

МОДЕЛЬ АДАПТИВНОГО ТЕСТИРОВАНИЯ КАК СПОСОБ РЕАЛИЗАЦИИ ИНДИВИДУАЛЬНОГО ПОДХОДА

Маслова О.В.
Сиб ГАУ, г. Красноярск

В настоящее время в образовательном процессе все более актуальными становятся такие способы обучения, которые максимально приближены к внутренним потребностям обучаемых, к их привычному способу взаимодействия с окружающим миром. Студенты современного вуза обладают хорошей подготовкой в области компьютерной грамотности, поэтому следует учитывать этот факт в условиях все возрастающей компьютеризации большинства сфер человеческой деятельности, распространения информационных и телекоммуникационных технологий. Использование таких технологий и различных технических средств в обучении позволяет повысить его интенсификацию, оптимизацию и реструктурирование для значительной экономии и более рационального использования учебного времени. При этом меняется и сам процесс обучения: он становится более разнообразным, живым, ярким; повышается мотивация и вырастает заинтересованность обучаемых в изучении предмета уже по причине того, что современное поколение студентов знакомится с компьютером и компьютерными технологиями достаточно рано. Поэтому актуальной задачей является разработка учебно-методических материалов, планирование проведения занятий и организация учебной и внеучебной деятельности обучаемых с учетом этой «особенной» характеристики сегодняшнего студента.

В современной практике преподавания такая компьютерная технология как тестирование, проводимая с определенной периодичностью и выполняющая как обучающую, так и контролирующую функцию, является одной из наиболее часто применяемых. Но, к сожалению, этот метод имеет существенный недостаток – потеря индивидуального подхода к каждому обучаемому, когда преподаватель не имеет возможности объективно оценить степень усвоения проверяемого материала каждым студентом [3].

Модели адаптивного тестирования, в которых сложность заданий меняется в зависимости от правильности ответов испытуемого, являются прогрессивными и наиболее интересными для исследования, они, частично, могут содействовать устранению этой проблемы.

В данной статье приведена модель адаптивного тестирования, использующая аппарат IRT (Item Response Theory, в русской литературе известная как математическая теория педагогических измерений) [4, 5]. Данная модель может помочь разработчикам систем тестирования и преподавателям, активно использующим технологии тестирования в обучении [6].

Для оценки трудности заданий и уровня знаний испытуемых используется аппарат IRT, в частности, однопараметрическая модель Раша

(Rasch Measurement Model) [7, 8]. Модель ИРТ имеет ряд преимуществ перед классической теорией тестов [9]. Во-первых, модель использует количественные методы оценки знаний испытуемых и трудности заданий. Во-вторых, модель является линейной, что позволяет использовать аппарат математической статистики для анализа результатов. В-третьих, оценка трудности заданий не зависит от выборки испытуемых, равно как оценка уровня подготовки испытуемых не зависит от выборки тестовых заданий. В-четвертых, существует возможность получить оценку уровня знаний испытуемых на выборке тестовых заданий, а не на полном множестве, что позволяет использовать ИРТ в адаптивном тестировании.

Основными переменными однопараметрической модели Раша являются уровень знаний испытуемого Θ и трудность задания β . Аргументом функции успеха испытуемого при ответе на задание теста является разность $\Theta - \beta$. Функция, связывающая успех испытуемого с уровнем его подготовленности и трудностью задания, в соответствии с моделью Раша, имеет следующий вид:

$$P(\Theta|\beta) = P(\beta|\Theta) = \frac{e^{(\Theta-\beta)}}{1 + e^{(\Theta-\beta)}} \quad (1)$$

На рисунке 1 слева представлен вид функции $P(\Theta)$ при значениях трудности задания β : -2, 0 и 2, а справа – вид функции $P(\beta)$ при соответствующих значениях β [10].

