

АНАЛИЗ И ОЦЕНКА РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ НА ПРИМЕРЕ МОДЕЛИ РАША

А.Б. Баяскаланов¹

¹ Байкальский государственный университет, г. Иркутск

Математическую статистику используют как одно из эффективных средств познания объективных законов воспитания, обучения и развития. Данное средство принесет какой-либо результат, если математические формулы станут конкретным выражением качественных особенностей сформированности умений и знаний. Для этого требуется установить определенные соотношения между показателями и факторными признаками, характеризующими различные стороны подготовки обучающихся к тестированию. Знание функциональной зависимости между ними дает возможность спрогнозировать степень подготовки для каждого обучающегося.

В данной работе «Анализ и оценка результатов тестирования на примере модели Раша» представлены результаты применения однопараметрической модели Раша на практике. При оценке результатов итоговой контрольной работы по математике учащихся 11-х классов МБОУ г. Иркутска СОШ №2.

Цель исследования: выявить степень знаний обучающихся и трудность выполненных ими задач тестирования на основе математических моделей IRT. Полученные результаты свидетельствуют о целесообразности применения представленного подхода.

Ключевые слова: латентные переменные, модель Раша, логит, тестирование, метод.

ВВЕДЕНИЕ

Тестирование занимает все большее место в образовательном процессе. Цели тестирования распределяются отбором технологических процессов обработки итогов тестирования. Сведения, которые содержатся в ответах испытуемых, позволяет помимо выводов о результатах каждого обучающегося, делать вполне аргументированные заключения о группе испытуемых, о формах и методах преподавания предмета [1].

Совокупности методов анализа заданий подразделяются на экспертные и математические. После завершения тестирования, используя полученные эмпирические материалы применяют математические методы.

Экспертные методы дают оценку заданиям с точки зрения их формы и, не считая этого, выявляют, в какой степени они определяют как раз те знания, умения и навыки обучающихся, для которых разрабатывается тест. В начальной базе экспертных методов находится соотношение содержания и форм задач с требованиями, определенными в спецификации. В случае, если к заданию ни один из экспертов не выразил замечаний к определенному заданию, то оно включается в тест, в противном случае оно переформулируется или исключается из теста. Первый вариант тестирования получают после исследования и проведения экспертного анализа задач. С целью увеличения качества заданий теста, применяются математические методы. Для применения этих методов требуются результаты экспериментальной апробации теста. Такое тестирование получило название предварительного.

При использовании математических методов обработки информации возникает вопрос о точности, достоверности педагогических заключений, следующих из математических формул.

Успешность выполнения тестовых заданий определяет выполнение их как верно, так и неверно. Так как любой тест представляет из себя комплекс вопросов, то перед нами стоит задача сопоставления результатов выполнения различных заданий между собой.

При решении заданий закрытой формы перед обучающимися встают две взаимосвязанные задачи: определить верные ответы и найти ошибки в неверных ответах.

Различия в границах трудности заданий в зависимости от уровня подготовленности испытуемых явились причиной возникновения словосочетания «sample-dependant item characteristics» [2]. Также можно говорить и о зависимости параметров уровня подготовленности испытуемых от степени трудности вопросов тестирования (item-dependend person characteristics). Система модель Раша и математическая теория измерений IRT предстали из стремления преодолеть отмеченные зависимости.

В модели Раша на начальном этапе можно выделить два взаимосвязанных объекта измерений – уровни трудности вопросов и уровни подготовленности обучающихся [3]. В модели эти объекты участвуют синхронно, в рамках общего исследования. По этой причине такое измерение в большинстве случаев именуют совместно проводимым (joint measurement) [4].

ОСНОВНАЯ ЧАСТЬ

Рассмотрим применение однопараметрической модели Раша на практике на примере анализа результатов итоговой контрольной работы по математике учащихся 11-х классов МБОУ г. Иркутска СОШ №2.

Было проведено исследование того, как на основе математических моделей IRT можно выявить степень знаний обучающихся и трудность выполненных ими задач тестирования.

Алгоритм расчета уровня подготовленности учеников по школьной дисциплине «Математика» (параметр θ) и уровня трудности тестовых вопросов по математике, представленных в итоговой работе (параметр β), можно разделить на несколько этапов.

Этап 1. На начальном этапе подсчитываются доли верных и неверных ответов каждого обучающегося на все вопросы теста. Доля правильных ответов i -го обучающегося находится по формуле:

$$p_i = \frac{x_i}{n}, \quad (1)$$

где x_i – количество заданий, правильно выполненных i -м учеником; $i = 1, 2, \dots, n$, n – число заданий в тесте.

