

УДК 37.091.26:004.9

ПРАКТИЧНИЙ АНАЛІЗ ЯКОСТІ ТЕСТІВ З ВИЩОЇ МАТЕМАТИКИ

Диховичний Олександр Олександрович
Дудко Анна Федорівна
м.Київ

В статті досліджено одне з актуальних питань у галузі комп'ютерного тестування – питання якості тестів. Авторами на підставі методів класичної теорії тестів (КТТ) та Item Response Theory (IRT) в статті [1] було запропоновано методику такого аналізу. В даній роботі розглянуто застосування вищезазначеної методики. У якості контрольного прикладу використано комп'ютерний тест на тему «Диференціальне числення функцій однієї змінної», який було проведено для студентів першого курсу НТУУ «КПІ». Дана методика у подальшому може бути використана для вдосконалення системи тестування з вищої математики в НТУУ «КПІ» а також, можливо, в інших навчальних закладах.

Ключові слова: тестування з вищої математики, аналіз тестових завдань, аналіз тесту в цілому, латентні параметри, характеристичні криві завдань, характеристичні криві рівнів завдань, IRT-моделі.

Постановка проблеми. Всебічно визнаним способом контролю результатів навчання у світовій практиці є тестування. При цьому одним з ключових питань роботи з комп'ютерними тестами є питання вибору інструментарію для аналізу якості створених тестів. Основу такого аналізу традиційно складають статистичні методи, а саме методи КТТ або IRT. При використанні цих напрямків природно виникають проблеми взаємодії цих підходів, поєд-

нання їх в єдину методику, а також постає питання особливостей впровадження такої методики.

Аналіз актуальних досліджень і публікацій. Як зазначалось вище основними методами аналізу якості тестів є методи КТТ та IRT. Застосування математичних методів IRT, основи яких закладені в роботах Г.Раша, А.Бірінбаума, Ф.М.Лорда та ін. [2], розглянуто в наукових працях В.С.Аванесова [3], Н.Ф.Єфремової [4], В.С.Кіма [5], Т.В.Лісової [6], Н.Б.Челишкової [7] та ін. Методи КТТ розглядалися в працях Н.Б.Челишкової [7], В.С.Кіма [5], Л.Крокера та Дж.Алґіни [8], Л.О.Кухар та В.П.Сергієнко [9] та ін.

Незважаючи на широке коло робіт, в яких розглядається питання аналізу якості комп'ютерних тестів, єдина комплексна методика аналізу тестів, яка б поєднувала взаємодоповнюючі методи КТТ та IRT в літературі представлено недостатньо.

Метою статті є продемонструвати практичне застосування комплексної методики аналізу якості педагогічних тестів, представленої в статті [1].

Застосування методики аналізу якості тестів. Розроблена методика аналізу якості тестів з вищої математики застосовано для аналізу контрольної роботи на тему «Диференціальне числення функцій однієї змінної», проведеної у формі тесту для 51 студента першого курсу РТФ та ІТС у 2013 році.

На першому етапі сформовано таблицю результатів тесту, фрагмент якої представлено на рис. 1.

ПІБ	Номер завдання j															X _i
	№15	№3	№6	№10	№1	№8	№4	№11	№5	№14	№2	№12	№9	№13	№7	
Альона Олександрівна Бикова	0,53	0,53	0,53	0,53	0,4	0,53	0,53	0,53	0,53	0,53	0,27	0,53	0,53	0,53	0,53	7,6
Наталія Андрівна Гвоздецька	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0	0,53	0,53	0,53	7,47
Юрій Вікторович Яшенко	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0	0,53	0,53	0,53	0,53	0,53	0,53	7,47
Максим Райз	0,53	0	0	0	0	0,53	0	0	0,27	0	0,27	0	0	0,53	0	2,13
Джанет Обенева Дансо	0	0	0,53	0	0,13	0,53	0	0	0,27	0	0	0,53	0	0	0	2
R_j	25,97	21,73	21,2	20,67	18,63	18,55	18,02	15,9	15,43	15,37	14,43	13,78	13,25	11,66	4,77	

Рис. 1. Фрагмент таблиці результатів тесту «Диференціальне числення функцій однієї змінної»

На наступному етапі проведено аналіз тестових завдань і тесту у цілому.