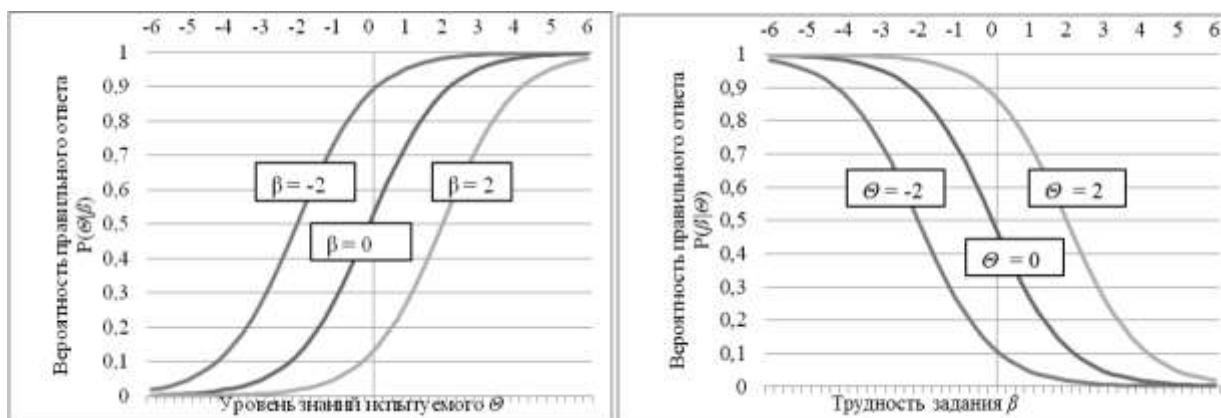


Рис. 1. Характеристические кривые заданий (*Item Characteristic Curve*) и характеристические кривые испытуемых (*Person Characteristic Curve*)

Чем выше уровень знаний испытуемого Θ , тем выше вероятность успеха в том или ином задании. В точках $\Theta = \beta$ вероятность правильного ответа равна 0,5. То есть, если трудность задания равна уровню знаний испытуемого, то он с равной вероятностью может справиться или не справиться с этим заданием [11].

На рисунке 2 представлен вид функции $P(\beta)$. Чем выше уровень знаний испытуемого, тем выше вероятность правильного ответа на задание.

Далее рассмотрим алгоритм получения значений Θ и β [12]. Пусть Q – индивидуальный балл испытуемого (количество правильных ответов), M – общее количество ответов, данных испытуемым, R – количество правильных ответов данных группой испытуемых на задание теста, N – общее количество ответов, данных на задание теста. Тогда начальные значения уровня знаний испытуемых и уровня трудности заданий вычисляются по следующим формулам:

$$\Theta_0 = \ln \frac{Q}{M - Q} \quad (2)$$

$$\beta_0 = \ln \frac{R}{N - R} \quad (3)$$

Затем значения Θ_0 и β_0 переводятся в единую интервальную шкалу стандартных оценок. Для этого необходимо вычислить среднее и несмещенную оценку дисперсии по множествам Θ_0 и β_0 . Пусть V – несмещенная оценка дисперсии по множеству Θ_0 , а U – несмещенная оценка дисперсии по множеству β_0 . $\bar{\beta}$ и $\bar{\Theta}$ – математическое ожидание соответственно Θ и β поправочные коэффициенты определяем по следующим формулам [13]:

$$C_Q = \sqrt{\frac{1 + \frac{U}{2,89}}{1 - \frac{UV}{8,35}}} \quad (4)$$

$$C_\beta = \sqrt{\frac{1 + \frac{V}{2,89}}{1 - \frac{UV}{8,35}}} \quad (5)$$

Θ и β в единой интервальной шкале вычисляются по следующим формулам:

$$\Theta = \bar{\beta} + C_Q * \Theta_0 \quad (6)$$

$$\beta = \bar{\Theta} + C_\beta * \beta_0 \quad (7)$$

Рассмотрим пример вычисления параметров Θ и β для группы испытуемых из восьми человек ($N = 8$) по результатам ответа на 10 заданий теста ($M = 10$). Исходная матрица ответов представлена в бинарном виде (таблица 1), т.е. 1 соответствует успешному выполнению задания испытуемым, 0 – неуспешному.

Таблица 1 – Матрица вычислений значений Θ и β

№ Фамилия	1	2	3	4	5	6	7	8	9	10	Q	Θ_0	Θ
1. Воробьев	1	1	0	0	1	0	0	0	1	1	5	0	0,051
2. Голубев	0	0	1	0	1	1	0	0	0	1	4	- 0,405	-0,387
3. Гусев	0	1	1	0	1	1	1	0	0	1	6	0,405	0,489
4. Журавлев	1	0	1	1	0	0	0	0	1	0	4	- 0,405	-0,387
5. Куликов	1	1	0	1	0	1	1	1	0	0	6	0,405	0,489
6. Лебедев	1	0	0	1	0	0	0	1	1	0	4	- 0,405	-0,387
7. Орлов	1	0	0	0	1	1	0	0	0	1	4	- 0,405	-0,387
8. Соколов	1	0	1	1	1	1	1	0	1	1	8	1,386	1,548
R	6	3	4	4	5	5	3	2	4	5			
β_0	1,09 9	0,51 1	0	0	0,51 1	0,51 1	0,51 1	1,09 9	0	0,5 11			
β	1,25 8	0,479 0,07 2	0,07 2	0,62 3	0,62 3	0,47 9	1,11 4	0,07 2	0,62 3				

