

УДК 378.176

DOI: <https://doi.org/10.53920/ITS-2023-1-7>

Олександр Анатолійович КОСТИКОВ,

кандидат фізико-математичних наук, доцент,
ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«МЕТІНВЕСТ ПОЛІТЕХНІКА»
ORCID ID: 0000-0003-3503-4836

Тетяна Юріївна СОЛОМКО,

кандидат технічних наук, доцент, проректор,
Донбаський інститут техніки та менеджменту Закладу вищої освіти
«Міжнародний науково-технічний університет
імені академіка Юрія Бугая»
ORCID ID: 0000-0002-3029-7920

ОЦІНКА ЯКОСТІ ТЕСТОВИХ ЗАВДАНЬ МЕТОДАМИ СУЧАСНОЇ ТЕОРІЇ ТЕСТУВАННЯ

У статті проведено аналіз якості тестових завдань для контролю знань студентів на основі однопараметричної моделі Раша з використанням інформаційних функцій. Аналіз ґрунтувався на досягненнях сучасної теорії тестування IRT (Item Response Theory). Метою дослідження було виявлення недоліків тесту та визначення засобів їх усунення. Для досягнення цієї мети було наведено алгоритм побудови інформаційних функцій тестових завдань та тесту в цілому для однопараметричної моделі Раша. На основі дослідження інформаційних функцій запропоновано методику аналізу і підвищення якості тестових завдань. Продемонстровано застосування запропонованої методики на прикладі комп'ютерного тестування з дисципліни «Паралельні та розподілені обчислення», яке було проведено для студентів 4-го курсу ЗВО «Міжнародний науково-технічний університет імені академіка Юрія Бугая».

Ключові слова: модель Раша, латентний параметр, інформаційна функція, сучасна теорія тестування IRT.

Alexander KOSTIKOV

PhD in Phys. and Math., associative Professor
associative Professor, LLC «TECHNICAL UNIVERSITY
«METINVEST POLYTECHNIC»

Tatyana SOLOMKO

PhD in Engineering, associative Professor,
Prorector, IHE «Academician Yuri Bugay
International science and technical university»
ORCID ID 0000-0002-3029-7920

ASSESMENT OF THE TEST ITEM QUALITY BY THE METODS OF ITEM RESPONSE THEORY

In this article the analysis of the test quality based on one-parameter Rush's model using information functions is performed. The analysis was based on advances in Item Response Theory. The purpose of the study was to identify the shortcomings of the test and determine the means to eliminate them. To achieve this goal, an algorithm for constructing information functions of test items and the test for the one-parameter Rush's model was given. Based on the study of information functions, a methodology for analyzing and improving the quality of test items is proposed. The application of the proposed methodology on the example of computer testing in the discipline «Parallel and distributed computing», which was conducted for 4th year students of the IHE «Academician Yuri Bugay International science and technical university», is demonstrated. As result of this demonstration, the non-informative test items were found. Ways to improve the quality of test items were proposed.

Keywords: *Rash's model, latent parameter, information function, modern testing theory IRT.*

Постановка проблеми. Нині, у зв'язку з розвитком дистанційного навчання, для оцінювання знань широко застосовується тестовий підхід. Для реалізації тестового підходу в e-Learning дуже активно використовуються комп'ютерні технології. Так, у Міжнародному науково-технічному університеті імені академіка Юрія Бугая для дистанційного навчання використовується Google Classroom. Використання цієї платформи вимагає від викладачів створення електронного курсу і розробки тестових завдань для проведення комп'ютерного тестування. При цьому виникає проблема адекватності оцінювання знань на основі тестових технологій. Використання методів сучасної теорії тестування IRT (Item Response Theory) дозволяє оцінити якість розроблених тестових

завдань і їх придатність для ефективного оцінювання знань студентів.

Аналіз останніх досліджень і публікацій. Нині для оцінювання результатів тестування використовуються два підходи: математичний апарат класичної (Classical Test Theory) та сучасної теорії тестування (Item response Theory – IRT).

Відносно переваг та недоліків кожного підходу серед дослідників немає єдиного підходу [1, 2]. Класична теорія тестів від сучасної відрізняється засобом оцінювання учнів, результатом якого є підсумковий бал по тестовому завданню з урахуванням похибки.

У випадку методів IRT підсумковий бал є результатом сукупної взаємодії латентних параметрів, а саме справжнього рівня підготовки учнів та складності запропонованих завдань [1]. Від цих параметрів залежить ймовірність вірної відповіді на тестове завдання.