Доля неверных ответов $q_i = 1 - p_i$, $i = 1, 2, \dots, n$.

Доля верных ответов p_i , для изучаемой нами совокупности изменяется на промежутке $[0,29; 0,93]$.

Этап 2. На втором этапе выполняется предварительная оценка значений параметра, определяющего степень подготовки обучающихся тестируемой категории. Начальные значения параметра оцениваются в логитах. Логит уровня подготовки i -го обучающегося θ_i^0 находят по формуле

$$\theta_i^0 = \ln \frac{p_i}{q_i}, \quad (2)$$

где p_i , q_i – доли верных и неверных ответов i -го обучающегося на задания теста.

В соответствии с частями верных и неверных ответов на задания теста для каждого обучающегося было найдено персональное начальное значение логитов уровня подготовки обучающихся θ_i^0 . На теоретическом уровне исходные значения θ и β могут меняться в интервале $(-\infty; \infty)$. Но на практике имеем при $\theta_i - \beta_j < -5$ значения P_{ij} (P_{ij} – вероятность того, что ученик со способностью θ_i даст верный ответ на тест трудности β_j) близки к нулю. Идентичная пограничная ситуация наблюдается, когда $\theta_i - \beta_j > 5$, тогда $P_{ij} \rightarrow 1$. В данном тестировании минимальному начальному значению логиту уровня знаний соответствует уровень подготовки обучающегося, равный «-0,92» логита. Самому высокому значению уровня знаний соответствует значение «2,56» логита.

Этап 3. На третьем этапе подсчитываются доли верных p_j и неверных q_j ответов на каждое задание теста:

$$p_j = \frac{R_j}{n}, \quad q_j = 1 - p_j \quad (3)$$

где R_j – количество верных ответов на j -е задание теста, $o = 1, 2, \dots, n$, n – число заданий в тесте.

Доля верных ответов R_j для каждого из тестового набора заданий в изучаемой нами совокупности варьируется от 0,19 (для заданий, которые выполнили 4 обучающегося из совокупности) до 1,00 (для заданий, которые правильно выполнили все обучающиеся).

Этап 4. На четвертом этапе производится предварительная оценка значений параметра β , характеризующего трудность заданий теста. В качестве меры трудности заданий выбирается единица измерения, называемая логитом. Логит трудности j -го задания равен:

$$\beta_j^0 = \ln \frac{q_j}{p_j}, \quad (4)$$

где p_j , q_j – доли верных и неверных ответов на j -е задание теста.

Минимальное значение логита трудности задания равно «-3», максимальное значение «1,4» логита.

Этап 5. На пятом этапе подсчитываются средние значения логитов уровня подготовки и логитов трудности заданий теста. Среднее значение уровня знаний обучающихся для множества θ_j^0 подсчитывают по формуле:

$$\theta' = \frac{\sum_{i=1}^n \theta_i}{n}, \quad (5)$$

Среднее значение уровня подготовки изучаемой выборочной совокупности составило «0,60» логита, среднее значение уровня трудности задания данного теста составило «-0,85» логита.

Этап 6. После завершения пятого этапа оценки параметров θ и β будут показаны в интервальной шкале, но с разными значениями средних и различными стандартными отклонениями. На шестом этапе начальные значения логитов уровней подготовки и трудности вопросов тестирования переводятся в унитарную интервальную шкалу стандартных оценок. Стандартизация достигается через ряд специальных преобразований, для проведения которых вычисляются:

Дисперсия по множеству значений θ_i^0 ($i = 1, 2, \dots, n$)

$$V = \frac{\sum_{i=1}^n (\theta_i^0)^2 - n(\theta')^2}{n-1} = 0,64$$

Дисперсия по множеству значений β_j^0 ($j = 1, 2, \dots, n$)

$$U = \frac{\sum_{i=1}^n (\beta_j^0)^2 - n(\beta)^2}{n-1} = 2,77$$

Поправочные коэффициенты

$$X = \sqrt{\frac{1+U/2,89}{1-UV/8,35}} = 1,58$$

$$Y = \sqrt{\frac{1+V/2,89}{1-UV/8,35}}$$

Оценки параметров θ и β в единой интервальной шкале вычисляются по формулам:

$$\Theta_i = \beta' + X \times \theta_i^0; \quad (6)$$

$$\beta = \theta' + Y \times \beta_j^0. \quad (7)$$

Функции двух последних формул в изменении IRT трудно переоценить, не смотря на то, что на первый взгляд они содержат ограниченную практическую направленность. Эти формулы позволяют преодолеть ряд значительных недостатков традиционной теории тестов, так как с их помощью приобретаются объективные оценки параметров обучающихся и заданий, не зависящие друг от друга и выраженные в единой интервальной шкале.

