподобия // Вестник Курганского государственного университета. Серия: Технические науки. 2011. № 6. С. 72–77.
12. Warm T.A. Weighted Likelihood Estimation of Ability in Item Response Theory // Psychometrika. 1989. Vol. 54. P. 427–450. DOI: 10.1007/BF02294627.
13. Linacre J.M. The Efficacy of Warm's Weighted Mean Likelihood Estimate (WLE) Correction to Maximum Likelihood Estimate (MLE) Bias. Rasch Measurement Transactions. 2009. Vol. 23. № 1. P. 1188–1189.
14. John M. Linacre. A User's Guide to WINSTEPS, MINISTEP. Rasch-Model Computer Programs, 2017. [Электрон. ресурс] Режим доступа: http://www.winsteps.com/winman/copyright.htm (Дата обращения: 07.04.2019).
15. Tao, J., Shi, N., & Chang, H. Item-weighted likelihood method for ability estimation in tests composed of both dichotomous and polytomous items // Journal of Educational and Behavioral Statistics. 2012. Vol. 37. № 2. P. 298–315. DOI: 10.3102/1076998610393969.
16. Кучумов А. И. Электроника и схемотехника. Учебное пособие. М.: Гелиос АРВ, 2011. 336 с.
17. MatLab. Функция fminunc. [Электрон. ресурс] Режим доступа: https://www.mathworks.com/help/optim/ug/fminunc.html?s_tid=srchtitle (Дата обращения: 07.04.2019).
18. Алексеев М.В., Белов Е.Б., Китаев Н.П., Кучумов А.И. Отличия в оценках уровней подготовленности экстремальных респондентов при использовании классической теории тестирования и дихотомической модели Раша // Информатизация образования и науки. 2018. 2(38). С. 99–107.
19. Wright B.D. Estimating measures for extreme scores // Rasch Measurement Transactions. 1998. 12 (2). P. 632–633.
20. Parameter MEAN in table 14.1 when using IWEIGHT. [Электрон. ресурс] Режим доступа: http://raschforum.boards.net/thread/796/parameter-mean-table-using-iweight (Дата обращения: 07.04.2019).

References

1. Neyman YU.M., KHlebnikov V.A. Vvedeniye v teoriyu modelirovaniya i parametrizatsii pedagogicheskikh testov = Introduction to the theory of modeling and parameterization of pedagogical tests. Moscow: Prometheus; 2000. 169 p. (In Russ.)
2. Kim V.S. Testirovaniye uchebnykh dostizheniy. Monografiya = Testing educational achievements. Monograph. Ussuriysk: Publisher UGPI; 2007. 214 p. (In Russ.)
3. Avanesov V: Item Response Theory: Basic Concepts and Provisions. Pedagogicheskiye izmereniya = Pedagogical Measurements. 2007; 2: 3–28. (In Russ.)
4. CHelyshkova M.B. Teoriya i praktika konstruirovaniya pedagogicheskikh testov: Ucheb. pos-

obiye = Theory and practice of constructing pedagogical tests: tutorial. Moscow: Logos; 2002. 432 p. (In Russ.)

5. Maslak A.A. Teoriya i praktika izmereniya latentnykh peremennykh v obrazovanii = Theory and practice of measuring latent variables in education. Moscow: Yurayt; 2016. 256 p. (In Russ.)

6. Popov A.P. Critical analysis of parametric models of Rush and Birnbaum. In: Materialy 4-y NMK «Innovatsionnyye metody i sredstva otsenki kachestva obrazovaniya» = Proceedings of the 4th NMC «Innovative methods and tools for assessing the quality of education.» Moscow: MGUP; 2006: 231-235. (In Russ.)

7. SHrayfel' I.S., Eliseyev I.N. Theoretical substantiation of a single iterative process of joint quantitative assessment of the difficulties of assignments and levels of student training. Sibirskiy zhurnal vychislitel'noy matematiki = Siberian Journal of Computational Mathematics. 2016; 19; 1: 107-123. (In Russ.)

8. Serbin V.I. Management of the learning process using a Markov model. Prikaspiyskiy zhurnal: upravleniye i vysokiye tekhnologii = Caspian Journal: Management and High Technologies. 2013; 4 (24): 29–35. (In Russ.)

9. Semov A.M., Semova M.A., KHlebnikov V.A. Edinyy iteratsionnyy protsess sovmestnoy kolichestvennoy otsenki trudnostey zadaniy i urovney podgotovlennosti uchastnikov testirovaniya. Trudy tsentra testirovaniya, vypusk 2 = A unified iterative process of joint quantitative assessment of the difficulties of tasks and levels of preparedness of testing participants. Works of the testing center, issue 2. Moscow: Prometheus Publishing House, 1999: 54–60. (In Russ.)