Можно выделить ряд утверждений о значениях Θ и β :

– при $\Theta > \beta$ вероятность того, что испытуемый справится с текущим заданием выше $\frac{1}{2}$ (чем больше разница между Θ и β , тем вероятность ближе к

1);

– при $\Theta < \beta$ вероятность того, что испытуемый справится с текущим заданием ниже $\frac{1}{2}$ (чем больше разница между Θ и β , тем вероятность ближе к 0);

– при $\Theta = \beta$ вероятность того, что испытуемый справится с текущим заданием равна 0,5.

Для применения однопараметрической модели Раша при адаптивном тестировании необходимо произвести начальное обучение системы. Для этого выбирается группа испытуемых с разным уровнем знаний в исследуемой области, предположительно 10 человек, для достижения достаточного уровня дифференцирования тестовых заданий по трудности ($M=10$). Рассмотрим основные этапы процедуры начального обучения системы.

1. Выбирается шкала оценки знаний испытуемого. Так как модель тестирования разрабатывается для проверки знаний студента по иностранному языку, то рекомендуется пятибалльная шкала оценки (неудовлетворительно, удовлетворительно, хорошо, отлично). Таким образом, получаем 4 группы заданий теста по уровню трудности.

2. После выбора шкалы оценки составляется база вопросов по исследуемой области, и испытуемым предлагается пройти тест. По результатам прохождения вычисляется оценка трудности каждого задания β_j .

3. Рекомендуется произвести корректировку базы вопросов, а именно:

– исключить задания, с которыми не справился ни один испытуемый, и задания, с которыми справились все испытуемые (эти задания несут мало информации об уровне знаний испытуемых в исследуемой области);

– при достаточно сильном отклонении среднего значения уровня трудности заданий $\bar{\beta}$ от 0 необходимо добавить вопросы в базы, при $\bar{\beta} \gg 0$ необходимо добавить «трудные» вопросы, а при $\bar{\beta} \ll 0$ «легкие» вопросы, при $\bar{\beta}$ близком к 0 набор тестовых заданий считается сбалансированным.

4. Затем, используя полученную оценку трудности заданий теста β , необходимо разделить задания на 4 группы сложности. При этом деление производится одним из двух способов:

– интервал делится на 4 равных подинтервала;

– интервал делится таким образом, чтобы в каждой группе оказалось равное количество вопросов.

Каждая группа сложности характеризуется минимальным и максимальным значением трудности задания в пределах данной группы.

Если используется первый способ деления, желательно, чтобы в каждой группе было достаточное количество вопросов для оценки знаний испытуемых.

Рассмотрим теперь алгоритм тестирования, который предлагается испытуемым. Введем обозначения: X – количество выданных заданий; Y – количество правильных ответов, данных испытуемым; Θ – текущий уровень подготовки испытуемого; i – номер текущего шага; R – оценка испытуемого (неудовлетворительно, удовлетворительно, хорошо или отлично); G – группа сложности, g – вопрос теста. Также для описания алгоритма необходимо ввести понятие критического количества вопросов $X_{\text{крит}}$ – количество вопросов для группы сложности, на которое испытуемый должен ответить (без изменения группы сложности), чтобы выставить ему оценку R . Значение $X_{\text{крит}}$ зависит от общего числа вопросов в базе и в каждой группе сложности и от времени тестирования; на начальном этапе определяется преподавателем. Во избежание постоянных смен уровней сложности при чередовании верных и ошибочных ответах тестируемого введём переменную η – коэффициент верных ответов для сохранения уровня сложности (изменяется от 0 до 1), также определяется преподавателем. Рекомендуются на начальном этапе использовать значение $\eta = 1/3$.

Исходные данные алгоритма: $i = 0$; $\Theta_0 = \Theta^-$; $Y_0 = 0$; $X_0 = 0$. Исходная группа сложности $G_0 = G(\Theta_0)$ – множество вопросов, входящих в группу сложности, содержащую значение Θ_0 .

Процедура оценки уровня подготовки испытуемых (рисунок 2):

1. Выбирается задание из текущей группы сложности вопросов $g_i \in G_i$ и предлагается для ответа испытуемому.