Значения расчетов по данным показателям для нашей выборочной совокупности приведены в таблице 1.

Табл. 1. Расчеты показателей для вычисления поправочных коэффициентов

Показатель	Значение	Показатель	Значение
V	0.64	U	2.77
X	1.58	Y	1.25

Таким образом, на основе одной из математических моделей IRT (в данном случае модели Раша) были произведены типовые оценки уровня подготовки рассматриваемой группы испытуемых (табл.2) и эталонные оценки параметра трудности заданий итогового теста по математике для этих испытуемых (табл.3).

Табл. 2. Стандартные оценки уровня подготовки испытуемых

i	Вектор уровня подготовки испытуемых в θ_i^0 логитах	I	Вектор уровня подготовки испытуемых в θ_i^0 логитах
1	1,30	12	0,29
2	0,92	13	0,59
3	0,29	14	1,30
4	0,59	15	1,30
5	0,29	16	1,30
6	0,00	17	0,59
7	0,92	18	1,30
8	-0,59	19	-0,92
9	0,29	20	0,92
10	0,29	21	-0,92
11	2,56		

$$P_j(\theta) = \frac{e^{1,7a_j(0-\beta_j)}}{1+e^{1,7a_j(0-\beta_j)}}$$

На основе этих значений строятся характеристические кривые заданий теста. Анализ их взаимного расположения позволяет наметить пути дальнейшего совершенствования теста и сформировать систему заданий, наиболее эффективных для оценки уровня подготовки каждого испытуемого выборки.

Табл. 3.

j	Вектор уровня трудности заданий β_j , логитов	j	Вектор уровня трудности заданий β_j , логитов
1	-3	8	-0,7
2	-2,99	9	-1,8
3	-0,5	10	-3
4	-3	11	-0,1
5	-1,2	12	1,2
6	1,4	13	1,4
7	-0,9	14	1,2

На рис.2 представлены характеристические кривые 14 заданий итогового теста по математике, построенные на значениях таблицы 4. Для заданий с одинаковыми функциями $P(j)$ характеристическая кривая совпадает.

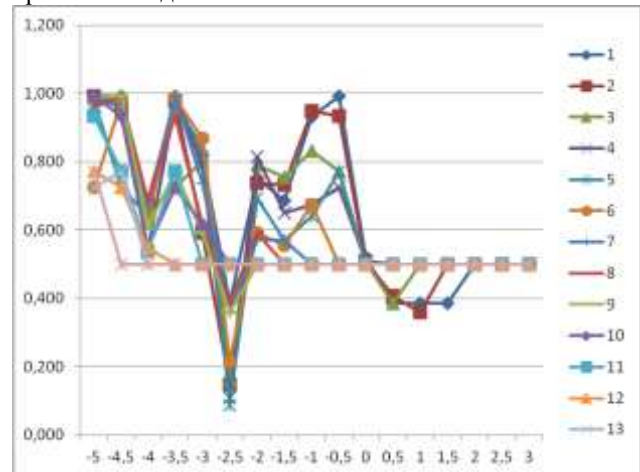


Рис. 1.

Операция улучшения заданий теста начинается с удаления ненужных заданий, нарушающих нормальный характер распределения значений β . При этом следует обратить внимание на случаи совпадения характеристических кривых и исключить ненужные задания, которые ничего не дают для теста как совокупности работающих заданий возрастающей трудности.

Это означает, что при использовании данного теста в перспективе одно из этих заданий необходимо снять, так как оно не дает ни единой ценности для оценивания уровня знаний обучающихся. В тест необходимо ввести задания, равные по трудности выделенным интервалам на оси латентной переменной θ . В идеале характеристические кривые должны запол-

нить более или менее равномерно практически весь интервал (-5; +3) шкалы логитов.

ЗАКЛЮЧЕНИЕ

Таким образом, использование данного подхода позволит обосновано выявить степень знаний обучающихся и трудность выполненных ими задач тестирования.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Тесленко, В.И. Методика анализа и оценка результатов тестирования [Текст] / В.И. Тесленко // Вестник Красноярского государственного педагогического университета им. В.П. Астафьева. – 2006. – Т. 1. – С. 78–95.

