

УДК 808.2: 159.937

М. В. МАКАРИЧ

## ПРИМЕНЕНИЕ ITEM RESPONSE THEORY ДЛЯ ОЦЕНКИ ЯЗЫКОВОГО ТЕСТА

*Белорусский национальный технический университет*

На современном этапе языковое тестирование является одним из разделов прикладной лингвистики и занимается определением уровня владения родным либо иностранным языком не только в учебных заведениях, но и при приёме на работу и получении гражданства. Целью предлагаемой работы является апробация Item Response Theory (IRT) для оценки языкового теста. В статье дается краткий исторический обзор методов оценки способностей и свойств личности. Приводятся фундаментальные принципы, положенные в основу стандартизации процедуры проведения тестирования, разработанные основоположником современной тестологии Дж. Кеттелом. Авторы дают подробное описание основных принципов классической теории тестов и ее современной интерпретации IRT, главной задачей которой является определение истинного тестового балла испытуемого, который зависит от многих условий – уровня трудности заданий, уровня подготовленности испытуемых, количества заданий и условий проведения тестирования. В статье авторы предлагают вариант использования IRT для оценки качества языкового теста, используемого в процессе контроля лексико-грамматических навыков по английскому языку. «Лингводидактический тест» рассматривается как классический тест, соответствующий четырем основным принципам, описанным В. А. Коккотой. В работе подробно описывается методика оценки качества языкового теста с использованием IRT.

**Ключевые слова:** языковой тест, классическая теория тестов, Item Response Theory (IRT), параллельные и эквивалентные тесты.

### Введение

Появление первых тестов относится к концу XIX – началу XX века. Родоначальником тестового движения считается известный английский ученый Френсис Гальтон. Хотя не все испытания Гальтона можно назвать тестами с позиции сегодняшнего дня, он сделал первый шаг на пути создания объективных методов оценки способностей и свойств личности [1]. Основоположником современной тестологии принято считать Дж. Кеттела. Он впервые ввел термин «умственные тесты» и сформулировал фундаментальные принципы, положенные в основу стандартизации процедуры проведения тестирования [2]. Рассматривая использование тестов в процессе обучения иностранным языкам, следует отметить, что сам термин «лингводидактический тест» был введен В. А. Коккотой, который соотносит лингводидактический тест со следующим комплексом заданий [3]:

1) задания, подготовленные в соответствии с определенными лингвистическими требованиями;

2) задания, прошедшие предварительную апробацию с целью выявления показателей качества;

3) задания, позволяющие определить у тестируемых степень их языковой (лингвистической) и/или речевой (коммуникативной) компетенции;

4) задания, результаты которых поддаются определенной оценке по заранее установленным критериям.

Данные дидактические принципы и сегодня являются основополагающими в процессе создания, апробации и оценки языковых тестов.

### Постановка проблемы

На современном этапе языковое тестирование является одним из разделов прикладной лингвистики и занимается определением уровня владения родным либо иностранным языком не только в учебных заведениях, но и при приёме на работу и получении гражданства. Фундаментальные основы данного научного направления, сконцентрированного на систематизации эмпирических данных, были заложены

создателем классической теории тестов (Classical Theory of mental tests) известным британским психологом Чарльзом Эдвардом Спирменом (1863–1945). Данная теория основывается на следующих пяти основных положениях [4]:

1. Эмпирически полученный результат измерения ( $X$ ) представляет собой сумму истинного результата измерения ( $T$ ) и ошибки измерения ( $E$ ):

$$X = T + E.$$

2. Истинный результат измерения можно выразить как математическое ожидание  $E(X)$ :

$$T = E(X).$$

3. Корреляция истинных и ошибочных компонентов по множеству испытуемых равна нулю, то есть  $\rho TE = 0$ .

