

# **Применение элементов IRT-теории для разработки автоматизированных систем тестирования**

**М.В. Васюков, Ю.С. Бахтин**

Рязанский государственный радиотехнический университет,

Россия, Рязань,

[lmaxter@gmail.com](mailto:lmaxter@gmail.com), [bekhtin@yandex.ru](mailto:bekhtin@yandex.ru)

Место публикации -  
«Международный научно-технический форум СТНО-2018.  
Сб. трудов, Т. 6. Стр. 42-45.

## *Аннотация*

Рассматривается применение элементов IRT-теории для  
разработки автоматизированных систем тестирования.

*Ключевые слова.* IRT-теория, автоматизированная система  
тестирования.

## **Application of IRT Theory elements for development of automated test systems**

**M.V. Vasyukov, Y.S. Bekhtin**

*Ryazan State Radio Engineering University,*

*Russia, Ryazan,*

[lmaxter@gmail.com](mailto:lmaxter@gmail.com), [yuri.bekhtin@yandex.ru](mailto:yuri.bekhtin@yandex.ru)

## *Abstract*

The application of elements of IRT-theory for the development of  
automated testing systems is considered.

*Keywords.* IRT-theory, automated testing system.

В настоящее время остро ощущается потребность в  
автоматизированных системах тестирования знаний обучаемых, которые  
позволяют, во-первых, обеспечить принцип саморегулирования  
(самоконтроля) в обучении, во-вторых, организовать образовательный

процесс на основе реальных и имитационных стимулирующих проблемных ситуациях, и, наконец, в-третьих, обеспечить самостоятельную познавательную деятельность обучаемого [1]. Эффективное использование подобных систем невозможно без наполнения их качественным, разноуровневым тестовым материалом. Как показывает практика предыдущего десятилетия, использование в качестве наполнителя системы обычного лекционного материала превращает тестирование в рутинную процедуру, которая не только не повышает индивидуализацию обучения, но и не даёт возможности объективно оценить достижения обучаемых. Реализация такой возможности позволит объективно измерять уровень обученности студентов, магистрантов, что в результате предоставляет преподавателям возможность управления процессом обучения и проведения коррекции. При этом особое внимание должно уделяться степени соответствия достижений обучаемого определенному образовательному стандарту и динамике развития достижений обучаемого на каждом этапе процесса обучения [1,2].

Для решения поставленной задачи известно не так много подходов. Один из них базируется на классической теории тестирования [1,2]. Здесь тестовый балл рассматривается как сумма двух компонент: истинного значения оценки и случайной ошибки. Два схожих параллельных теста будут давать одну и ту же истинную оценку, хотя реальные баллы респондентов (обучаемых) не совпадут из-за непредсказуемой случайной ошибки. Тем не менее, величина данной ошибки может быть сведена к нулю при достаточном увеличении числа заданий (длины теста), что не всегда приемлемо в условиях временных ограничений образовательного процесса. Другой недостаток в том, что если имеется два идентичных теста, один из которых состоит из более простых заданий, то результаты использования тестов не являются сравнимыми, так как кривая распределения итоговых баллов будет иметь различную форму даже в одинаковых группах. Классическая теория не дает иного метода для сравнения различных тестов, кроме метода контрольных групп.

Другой подход в теории тестирования основан на использовании

статистической теории тестирования IRT (Item Response Theory). Построение теории IRT [1-4] основано на предположении о существовании функциональной связи между латентными параметрами обучаемых и наблюдаемыми результатами выполнения заданий теста. Главный постулат здесь в том, что латентные параметры обучаемых взаимодействуют с заданиями в процессе тестирования, а это, в свою очередь, приводит к получению наблюдаемых результатов выполнения теста. На практике всегда ставится обратная задача: по ответам обучаемых на задания теста требуется найти оценки значения латентного параметра  $\theta_i$  ( $i = 1, 2, \dots, N$ ), определяющие уровень подготовки  $N$  обучаемых, и латентного параметра  $\beta_j$  ( $j = 1, 2, \dots, n$ ), равные оценкам трудности  $n$  заданий теста [2, 4].

Для решения этой задачи была предложена математическая модель связи между латентными параметрами и наблюдаемыми результатами тестирования [1, 3, 4], которая содержала разность  $\theta - \beta$  при условии, что параметры  $\theta$  и  $\beta$  оцениваются в одной и той же шкале. В качестве такой единой шкалы введена интервальная шкала логитов [3,4].

Если значение параметра  $\theta_i$ , является положением  $i$ -го обучаемого на шкале логитов, а значение  $\beta_j$  - положением  $j$ -го задания на той же шкале, то абсолютная величина разности  $|\theta_i - \beta_j|$  показывает расстояние, на котором находится обучаемый с уровнем подготовки  $\theta_i$  от задания с трудностью  $\beta_j$ . Если данная разность отрицательна и относительно велика по модулю, то тестовое задание бесполезно для измерения уровня подготовленности  $i$ -го обучаемого (он не сможет успешно пройти тест). С другой стороны, если разности  $|\theta_i - \beta_j|$  принимают относительно большие положительные значения, то нет смысла использовать такой тест ни для процесса контроля, ни для обучения  $i$ -го испытуемого, поскольку знания освоены обучаемым и он справится с ними успешно при выполнении теста. Таким образом, наименьшую ошибку измерения уровня подготовленности обучаемых обеспечивают задания, трудность которых подобрана по критерию  $\theta = \beta$ .