2. Аванесов, В.С. Понятия и методы математической теории педагогических измерений (Item Response Theory, IRT). Статья третья [Текст] / В.С. Аванесов // Педагогические Измерения. – №4. – 2009. – С. 5.

3. Аванесов, В.С. Метрическая система Георга Раша [Текст] / В.С. Аванесов // Педагогические Измерения. – №2. – 2010. – С. 57–80.

4. Чельшкова, М.Б. Теория и практика конструирования педагогических тестов [Текст] / М.Б. Чельшкова. – М.: Логос, 2002. – 432 с.

Баяскаланов Алексей Баиртуевич – аспирант, ФГБОУ ВО «Байкальский государственный университет», тел. 89248208417, e-mail: alexrinch@mail.ru.

Табл. 4. Значения функции $P_j(\theta)$ на основе модели Раша

	-5	-4,5	-4	-3,5	-3	-2,5	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5	3
1	0,992	0,992	0,625	0,992	0,818	0,385	0,752	0,686	0,931	0,992	0,521	0,385	0,385	0,385	0,500	0,500	0,500
2	0,992	0,959	0,686	0,980	0,593	0,148	0,738	0,732	0,947	0,935	0,510	0,407	0,358	0,500	0,500	0,500	0,500
3	0,959	0,992	0,652	0,725	0,803	0,163	0,791	0,754	0,831	0,770	0,508	0,385	0,500	0,500	0,500	0,500	0,500
4	0,992	0,980	0,539	0,974	0,788	0,109	0,814	0,649	0,673	0,725	0,510	0,500	0,500	0,500	0,500	0,500	0,500
5	0,980	0,725	0,643	0,967	0,844	0,088	0,693	0,569	0,641	0,770	0,500	0,500	0,500	0,500	0,500	0,500	0,500
6	0,725	0,974	0,634	0,987	0,867	0,217	0,591	0,556	0,673	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
7	0,974	0,967	0,669	0,992	0,737	0,358	0,573	0,569	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
8	0,967	0,987	0,686	0,935	0,615	0,385	0,591	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
9	0,987	0,992	0,606	0,770	0,593	0,358	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
10	0,992	0,935	0,549	0,725	0,615	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
11	0,935	0,770	0,539	0,770	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
12	0,770	0,725	0,549	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
13	0,725	0,770	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
14	0,770	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500

ANALYSIS AND EVALUATION OF TEST RESULTS ON THE EXAMPLE OF THE MODEL RASHA

A.B. Bayaskalanov

Baikal State University, Irkutsk

Abstract – Mathematical statistics is used as one of the effective means of learning the objective laws of education, training and development. This tool will bring any result if the mathematical formulas become a concrete expression of the qualitative features of the formation of skills and knowledge. To do this, it is necessary to establish certain ratios between indicators and factor signs characterizing various aspects of preparing students for testing. Knowledge of the functional relationship between them makes it possible to predict the degree of preparation for each student.

In this paper, “Analysis and Evaluation of Test Results on the Example of the Rush Model”, the results of applying the one-parameter Rush model in practice are presented. when evaluating the results of the final test work in mathematics of students of 11th grades of MBOU of the city of Irkutsk Secondary School No. 2.

Objective: to identify the degree of knowledge of students and the difficulty of their testing tasks based on mathematical models of IRT. The results indicate the feasibility of applying the presented approach.

Index terms: testing, method, analysis, evaluation, educational process.

REFERENCES

1. Teslenko, V.I., “Methods of analysis and evaluation of test results,” *Investor of the Krasnoyarsk State Pedagogical University*. V.P. Astafieva, V. 1, p. 78–95, 2006.
2. Avanesov, V.S., “Concepts and methods of the mathematical theory of pedagogical measurements (Item Response Theory, IRT). Article Three,” *Pedagogical Measurements*, №4, p. 5, 2009.
3. Avanesov, V.S., “George Rush metric system,” *Pedagogical Measurements*, №2, pp. 57–80, 2010.
4. Chelyshkova, M.B., *Theory and practice of constructing pedagogical tests*. M.: Logos, 2002.

Bayaskalanov Alexey Bairtuyevich – Postgraduate Student, Baikal State University, tel. 89248208417, e-mail: alexrinch@mail.ru.