4. Ошибочные компоненты двух любых тестов не коррелируют:

$$\rho E_1, E_2 = 0.$$

5. Ошибочные компоненты одного теста не коррелируют с истинными компонентами любого другого теста:

$$\rho E_1, T_2 = 0.$$

Кроме этого, основу классической теории тестов составляют два определения – *параллельных* и *эквивалентных* тестов. *Параллельные тесты* должны соответствовать требованиям (1–4), то есть истинные компоненты одного параллельного теста  $T_1$  должны быть равны истинным компонентам другого –  $T_2$  в каждой выборке испытуемых, отвечающих на оба теста. Предполагается, что  $T_1 = T_2$ . *Эквивалентные тесты* должны соответствовать всем требованиям параллельных тестов за исключением одного: истинные компоненты одного теста не обязательно должны равняться истинным компонентам другого параллельного теста, но отличаться они должны на одну и ту же константу  $c$ :

$$T_1 = T_2 + c_{12}$$

где  $c_{12}$  – константа различий результатов первого и второго тестов.

В классической теории тестов важнейшей проблемой является определение истинного тестового балла испытуемого ( $T$ ). Эмпирический тестовый балл ( $X$ ) зависит от многих условий – уровня трудности заданий, уровня подготовленности испытуемых, количества за-

даний, условий проведения тестирования и т. д. В этой связи возникает вопрос о возможности создания такого языкового теста, который можно было бы рассматривать не только как средство педагогического контроля, но и как средство измерений, представляющее собой систему тестовых заданий возрастающей трудности и специфической формы, который позволит надежно и объективно определить уровень усвоения знаний, сформированности языковых навыков и умений испытуемых и выразить результат в числовом эквиваленте. Другими словами возможность интерпретации числового результата теста позволяет адекватно оценить информативную валидность теста с тем, чтобы при необходимости произвести необходимую доработку тестовых заданий.

В современной практике тестирования анализ результатов теста проводится на основе «Item Response Theory» (IRT), современной интерпретации классической теории тестов. IRT является частью более общей теории латентно-структурного анализа Георга Раша – Rasch measurement [5]. На русский язык название «Item Response Theory» переводится различным образом. Ю. Нейман и В. Хлебников предлагают называть ее «Теория моделирования и параметризации педагогических тестов» (ТМППТ) [6]. Теория IRT обладает рядом преимуществ перед классической теорией тестов [7]:

1) превращает измерения, выполненные в дихотомических и порядковых шкалах, в линейные измерения, в результате чего качественные данные анализируются с помощью количественных методов;

2) мера измерения параметров является линейной, что позволяет использовать широкий спектр статистических процедур для анализа результатов измерений;

3) оценка трудности тестовых заданий не зависит от выборки испытуемых, на которых она была получена;

4) оценка уровня подготовленности испытуемых не зависит от используемого набора тестовых заданий;

5) неполнота данных (пропуск некоторых комбинаций: испытуемый – тестовое задание) не является критичным.

Полный перечень преимуществ модели IRT приведен в работе «Constructing Measures: An Item Response Modeling Approach» [8].

### Результаты исследования и их обсуждение

В настоящей работе приводится методика применения ИРТ для оценки качества тестовых заданий по английскому языку. Эксперимент проводился на кафедре английского языка № 2 БНТУ. Участники эксперимента – студенты 1 курса факультета энергетического строительства. Материалом для исследования послужили тесты для контроля лексико-грам-матических навыков, созданные в соответствии с учебной программой УО РБ [9].

На начальном этапе строится бинарная матрица результатов тестирования (табл. 1).

Таблица 1. Бинарная матрица (11×9)

	1	2	3	4	5	6	7	8	9
1. Шкурко	0	0	1	0	1	0	1	1	1
2. Скобялко	1	0	0	1	1	0	0	1	1
3. Руль	0	0	0	0	0	1	0	1	0
4. Сергейчик	0	0	0	0	1	0	1	0	1
5. Косько	1	0	0	0	1	0	0	1	0
6. Шнягина	1	1	1	0	0	0	1	1	1
7. Чулада	1	1	1	1	1	1	1	1	1
8. Ивашкевич	1	0	1	1	0	1	1	1	1
9. Толстикова	0	0	0	0	0	0	1	1	0
10. Кайрович	1	0	1	0	1	1	1	1	1
11. Федкович	1	1	1	0	1	1	1	1	1

Матрица состоит векторов-строк, содержащих значения индикатора для испытуемого. Профиль испытуемого – это последовательность значений индикатора в упорядоченной матрице тестовых результатов. В дихотомическом случае ответы испытуемого характеризуются двумя символами (цифрами) – 0 и 1. Нулю соответствует неверный ответ, единице – верный ответ.