Дані представлено у вигляді частотного розподілу індивідуальних балів на рис. 2.

Бал	[0;1]	(1;2]	(2;3]	(3;4]	(4;5]	(5;6]	(6;7]	(7;8]
Частота	0	1	7	6	11	11	11	4

Рис. 2. Частотний розподіл індивідуальних балів

Гіпотеза про нормальність розподілу індивідуальних балів тесту підтверджено за критерієм Пірсона.

Гістограму частотного розподілу індивідуальних балів представлено на рис. 3.

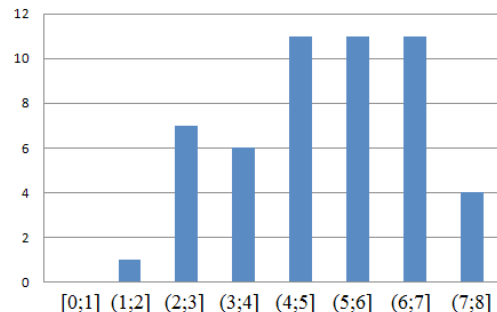


Рис. 3. Гістограма частотного розподілу індивідуальних балів

Значення асиметрії вибірки дорівнює -0,23, що говорить про дещо завищену складність тесту.

Коефіцієнт надійності тесту, знайдений на основі методу розщеплення, складає $r=0,78$; коефіцієнт надійності Кронбаха $\alpha=0,72$. В якості нижньої границі припустимих значень коефіцієнта надій-

Мода частотного розподілу індивідуальних балів дорівнює 6, середнє вибіркоче – 4,88. Правильно складений тест має бути унімодальним, і значення моди має бути близьким до значення вибіркового середнього [6]. Отже, даний тест потребує деякої корекції.

ності обрано 0,7 [1]. Отже, можна говорити про високу якість тесту.

Оцінено критеріальну валідність тесту. Критерієм обрано оцінки експертів, виставлені при традиційній перевірці знань учнів. Коefіцієнт валідності $\rho=0,82$. В якості нижньої границі припустимих значень коефіцієнта валідності обрано 0,7 [1]. Отже, даний тест володіє високою валідністю.

Побудовано інформаційну функція тесту, яка має не занадто загострену дзвоноподібну форму (рис. 4). Можна зробити висновок, що, в цілому, вигляд інформаційної функції відповідає правильно складеному тесту. Найбільш інформативним тест є для іспитників з рівнем підготовленості від -2 до 2,5 логітів.

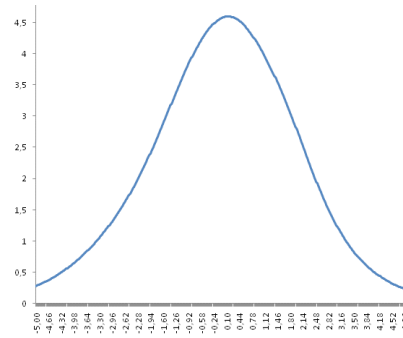


Рис. 4. Інформаційна функція тесту

Обчислено кореляційну матрицю завдань. (рис. 5).