10. Hu F., Zidek J.V. The weighted likelihood. The Canadian Journal of Statistics. 2002; 30; 3: 347–371. DOI: 10.2307/3316141.

11. Simakhin V.A., Investigation of weighted maximum likelihood estimates. Vestnik Kurganskogo gosudarstvennogo universiteta. Seriya: Tekhnicheskiye nauki = Bulletin of Kurgan State University. Series: Engineering. 2011; 6: 72–77. (In Russ.)

12. Warm T.A. Weighted Likelihood Estimation of Ability in Item Response Theory. Psychometrika. 1989; 54: 427–450. DOI: 10.1007/BF02294627.

13. Linacre J.M. The Efficacy of Warm's Weighted Mean Likelihood Estimate (WLE) Correction to Maximum Likelihood Estimate (MLE) Bias. Rasch Measurement Transactions. 2009; 23; 1: 1188–1189

14. John M. Linacre. A User's Guide to WINSTEPS, MINISTEP. Rasch-Model Computer Programs; 2017. [Internet] URL: http://www.winsteps.com/winman/copyright.htm (Cited: 07.04.2019).

15. Tao, J., Shi, N., & Chang, H. Item-weighted likelihood method for ability estimation in tests composed of both dichotomous and polytomous items. Journal of Educational and Behavioral Statistics. 2012; 37; 2: 298-315. DOI: 10.3102/1076998610393969.

16. Kuchumov A. I. Elektronika i skhemotekhnika. Uchebnoye posobiye = Electronics and circuitry. Tutorial. Moscow: Helios ARV; 2011. 336 p. (In Russ.)

17. MatLab. Fminunc function [Internet] URL: https://www.mathworks.com/help/optim/ug/fminunc.html?s_tid=srchtitle (Cited: 07.04.2019).

18. Alekseyev M.V., Belov E.B., Kitayev N.P., Kuchumov A.I. ferences in assessing the levels of preparedness of extreme respondents using the classical theory of testing and the dichotomous Rasch model. Informatizatsiya obrazovaniya i nauki = Informatization of education and science. 2018. 2(38): 99-107. (In Russ.)

19. Wright B.D. Estimating measures for extreme scores. Rasch Measurement Transactions. 1998. 12 (2): 632-633.

20. Parameter MEAN in table 14.1 when using IWEIGHT. [Internet] URL: http://raschforum.boards.net/thread/796/parameter-mean-table-using-iweight (Cited: 07.04.2019).

Сведения об авторах

Евгений Борисович Белов
Заместитель председателя
Федеральное учебно-методическое объединение в системе высшего образования по укрупненной группе специальностей и направлений подготовки «Информационная безопасность», Москва, Россия
Эл. почта: umoib@yandex.ru

Михаил Витальевич Алексеев
К.ф.-м.н., эксперт
Федеральное учебно-методическое объединение в системе высшего образования по укрупненной группе специальностей и направлений подготовки «Информационная безопасность», Москва, Россия
Эл. почта: umoib@yandex.ru

Information about the authors

Evgeniy B. Belov
Vice-Chairman
Federal Educational and Methodical Association in the System of Higher Education on the Enlarged Group of Specialties and Directions of Training «Information Security», Moscow, Russia
E-mail: umoib@yandex.ru

Mikhail V. Alekseev
Cand. Sci. (Physics and Mathematics), Expert
Federal Educational and Methodical Association in the System of Higher Education on the Enlarged Group of Specialties and Directions of Training «Information Security», Moscow, Russia
E-mail: umoib@yandex.ru

Николай Павлович Китаев
К.ф.-м.н., эксперт
Федеральное учебно-методическое объединение в системе высшего образования по укрупненной группе специальностей и направлений подготовки «Информационная безопасность», Москва, Россия
Эл. почта: umoib@yandex.ru

Александр Иванович Кучумов
К.т.н., эксперт
Федеральное учебно-методическое объединение в системе высшего образования по укрупненной группе специальностей и направлений подготовки «Информационная безопасность», Москва, Россия
Эл. почта: umoib@yandex.ru

Nikolay P. Kitaev
Cand. Sci. (Physics and Mathematics), Expert
Federal Educational and Methodical Association in the System of Higher Education on the Enlarged Group of Specialties and Directions of Training «Information Security», Moscow, Russia
E-mail: umoib@yandex.ru

Aleksander I. Kuchumov
Cand. Sci. (Engineering), Expert
Federal Educational and Methodical Association in the System of Higher Education on the Enlarged Group of Specialties and Directions of Training «Information Security», Moscow, Russia
E-mail: umoib@yandex.ru