Возникает задача в подборе некоторой функции  $P_j(\theta) = f(\theta - \beta_j)$ , описывающей вероятность правильного ответа обучаемых любого уровня подготовленности на задание определённого уровня трудности под номером  $j$ . В настоящее время, в теории рассматриваются одно-, двух- и трехпараметрическая модели [4, 5]. Трехпараметрическая модель отличается введением в рассмотрение вероятности угадывания правильного ответа обучаемым, определяемая, в общем случае, как  $c_j = m/M$ , где  $M$  - число ответов в тестовом задании,  $m$  - число правильных ответов [5]:

$$P_j\{X_{ij} = 1 | \beta_j, a_j, c_j\} = c_j + (1 - c_j) \exp a_j(\theta - \beta_j) / (1 + \exp a_j(\theta - \beta_j)) \quad (1)$$

здесь параметр  $a_j$  даёт информацию о тестовом задании с точки зрения оценки его дифференцирующей способности, то есть способности различать уровни подготовленности обучаемых.

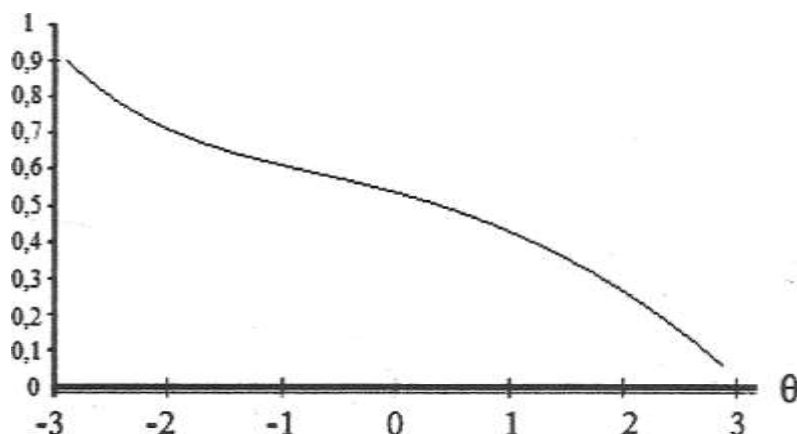
Тем не менее, модель (1) не учитывает всех возможных ситуаций при прохождении теста обучаемыми, а именно, их психологическую составляющую. В работе предлагается усложнить модель (1) введением параметра  $d_j$ , который учитывал бы наличие т.н. «психологического барьера» при решении ряда тестовых задач. Например, хорошо известны случаи, когда обучаемый при ответе на вопросы искусственно возводит перед собой данный «психологический барьер», то есть он не может выйти из ограниченного пространства решений, или, по-другому, посмотреть на решение задачи из другого пространства решений. То или иное пространство решений формируется у проходящего тестирование обучаемого в процессе ответа на серию предшествующих заданий. С другой стороны, известны случаи, когда отдельные обучаемые относительно легко справляются со сложными заданиями, а на относительно простых заданиях дают неправильные ответы. В данной работе вышеуказанный момент предлагается учитывать в модели (1) следующим образом:

$$P_j\{X_{ij} = 1 | \beta_j, a_j, c_j\} = c_j + (1 - c_j) \exp a_j(\theta - \beta_j) / (1 + \exp a_j(\theta - \beta_j)) + d_j \exp a_j\{\theta - \beta_j\} \quad (2)$$

где  $d_j$  есть вероятность появления «психологического барьера» в данном тесте.

В работе были проведены исследования модели (2) при разных значениях параметра  $d_j$ . Исследования показали, что при относительно больших значениях параметра  $d_j$  наблюдается изменение вида кривой зависимости  $P(\theta)$ , подобно изображенному на рисунке примеру, полученному при  $d_j = 0,75$ . При этом остальные параметры модели (2) определяли задание с параметрами трудности  $b = 1,5$ , крутизны кривой  $a_j = 1,3$  и вероятностью правильного угадывания  $c_j = 0,2$ . Из рисунка видно, что кривая зависимости  $P(\theta)$  по внешнему виду совпадает с кривыми для графиков, направленными от «сверху-слева» к «вниз-направо». Это соответствует ситуациям, обусловленным нарушением содержательных, формальных, организационных и этических требований. «Правильные» графики имеют противоположный тренд, то есть «снизу-вверх» [1, 3].

Очевидно, данный факт указывает на совершенно непригодный и абсурдный для проведения тестирования случай: чем выше уровень подготовленности обучаемых, тем ниже оказывается вероятность правильного ответа. Приведенный ниже рисунок отражает случай абсолютно не тестового задания и означает, что такое задание ни при каких обстоятельствах не должно быть применено. Похожий график зависимости  $P(\theta)$  могут иметь задания, у которых коэффициент их корреляции с суммой баллов проектируемого теста является отрицательным и близким по модулю к единице величиной.



Пример графика оценки задания с отрицательной характеристикой тестируемости

Таким образом, результаты исследований позволяют сделать вывод о том, что возможно совершенствование тестовых заданий для избегания возникновения «психологических барьеров». В докладе также изложена методика расчета параметра  $d_j$  на ряде примеров тестовых заданий по дисциплинам подготовки магистров по направлению 27.04.04 «Управление в технических системах» в рамках образовательной программы «Обработка сигналов и изображений в информационно-управляющих системах».

1. Аванесов В.С. Основы теории педагогических заданий. Педагогические Измерения №2, 2006 г. С. 26-62.
2. Дюк В.А. Компьютерная психодиагностика,- СПб., издательство «Братство», 1994. - 364 с.
3. Baker, Frank *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. 2001.
4. Haley D.C. Cited in: Hambleton R.K. Swaminathan H. *Item Response Theory: Principles and Applications*, Boston 1985. - 327 p.
5. Birnbaum A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability / In: F.M. Lord and M.R. Novick. *Statistical Theories of Mental Test Scores*. Reading, Mass: Addison - Wesly, 1968. - 568p.