№ завдання	№ 15	№ 3	№ 6	№ 10	№ 1	№ 8	№ 4	№ 11	№ 5	№ 14	№ 2	№ 12	№ 9	№ 13	№ 7
№ 15	1,00	0,15	0,14	0,36	0,05	0,08	0,29	0,24	-0,09	0,03	0,21	-0,20	0,20	0,18	0,09
№ 3		1,00	0,22	0,19	0,34	0,20	0,17	0,09	0,36	0,27	0,10	0,01	0,19	0,23	0,23
№ 6			1,00	0,50	0,27	0,36	0,03	-0,05	-0,08	0,02	0,05	0,06	0,23	0,07	0,24
№ 10				1,00	0,29	0,22	0,20	0,10	0,09	0,26	0,19	0,10	0,36	0,30	0,26
№ 1					1,00	0,19	0,33	0,13	-0,05	0,19	0,32	0,09	0,16	0,09	0,11
№ 8						1,00	0,15	0,04	0,21	-0,08	0,03	-0,07	0,24	0,08	0,09
№ 4							1,00	0,17	-0,03	0,22	0,27	0,06	0,36	0,28	0,11
№ 11								1,00	-0,14	0,32	0,35	0,22	0,18	0,17	0,07
№ 5									1,00	0,19	-0,14	-0,03	0,03	0,18	0,16
№ 14										1,00	0,31	0,10	0,22	0,44	0,20
№ 2											1,00	-0,11	0,29	0,38	0,05
№ 12												1	0,098462	-0,01708	0,042366
№ 9													1	0,254669	0,163411
№ 13														1	0,323773
№ 7															1

Рис. 5. Кореляційна матриця завдань тесту

Обчислено показники валідності окремих завдань тесту (рис. 6). Знайдено значення індексу розрізняльної здатності D_j . На основі практичного досвіду обрано наступні інтерпретації значень D_j : $D_j \geq 0,4$ – завдання добре функціонує; $0,3 \leq D_j \leq 0,39$ – можливо необхідна деяка корекція завдання; $0,2 \leq D_j \leq 0,29$ – завдання, яке потре-

бує перегляду; $D_j \leq 0,19$ – завдання, яке є складеним неправильно. Завдання №15 з індексом розрізняльної здатності $D_1=0,14$ є неправильно складеним. Привертають до себе увагу завдання №6, 7, 12 з $D_6 = D_7 = D_{12} = 0,36$. Текст завдань № 15, 6, 7, 12 відповідно на рис 7, 8,9,10. Далі розглянемо їх детальніше.

	#15	#3	#6	#10	#1	#8	#4	#11	#5	#14	#2	#12	#9	#13	#7
D_j	0,14	0,50	0,36	0,50	0,36	0,50	0,64	0,71	0,36	0,64	0,43	0,36	0,71	0,71	0,36
ρ_j	0,32	0,53	0,43	0,62	0,46	0,40	0,52	0,43	0,27	0,56	0,43	0,25	0,59	0,59	0,45
ρ_{bisj}	0,22	0,42	0,39	0,54	0,51	0,39	0,49	0,44	0,33	0,54	0,56	0,28	0,66	0,67	0,78
$t: t \leq \rho_{bisj}$	0,32	0,40	0,40	0,38	0,57	0,37	0,37	0,36	0,40	0,36	0,41	0,35	0,35	0,36	0,43

Рис. 6. Показники валідності завдань тесту

15 «
Баллов: 1

Похідну параметрично заданої функції $\begin{cases} x = x(t), \\ y = y(t) \end{cases}$ знаходять за формулою

Виберіть один ответ.

a. $\begin{cases} x = x(t), \\ y'_x = \frac{y'_t}{x'_t} \end{cases}$ ✓

b. $\begin{cases} x = x(t), \\ y'_x = \frac{x'_t}{y'_t} \end{cases}$ ✗

c. $\begin{cases} x = x(t), \\ y'_x = -\frac{x'_t}{y'_t} \end{cases}$ ✗

d. $\begin{cases} x = x(t), \\ y'_x = -\frac{y'_t}{x'_t} \end{cases}$ ✗

Рис. 7. Текст завдання №15

6 Баллов: 1 Закінчіть теорему Коші: Якщо функції $f(x)$ і $g(x)$ неперервні на $[a; b]$, диференційовні в $(a; b)$, причому $g'(x) \neq 0$, коли $x \in (a; b)$, то існує точка $c \in (a; b)$ така, що правдива рівність