2. При успешном ответе испытуемого, увеличивается количество правильных ответов.

3. Увеличивается количество выданных заданий, выданное задание исключается из набора заданий для данного испытуемого.

4. Вычисляется новый уровень подготовки испытуемого Θ_i по модели Раша ранее описанным способом.

5. Вычисляется группа сложности в зависимости от значения Θ_i . Принадлежность Θ_i к группе сложности определяется путем сравнения значения Θ_i с максимальным и минимальным значениями трудности заданий для каждой группы сложности. Если $\Theta_i \geq \beta_{\text{мин}}(G_k)$ и $\Theta_i < \beta_{\text{макс}}(G_k)$, то $G_i = G_k$ (k – индекс перебора по всем группам сложности, k изменяется от 1 до 4).

6. Если группа сложности изменилась, количество выданных заданий превышает η -ую часть от $X_{\text{крит}}$ и новая группа сложности больше максимальной

(оценка 5) или меньше минимальной (оценка 2), то выставляется оценка в соответствии с группой сложности на предыдущем шаге (5 или 2).

7. Если группа сложности изменилась, количество выданных заданий превышает η -ую часть от $X_{\text{крит}}$ и новая группа сложности не больше максимальной (оценка 5) и не меньше минимальной (оценка 2), то количество выданных вопросов и правильных ответов обнуляется и происходит возврат к шагу 1.

8. Если группа сложности изменилась, а количество выданных заданий не превышает η -ую часть от $X_{\text{крит}}$, то группа сложности возвращается к предыдущему значению (изменения не принимаются).

9. Если группа сложности не изменилась, не превышено количество критических вопросов и текущая группа сложности содержит еще хотя бы 1 вопрос, то происходит возврат к шагу 1.

10. Если группа сложности не изменилась, но при этом превышено количество критических вопросов или все вопросы текущей группы сложности уже были выданы испытуемому ранее, то выставляется оценка в соответствии с текущей группой сложности.

После ответа испытуемым на текущий вопрос результаты должны сохраняться в базе для вычисления Θ_i .

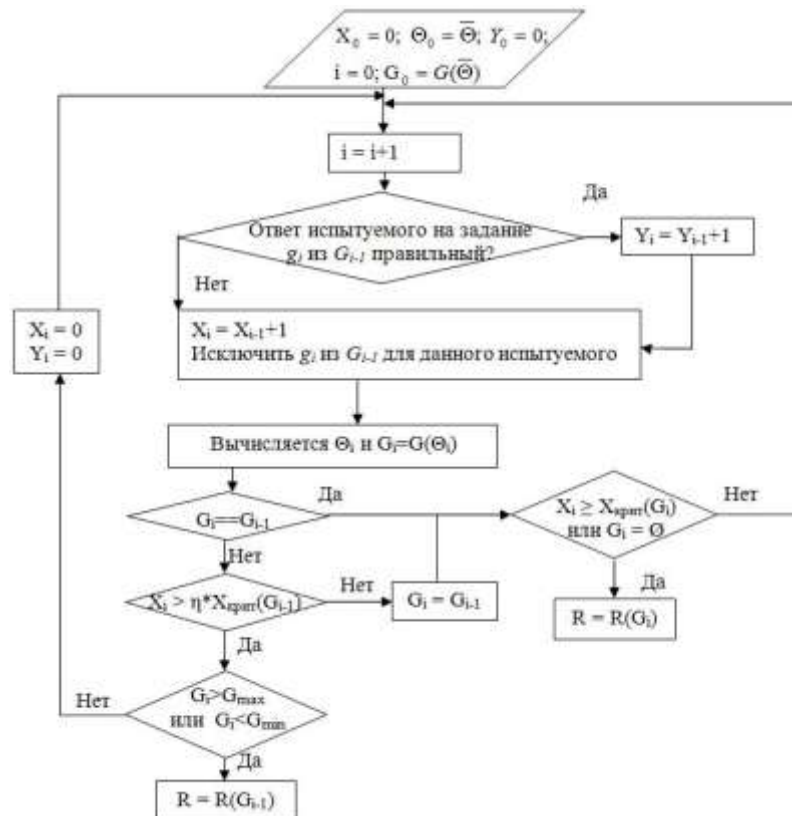


Рис. 2. Блок-схема процедуры тестирования

Модель Раша также позволяет изменять уровень сложности заданий в зависимости от результатов прохождения теста. Рекомендуется сохранять результаты ответов каждого испытуемого, и после того, как вся группа пройдет тест, вычислить новые уровни трудности заданий β для всех вопросов, которые были выданы одному или нескольким испытуемым.