Для знаходження залежності ймовірності правильної відповіді від вищезазначених латентних параметрів в IRT використовуються різні математичні моделі. Серед них найбільш відомими є однопараметрична модель Раша [3], двохпараметрична та трьохпараметрична моделі Бірнбаума [4].

Однією із складових аналізу на основі IRT є дослідження інформаційної функції як тестового завдання, так і тесту в цілому. На основі інформаційної функції можна зробити висновки про адекватність тестових завдань для оцінки рівня підготовленості студентів.

Поняття інформаційної функції було введено у 1968 році в роботі [4] А. Бірнбаума. В цій роботі була виведена загальна формула інформаційної функції для дихотомічних завдань, тобто для завдань, відповідь на які має вигляд «так» чи «ні». Дослідженню цієї функції присвячено роботи багатьох авторів, зокрема, роботи Ф. Бейкера [5]. Для політомічних завдань (для яких відповіді мають декілька градацій) формула для інформаційної функції була виведена Ф. Сімеємою [6].

Аналіз тестових завдань на основі моделі Раша з використанням інформаційної функції було проведено в роботах авторів [7, 8].

Мета статті – провести аналіз розроблених тестових завдань для дистанційного навчання та надати рекомендації по підвищенню їх якості для адекватного оцінювання рівня знань студентів.

Виклад основного матеріалу дослідження. Аналіз якості тестових завдань проводився на основі моделі Раша [3], яка виражає залежність між ймовірністю правильної відповіді учня на тестове завдання і латентними параметрами за допомогою формули:

$$P_{ni} = \frac{\exp(\theta_n - \beta_i)}{1 + \exp(\theta_n - \beta_i)}$$

де P_{ni} – ймовірність того, що учасник n , $n = 1, \dots, N$ з рівнем підготовки θ_n вірно виконає завдання i , $i = 1, \dots, l$, з рівнем складності β_i .

В якості вихідних даних для аналізу було взято результати комп'ютерного тестування з курсу «Паралельні та розподілені обчислення» кафедри комп'ютерних наук та інженерії програмного забезпечення Міжнародного науково-технічний університету імені академіка Юрія Бугая.

В табл. 1 наведено статистичні характеристики відібраних для аналізу десяти тестових завдань для демонстрації запропонованої технології.

Таблиця 1. Статистичні характеристики тестових завдань

№ пит.	Успішність	Станд. відхилення	Призначена вага	Ефективна вага	Розрізнення	Ефективність розрізнення
1	33.33%	40.32%	5.0%	0,00%	-32.64%	-35.00%
2	76.00%	41.83%	6.0%	10.57%	34.34%	87.76%
3	66.67%	51.64%	6.0%	11.19%	73.29%	100.00%
4	76.00%	41.83%	6.0%	10.57%	34.34%	87.76%
5	33.33%	40.82%	6.0%	10.78%	92.20%	100.00%
6	100.00%	0.00%	5.0%	0.00%	0,00%	0,00%
7	83.33%	40.82%	5.0%	10.78%	92.20%	100.00%
8	100.00%	0.00%	5.0%	0.00%	0,00%	0,00%
9	50.00%	54.77%	5,0%	10.78%	60.19%	100.00%
10	100.00%	0.00%	5.0%	0.00%	0,00%	0,00%

Джерело: Розроблено авторами

Ці дані використовувались для аналізу якості тестових завдань.

На першому етапі проводилася оцінка рівня підготовки i -го студента (в логітах) за формулою:

$$\theta_i^0 = \ln\left(\frac{p_i}{q_i}\right), \quad i = 1, 2 \dots N,$$

де N – кількість учасників тестування, p_i – доля правильних відповідей i -го учасника на всі завдання, q_i – доля неправильних відповідей ($q_i = 1 - p_i$).

Далі обчислювався рівень складності тестових завдань β_j^0 в логітах за формулою:

$$\beta_j^0 = \ln\left(\frac{q_j}{p_j}\right), \quad j = 1, 2 \dots M,$$

де M – кількість тестових завдань, p_j – доля правильних відповідей всіх учасників на j -те завдання тесту, q_j – доля неправильних відповідей.

Далі виконувалося зведення рівня підготовки студентів θ_i^0 і рівня складності тестового завдання β_j^0 до єдиної інтервальної шкали для зниження впливу складності завдань на оцінки учасників тестування.