Виберіть один ответ.

a. $\frac{f(b)f(a)}{g(b)g(a)} = \frac{f'(c)}{g'(c)}$ x
 b. $\frac{f(b) - g(b)}{f(a) - g(a)} = \frac{f'(c)}{g'(c)}$ x
 c. $\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}$ ✓
 d. $\frac{f(b) + f(a)}{g(b) + g(a)} = \frac{f'(c)}{g'(c)}$ x

Рис. 8. Текст завдання №6

7 Баллов: 1 Знайдіть $\lim_{x \rightarrow +\infty} (x + 3^x)^{1/x}$.

Виберіть один ответ.

a. **3** ✓
 b. ∞ x
 c. $\ln 3$ x
 d. e^3 x

Рис. 9. Текст завдання №7

12 Баллов: 1 Знайдіть Тейлорів многочлен 3-го порядку в точці $x = 0$ для функції $f(x) = \sin 2x$.

Виберіть один ответ.

a. $2x + \frac{4x^3}{3}$ x
 b. $x - \frac{x^3}{6}$ x
 c. $2x - 8x^3$ x
 d. $2x - \frac{4x^3}{3}$ ✓

Рис. 10. Текст завдання №12

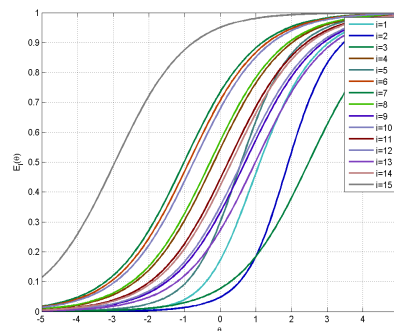
Обчислено коефіцієнти точкової бісеріальної кореляції між дихотомічними завданнями та індивідуальними балами іспитників і коефіцієнти кореляції Пірсона між політомічними завданнями та індивідуальними балами іспитників ρ . В якості нижньої границі припустимих значень коефіцієнта точкової бісеріальної кореляції 0,28 [1]. Завдання №5 і №12 потребують детального вивчення: $\rho_5 = 0,27$, $\rho_{12} = 0,25$.

Обчислено бісеріальний коефіцієнт кореляції між дихотомічним завданням та індивідуальними балами іспитників $\rho_{\text{bis}j}$. Для кожного завдання встановлено нижню границю припустимих значень бісеріальних коефіцієнтів кореляції [1]. Підтверджено неправильність складання завдань №12: $\rho_{\text{bis}12} = 0,28$ (при встановленому значенні нижньої границі 0,35) та №15: $\rho_{\text{bis}15} = 0,22$ (при встановленому значенні нижньої границі 0,32). Деяко виділяється завдання №6: $\rho_{\text{bis}6} = 0,39$ (при встановленому значенні нижньої границі 0,4).

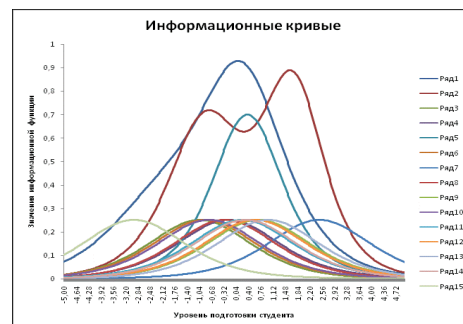
Побудовано ансамбль характеристичних кривих завдань тесту (рис. 11а). Можна побачити, що характеристичні криві досить рівномірно покривають інтервал від -2 до 2,5 логітів. Рівномірність покриття порушує лише завдання під номером 15 (рис. 7), яке дійсно виявилось занадто легким. Це тільки підтверджує неправильність його складання, виявлену раніше

Побудовано інформаційні криві всіх завдань тесту (рис. 11б).