Для дальнейшего анализа, нам потребуются значения  $X_i$  – индивидуального балла  $i$ -го испытуемого, количество верных ответов  $R_j$  на  $j$ -е задание. В табл. 2 приведены вычисленные значения  $X_i$  и  $R_j$ .

Таблица 2. Бинарная матрица (11×9) с индивидуальными тестовыми баллами испытуемых

	1	2	3	4	5	6	7	8	9	$X_i$
1. Шкурко	0	0	1	0	1	0	1	1	1	5
2. Скобялко	1	0	0	1	1	0	0	1	1	4
3. Руль	0	0	0	0	0	1	0	1	0	2
4. Сергейчик	0	0	0	0	1	0	1	0	1	4
5. Косько	1	0	0	0	1	0	0	1	0	3

Окончание табл. 2

	1	2	3	4	5	6	7	8	9	$X_i$
6. Шнягина	1	1	1	0	0	0	1	1	1	6
7. Чулада	1	1	1	1	1	1	1	1	1	9
8. Ивашкев.	1	0	1	1	0	1	1	1	1	7
9. Толстиков	0	0	0	0	0	0	1	1	0	2
10. Кайрович	1	0	1	0	1	1	1	1	1	7
11. Федков	1	1	1	0	1	1	1	1	1	8
$R_j$	6	3	6	2	7	5	8	9	8	55

Следует отметить, что если какое-либо задание успешно выполнили все испытуемые, его необходимо удалить из теста, так как оно не позволяет дифференцировать студентов и тест не может считаться нормативно-ориентированным. Соответственно, если бы на какое-то задание не ответил ни один испытуемый ( $R_j = 0$ ), то это задание тоже должно быть удалено из теста.

Далее рассчитываем количество неверных ответов  $W_j$  на  $j$ -е задание, долю верных ответов  $p_j$  и долю неверных ответов  $q_j$ .

$$X_i = \sum_{j=1}^M a_{ij} n.$$

В нашем случае ( $M = 9$ ) индивидуальный тестовый балл, например, для второго испытуемого ( $i = 2$ ) равен:

$$X_2 = \sum_{j=1}^9 a_{2j} = a_{21} + a_{22} + a_{23} + \dots + a_{29} = 1 + 0 + 0 + 1 + 0 + 0 + 0 + 1 + 1 = 4,$$

$$R_j = \sum_{i=1}^N a_{ij},$$

$$W_j = N - R_j.$$

В нашем случае ( $N = 11$ ) для третьего задания ( $j = 3$ ) получаем:

$$R_3 = \sum_{i=1}^{11} a_{i3} = a_{13} + a_{23} + \dots + a_{113} = 1 + 0 + 0 + 0 + 0 + 1 + 1 + 1 + 0 + 1 + 1 = 6,$$

$$W_3 = 11 - R_3 = 11 - 6 = 5.$$

Доля верных ответов на  $p_j$  на  $j$ -е задание равна:

$$p_j = \frac{R_j}{N},$$

где  $p_j$  – мера трудности задания. Доля неверных ответов равна:

$$q_j = 1 - p_j.$$

В нашем случае для третьего задания получим:

$$p_3 = \frac{R_3}{N} = \frac{6}{11} = 0,545,$$

$$q_3 = 1 - p_3 = 1 - 0,545 = 0,455.$$

В табл. 3 приведены расчетные значения,  $W_j$ ,  $p_j$ ,  $q_j$ . Для удобства визуального анализа фамилии испытуемых заменены номерами.