Попередньо обчисливши середнє значення початкових логітів рівня знань студентів

$$\bar{\theta} = \frac{\sum_{i=1}^N \theta_i^0}{N}$$

та стандартне відхилення V розподілу початкових значень параметра θ

$$V^2 = \frac{\sum_{i=1}^N (\theta_i^0 - \bar{\theta})^2}{N - 1},$$

отримаємо формулу для обчислення логіта складності j -ого завдання

$$\beta_j = \bar{\theta} + Y \cdot \beta_j^0, \quad j = \overline{1, M},$$

$$\text{де } Y = \left(1 + \frac{V^2}{2.89}\right)^{\frac{1}{2}}$$

Аналогічно, обчисливши $\bar{\beta} = \frac{\sum_{j=1}^M \beta_j^0}{M}$, $W = \sqrt{\frac{\sum_{j=1}^M (\beta_j^0 - \bar{\beta})^2}{M-1}}$, отримаємо формулу для обчислення логіта рівня знань i -го студента:

$$\theta_i = \bar{\beta} + X \cdot \theta_i^0, \quad i = \overline{1, N},$$

За допомогою отриманих значень можна співставити рівень знань здобувачів із рівнем складності завдань. Від'ємне і велике за модулем значення різниці $\theta_i - \beta_j$ означає, що завдання β_j є дуже важким для учасника тестування і не може використовуватися для виміру його рівня знань. У випадку додатної і великої за модулем різниці завдання вважається занадто легким і також воно не придатне для оцінювання рівня знань цього студента. Якщо $\theta_i = \beta_j$, то ймовірність того, що студент вірно виконає завдання, дорівнює 0,5.

На наступному етапі знаходяться інформаційні функції для тестового завдання та тесту в цілому.

Для обчислення інформаційної функції i -го завдання використовується формула:

$$I_i(\theta) = P_i(\theta) \cdot Q_i(\theta)$$

Інформаційна функція тесту обчислюється як сума інформаційних функцій тестових завдань:

$$I(\theta) = D^2 \cdot \sum_{j=1}^M I_j(\theta)$$

де D – поправний коефіцієнт ($D=1.7$), необхідний для наближення розподілу логістичної ймовірності до закону нормального розподілу.

На основі інформаційної функції обчислюється похибка вимірювання за формулою:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

Вона використовується для завершення комп'ютерного тестування в адаптивних алгоритмах оцінювання знань з використанням IRT.

Графік інформаційних функцій тестового завдання та тесту в цілому для використовуваних вихідних даних наведено на рис. 1.

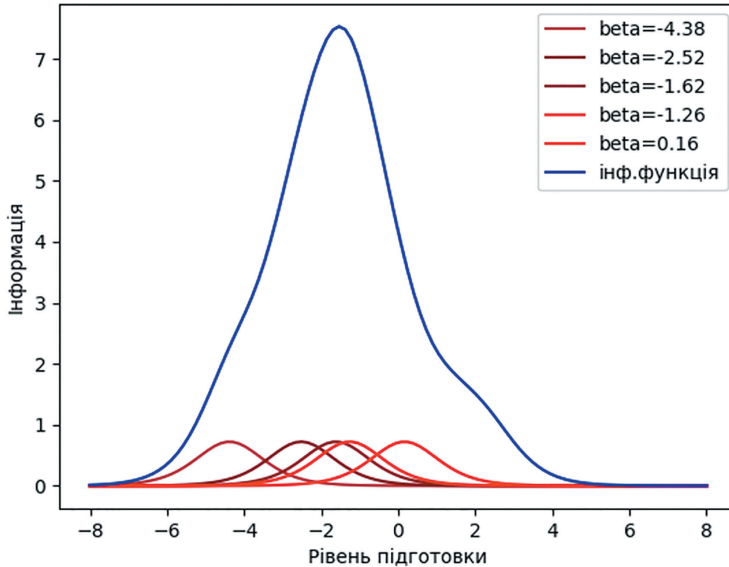


Рис. 1. Інформаційна функція тесту та тестових завдань

Джерело: Розроблено авторами

Аналізуючи рис.1, можна помітити, що інформаційна функція має один максимум, що означає збалансованість тесту. Також тест має багато легких завдань, про що свідчить розташування інформаційних функцій багатьох тестових завдань в інтервалі $(-3,1)$. Тест містить мало завдань підвищеної складності. Інформаційні функції тестових завдань розташовані нерівномірно, є задачі з близькою складністю, які можна видалити. В той же час необхідно додати до тесту задачі підвищеної складності.

Висновки та пропозиції. Аналіз якості тестових завдань на основі IRT виявив такі проблеми:

1) у тесті багато легких завдань та мало складних завдань, що приводить до завищення оцінки рівня навчальних досягнень здобувача освіти;

2) має місце нерівномірність заповнення інтервалів рівнів підготовки інформаційними кривими, що свідчить про недостатню структурованість тестових завдань за рівнем складності, а це, у свою чергу, не дозволяє здійснити об'єктивну диференціацію студентів за рівнем сформованості компетентностей.