После корректировки значений β необходимо произвести расчет среднего уровня подготовки испытуемых заново. Затем заново формируются группы сложности. Возможно, что некоторые задания изменят группу сложности. Также существует возможность добавления в базу новых вопросов, желательно небольшими группами. В противном случае необходимо будет выполнить процедуру переобучения системы для балансировки уровня сложности заданий. При единичных добавлениях новым добавляемым вопросам присваивается средний уровень трудности.

Таким образом, в работе предложена процедура адаптивного тестирования, основанная на модели IRT. Данная модель направлена на решение проблемы потери «индивидуального подхода» при процедуре автоматизированного тестирования. Эта проблема снижает эффективность применения метода тестирования как способа обучения и контроля знаний и, в качестве ее возможного решения, авторами предлагается интеллектуальный подход к формированию тестовых заданий. Общее число вопросов при оценке знаний студентов в целом будет уменьшаться, испытуемому будут предлагаться задания с понижающейся или повышающейся трудностью, в зависимости от данных ответов. При этом если ответы испытуемого нестабильны, то число заданных вопросов в пределах одного уровня сложности будет выше среднего, что позволит получить более точное представление об уровне его подготовленности.

Данная модель может помочь создателям систем тестирования и преподавателям, активно использующим технологии тестирования в обучении, усовершенствовать и повысить эффективность проведения тестового контроля знаний [14, 15].

Список литературы

1. Девтерова З.Р., Новые информационные технологии в преподавании иностранного языка в ВУЗе // Вестник Адыгейского государственного университета, № 4, 2006. С.157-159.

2. Шерматов Ш.М., Гафорова М.С., Использование компьютерных технологий и методов обучения в преподавании гуманитарных дисциплин // Вестник ТГУПБП, № 3 (51), 2012. С.262-267.

3. Маслова О.В., Медведева Т.Е., Пятков А.Г., Трудности использования online-тестирования как формы контроля знаний // сборник «Тенденции, наработки, инновации, практика в науке», Гданьск, 2014. – С.124-128.

4. Аванесов В.С. *Item Response Theory: Основные понятия и положения // Педагогические измерения. № 2, 2007. – С. 3-28.*
5. Аванесов В.С., *Основные понятия и положения математической теории измерений (Item Response Theory) // Педагогические измерения [электронный ресурс]. URL: <http://testolog.narod.ru/Theory60.html> (01.09.14).*
6. *Использование модели Раша в пересчете баллов ЕГЭ // Сайт по менеджменту [электронный ресурс]. URL: <http://baguzin.ru/wp/?p=3370> (27.08.14).*
7. Ким В.С. *Тестирование учебных достижений // монография. – Уссурийск: Издательство УГПИ, 2007. – 214 с.*
8. *What is Rasch Analysis // Official site of Rasch analysis method [электронный ресурс]. URL: <http://www.rasch-analysis.com/rasch-analysis.htm> (27.08.14).*
9. Nordin A.R., Ahmad Z.K, Lei M.T., *Examining Quality of Mathematics Test Items Using Rasch Model: Preliminary Analysis // International Conference on Education & Educational Psychology (ICEEPSY 2012), vol. 69, 2012. – С.2205-2214.*
10. Malcolm J.R., *Estimating Item Characteristic Curves // Applied psychological measurement vol. 3. № 3, 1979. – С.371-385.*
11. Dodeen, H. (2004). *The relationship between item parameters and item fit // Journal of Educational Measurement, vol. 41. – С.261-270.*
12. B. N. Wright and G. N. Masters, *Rating Scale Analysis: Rasch Measurement, MESA Press, Chicago, Ill, USA, 1982. – 206 с.*
13. Чельшикова М.Б. *Теория и практика конструирования педагогических тестов // учебное пособие. – М.: Логос, 2002. – 432 с.*
14. Wagner-Menghin M., Preusche I., Schmidts M., *The Effects of Reusing Written Test Items: A Study Using the Rasch Model // International Scholarly Research Notices, vol. 2013, 2013. – С.7-14.*
15. Lukman N., Ibrahim N., Utaberta N., Hassanpour B., *Rasch Modeling Analysis in Assessing Student's Ability and Questions Reliability in Architecture Environmental Science Examination // Journal of Applied Sciences Research, vol. 8(3), 2012. – С.1797-1801.*