Завдання №1 є інформативним (рис. 11б) для іспитників з рівнем підготовленості від -3,5 до 2,5 логітів. Аналіз характеристичних кривих рівнів завдання (рис 12а) з параметрами складності $\beta_1 = (-2,47; -0,58; 0,27; 1,02)$ підтверджує, що останнє є правильно складеним. Текст завдання №1 наведений на рис. 12б.

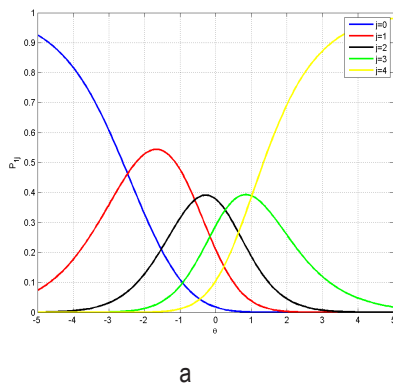


а



б

Рис. 11. Ансамблі характеристичних та інформаційних кривих завдань тесту



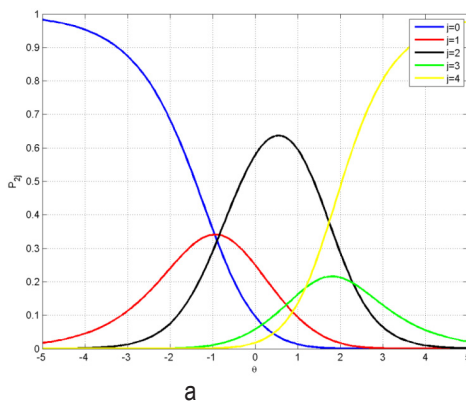
Встановіть відповідність між графіками функцій і типами поведінки функцій.

	скризь спадає	✓
	має одну точку перегину	✓
	має єдиний локальний максимум	✓
	має єдиний локальний мінімум	✓

Рис. 12. Характеристичні криві рівнів та текст завдання №1

Інформаційна крива завдання №2 (рис. 116) має провал, що означає, що рівень підготовленості студентів з різним значенням θ дане завдання оцінює з різною точністю. «Недосконалість» завдання підтверджено і виглядом характеристичних кривих рівнів, зображених на рис. 12а. Завдання №2 має 5 рівнів з параметрами складності $\beta_2 = (0; -0,95; -0,84; 2,3; 1,18)$. Як

бачимо 3 рівень цього завдання не спрацьовує (рис. 13а). Отже, це завдання не може бути ефективним для оцінювання рівня підготовленості іспитників з широкого інтервалу θ . Отже, у подальшому доречно зменшити кількість рівнів цього завдання до чотирьох. Текст завдання №2 наведено на рис. 13б.



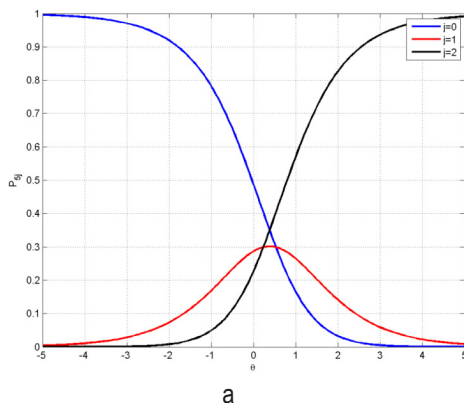
На рисунку зображено графік функції $y = f(x)$. Чому дорівнює:

диференціал $df(x)$	NK	✓
приріст аргументу Δx	MK	✓
приріст функції Δf	TK	✓
похідна $f'(x)$	NK/MK	✓

Рис. 13. Характеристичні криві рівнів та текст завдання №2

Інформаційна крива завдання №5 (рис. 116) має «загострену» дзвоноподібну форму, що говорить про різну точність оцінювання знань студентів з різними значеннями параметра підготовленості в інтервалі від -1,5 до 2,5 логітів. Розглянемо вигляд характеристичних кривих завдання №5 (рис. 14а). Дане завдання має

3 рівні з параметрами складності $\beta_5 = (0; 0,53; 0,23)$. Перший рівень цього завдання не спрацьовує. Отже, враховуючи «невдалість» завдання №5, виявлену на 4-ому і 5-ому етапах, його у подальшому доречно замінити іншим. Текст завдання №2 наведено на рис. 14б.