Запропоновану методику можна застосовувати для створення тестів різної складності та підвищення якості тестових завдань, що може суттєво підвищити ефективність освітнього процесу з використанням платформ e-Learning.

© **Костіков О.А., Соломко Т.Ю., 2023**

ЛІТЕРАТУРА

1. Wiberg, M. Classical test theory vs. item response theory [Electronic resource] // *EM* No50. 2004. Way of access: http://www.edusci.umu.se/digitalAssets/59/59529_em-no-50.pdf. – Title from the screen
2. Ronald, K. Comparison of classical test theory and item response theory and their applications to test development // *Educational Measurement: issues and practice*. 1993. Way of access: http://www.internationalgme.org/Resources/Pubs/ITEMS_Module_16.pdf – Title from the screen.
3. Rasch G. Probabilistic models for some intelligence and attainment tests. – MESA Press, 5835 S. Kimbark Ave., Chicago, IL 60637; e-mail: MESA@uchicago.edu; web address: www.rasch.org; tele, 1993.
4. Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley. 1968. Chapters 17 – 20.
5. Baker F. *The Basics of Item Response Theory* / F. Baker. – ERIC Clearinghouse on Assessment and Evaluation. 2001. 172 p.
6. Samejima F. Estimation of Latent Ability Using a Response Pattern of Graded Scores / F. Samejima // *Psychometrika Monograph*. № 17. 1969. Pp. 1 – 100.
7. Kostikov A. A. The algorithm for knowledge assessment based on the Rasch model / A. A. Kostikov, K. V. Vlasenko, I. V. Lovianova, S. V. Volkov, E. O. Avramov // *CEUR Workshop Proceedings (9th Illia O. Teplytskyi Workshop on Computer Simulation in Education, CoSinE 2021, Kherson, 1 October 2021, Code 177072)*. 2022. Vol. 3083. pp. 28–42. Режим доступу до повного тексту статті: <http://ceur-ws.org/Vol-3083/paper268.pdf>.
8. Kostikov A, Vlasenko K, Lovianova I, Volkov S, Kovalova D, Zhuravlov M. Assessment of Test Items Quality and Adaptive Testing on the Rasch Model. *Communications in Computer and Information Science*. 2022 Nov 18;1698:252–71. doi:10.1007/978-3-031-20834-8_12. Режим доступу до повного тексту статті на сайті видання: https://doi.org/10.1007/978-3-031-20834-8_12.

REFERENCES

1. Wiberg, M. Classical test theory vs. item response theory [Electronic resource] // *EM* No50. 2004. Way of access: http://www.edusci.umu.se/digitalAssets/59/59529_em-no-50.pdf. – Title from the screen.
2. Ronald, K. Comparison of classical test theory and item response theory and their applications to test development // *Educational Measurement: issues and practice*. 1993. Way of access: http://www.internationalgme.org/Resources/Pubs/ITEMS_Module_16.pdf – Title from the screen.
3. Rasch G. Probabilistic models for some intelligence and attainment tests. – MESA Press, 5835 S. Kimbark Ave., Chicago, IL 60637; e-mail: MESA@uchicago.edu; web address: www.rasch.org; tele, 1993.
4. Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley. 1968. Chapters 17 – 20.
5. Baker F. *The Basics of Item Response Theory* / F. Baker. – ERIC Clearinghouse on Assessment and Evaluation. 2001. 172 p.
6. Samejima F. Estimation of Latent Ability Using a Response Pattern of Graded Scores / F. Samejima // *Psychometrika Monograph*. № 17. 1969. Pp. 1 – 100.
7. Kostikov A. A. The algorithm for knowledge assessment based on the Rasch model / A. A. Kostikov, K. V. Vlasenko, I. V. Lovianova, S. V. Volkov, E. O. Avramov // *CEUR Workshop Proceedings (9th Illia O. Teplytskyi Workshop on Computer Simulation in Education, CoSinE 2021, Kherson, 1 October 2021, Code 177072)*. 2022. Vol. 3083. Pp. 28 – 42. Way of access: <http://ceur-ws.org/Vol-3083/paper268.pdf>.
8. Kostikov A, Vlasenko K, Lovianova I, Volkov S, Kovalova D, Zhuravlov M. Assessment of Test Items Quality and Adaptive Testing on the Rasch Model. *Communications in Computer and Information Science*. 2022 Nov 18;1698:252–71. doi:10.1007/978-3-031-20834-8_12. Way of access: https://doi.org/10.1007/978-3-031-20834-8_12.

СТАТТЯ НАДІЙШЛА ДО РЕДАКЦІЇ 07.06.2023