Таблица 3. Показатели меры трудности заданий теста

	1	2	3	4	5	6	7	8	9	$X_i$
1	0	0	1	0	1	0	1	1	1	5
2	1	0	0	1	1	0	0	1	1	4
3	0	0	0	0	0	1	0	1	0	2
4	0	0	0	0	1	0	1	0	1	4
5	1	0	0	0	1	0	0	1	0	3
6	1	1	1	0	0	0	1	1	1	6
7	1	1	1	1	1	1	1	1	1	9
8	1	0	1	1	0	1	1	1	1	7
9	0	0	0	0	0	0	1	1	0	2
10	1	0	1	0	1	1	1	1	1	7
11	1	1	1	0	1	1	1	1	1	8
$R_j$	6	3	4	2	7	5	8	9	8	55
$W_j$	3	3	5	5	3	5	2	2	2	
$p_j$	0,6	0,3	0,4	0,2	0,7	0,5	0,8	0,9	0,8	
$q_j$	0,3	0,3	0,5	0,5	0,3	0,5	0,2	0,2	0,2	
	0,18	0,09	0,2	0,1	0,21	0,25	0,16	0,18	0,16	

Важным параметром тестового задания является вариация (дисперсия) тестовых баллов  $p_j$ ,  $q_j$ . Чем больше вариация, тем лучше задание дифференцирует испытуемых. В табл. 3 последняя строка отражает зависимость вариации тестовых баллов от трудности задания. На рис. 1 приведен график зависимости вариации тестовых баллов от трудности задания. Видно, что максимальное значение, равное 0,25 достигается при  $p_j = 0,5$ . При  $p_j = 0$  и  $p_j = 1$  дисперсия задания равна нулю. Таким образом, если на задание не ответил ни один испытуемый или успешно ответили все, то задание не может их дифференцировать по уровню подготовленности.

Результаты нормативно-ориентированного тестирования при больших выборках обычно имеют распределение, близкое к нормальному



Рис. 1. Зависимость вариации тестовых баллов от трудности задания

(закон Гаусса). Другими словами эмпирическая кривая, полученная нами по результатам эксперимента соответствует нормальному закону распределения, что позволяет считать рассматриваемый языковой тест нормативно-ориентированным: тест хорошо дифференцирует испытуемых и их индивидуальные тестовые баллы в достаточной степени отличаются друг от друга. В случае отклонения полученной кривой от кривой Гаусса, задания исследуемого теста подлежат доработке.

### Заключение

Предложенная нами методика оценки качества языкового теста основана на классической теории тестов и может быть использована преподавателями иностранных языков для создания нормативно-ориентированных тестов с тем, чтобы объективно оценить уровень лексико-грамматических навыков обучаемых. Основные этапы данной методики: построение бинарной матрицы результатов тестирования исследуемого теста, отражающей профили испытуемых; вычисление значений  $X_i$  – индивидуального балла каждого испытуемого,  $R_j$  – количества верных ответов на каждое задание,  $W_j$  – количества неверных ответов,  $p_j$  – доли верных ответов,  $q_j$  – доли неверных ответов; построение графика зависимости вариации тестовых баллов от трудности задания для сопоставления его с кривой Гаусса.

### Литература

1. Аванесов, В. С. Проблема психологических тестов / В. С. Аванесов // Вопросы психологии, 1978. – № 5. – С. 97–107.
2. Drevdahl, J. E. Personality and Creativity in Artists and Writers / J. E. Drevdahl, R. B. Cattell // Journ. of Clinical Psychology, 1958. – vol. XIV. – N 2.
3. Коккота, В. А. Лингводидактическое тестирование / В. А. Коккота. – М.: Высш. шк., 1989. – 130 с.

4. **Crocker, L.** Introduction to Classical and Modern Test Theory / L. Crocker, J. Algina. – New-York: Holt, Rinehart and Wilson, 1986. – 527 p.
5. **Rasch, G.** Probabilistic Models for Some Intelligence and Attainment Tests / G. Rash. – Copenhagen, 1960, Danish Institute of Educational Research. (Expanded edition, Chicago, 1980, The University of Chicago Press).
6. **Нейман, Ю. М.** Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. – М.: Прометей, 2000. – 169 с.
7. **Маслак, А. А.** Измерение латентных переменных в социально-экономических системах: Монография. – Славянск-на-Кубани: Изд. центр СГПИ, 2006. – 333 с.
8. **Wilson, M.** Constructing Measures: An Item Response Modeling Approach / M. Wilson. – Mahwah, New Jersey: Lawrence Erlbaum associates, 2005. – 228 p.
9. **Мартысюк, Н. П.** Английский язык: пособие-репетитор для подготовки к централизованному тестированию / Н. П. Мартысюк, Т. Н. Руденко. – 2-е изд. – Минск: Аверсэв, 2012. – 415 с.

### References

1. **Avanesov, V. S.** The problem of psychological tests B. C. / V. S. Avanesov // The issues of psychology, 1978. – № 5. – P. 97–107.
2. **Drevdahl J. E., Cattell R. B.** Personality and Creativity in Artists and Writers. – Journ. of Clinical Psychology; April, 1958, vol. XIV, N 2.
3. **Kockota V. A.** Language testing / V. A. Kockota. – М.: Vys. shk., 1989. – 130 p.
4. **Crocker, L.** Introduction to Classical and Modern Test Theory / L. Crocker, J. Algina. – New-York: Holt, Rinehart and Wilson, 1986. – 527 p.
5. **Rasch, G.** Probabilistic Models for Some Intelligence and Attainment Tests / G. Rash. – Co-penhagen, 1960, Danish Institute of Educational Research. (Expanded edition, Chicago, 1980, The University of Chicago Press).
6. **Neiman, Y. M.** Preamble for modeling test theory and test parameterization in pedagogy / Y. M. Neiman, V. A. Chlebnikov. – М.: Prometey, 2000. – 169 p.
7. **Maslak, A. A.** Measuring of latent variables in social economic systems: Monograph / A. A. Maslak. – Slavyansk-na-Kubany: Publishing centre SGPI, 2006. – 333 p.
8. **Wilson, M.** Constructing Measures: An Item Response Modeling Approach / M. Wilson. – Mahwah, New Jersey: Lawrence Erlbaum associates, 2005. – 228 p.
9. **Martyssyk, N. P.** English language: special training for testing / N. P. Martyssyk, T. N. Rudenko. – 2-nd edit. – Minsk: Aversev, 2012. – 415p.

Поступила  
18.12.2017

После доработки  
20.12.2017

Принята к печати  
15.03.2018

*Makarych M. V.*

## ITEM RESPONSE THEORY AS A TOOL FOR THE ESTIMATION OF A LANGUAGE TEST

*Belarusian National Technical University*

*Being a part of applied linguistics a language test plays an important role in a modern society. It is used for evaluating person's language skills not only at educational establishments but also while acquiring citizenship or getting a job. The purpose of the article is the development of a special technique for language test estimation using Item Response Theory (IRT). Short historical review of methods for learning ability estimation is represented. Fundamental principles of Classical and Modern Test Theory is described. The author considers a linguodidactical test as a classical test that patterns four main principles by Kockota V. A. The methodology for the evaluation process of English linguodidactical test with the help of IRT is described in details. Such a linguodidactical test will take into account the level of tasks difficulty, the total amount of tasks, the level of person's knowledge and testing conditions. This methodology helps to create a test that can assess the real grade of a person being tested.*

**Keywords:** *a language test, Item Response Theory, Classical and Modern Test Theory.*



**Макарич Марина Васильевна**

Минский р-н, п. Валерьяново, ул. Янтарная, д. 13.

Тел. дом. (017) 510 92 28, тел. моб. + 37529 634 99 29.

E-mail: 2348843@tut.by.

Доцент кафедры английского языка № 2 БНТУ, кандидат филологических наук (специальность 10.02.21 – прикладная и математическая лингвистика), доцент. Также имеет диплом БНТУ (БПИ, факультет роботов) по специальности – автоматизация и комплексная механизация машиностроения. Научные интересы: педагогический аспект технического образования и лингвистическое обеспечение информационных систем.

**Makarych Marina**, Associate Professor of the 2<sup>nd</sup> English Department of the Belarusian National Technical University, PhD in Applied and mathematical linguistics. In addition she has B. Sc. in robotics.

E-mail: 2348843@tut.by.

Her scientific interests focus on engineering education and interdisciplinary education that combines linguistics and computer science.