Вкажіть усі правильні твердження.

Выберите по крайней мере один ответ:

- а. Якщо $f'(x_0) = 0$ і під час переходу через точку x_0 похідна змінює знак, то в точці x_0 функція $f(x)$ має локальний екстремум ✓
- б. Для того, щоб у точці x_0 функція $f(x)$ мала локальний екстремум достатньо, щоб у в цій точці похідна функції або дорівнювала нулю, або не існувала ✗
- с. Якщо функція $f(x)$ диференційовна в точці x_0 , то вона неперервна в цій точці. ✓
- д. Кожна функція, неперервна на відрізку $[a, b]$, має точки локального екстремуму. ✗


Рис. 14. Характеристичні криві рівнів та текст завдання №5

Проведено аналіз дистракторів завдань з множинним вибором. Дистрактори, які вибирають менше 5% іспитників, повинні бути видалені або замінені.

В завданні №15 кожний дистрактор обрали лише 2%

іспитників. Отже, це завдання складене неправильно.

В завданні №13 (рис. 15) дистрактор а) обрали лише 4% іспитників, отже, його необхідно замінити іншим.

13  Баллов: 1

Знайдіть усі асимптоти графіка функції $y = \frac{x^2 - 2x}{x - 1}$.

Выберите один ответ.

a. $y = 1, y = x$ x

b. $x = 1, y = x$ x


c. $x = 1, y = x - 1$ ✓

d. $x = 1$ x

Рис. 15. Текст завдання №13

В завданні №6 (рис. 8) дистрактор а) відповідно обрали 4%. Отже, його необхідно замінити іншим.

В завданні №8 (рис. 16) дистрактор b), який обрали лише 4% іспитників, необхідно замінити іншим, а саме -1.

8  Баллов: 1

Знайдіть y' у точці (1;1), якщо $x^4 + x^2y^2 + y - 3 = 0$.

Выберите один ответ.

a. 1 x


b. 0 x

c. -2 ✓

d. 2 x

Рис. 16. Текст завдання №13

В завданні №3 (рис. 17) дистрактор с) обрали лише 4% іспитників, отже, його необхідно замінити іншим.

3  Баллов: 1

Визначте точки локального екстремуму функції $f(x) = 1 - 3x^2 - x^3$.

Выберите один ответ.

a. $x_{\max} = -1, x_{\min} = 0$ x

b. $x_{\max} = 0, x_{\min} = -2$ ✓

c. $x_{\max} = 0, x_{\min} = 1$ x

d. $x_{\max} = 1, x_{\min} = -2$ x

Рис. 17. Текст завдання №13

Підведемо підсумки щодо якості тесту. Розподіл вибірки індивідуальних балів близький до нормального, коефіцієнти надійності та критеріальної валідності більші за 0,7, графік інформаційної функції має дзвоноподібну не занадто загострену форму. Отже, в цілому, тест складений правильно. Найінформативнішим він є для іспитників з рівнем підготовленості від -2 до 2,5 логітів.

На неправильність завдання множинного вибору №15, що виявилось занадто легким на етапі аналізу ансамблю характеристичних кривих, вказують значення індексу розрізняльної здатності та коефіцієнту бісеріальної кореляції, а також аналіз дистракторів. Отже, завдання №15 необхідно замінити іншим зі складністю $\beta \in (-2; -1) \cup (1; 2)$.

На неправильність завдання множинного вибору №12 вказують значення індексу розрізняльної здатності, коефіцієнтів бісеріальної кореляції та точкової бісеріальної кореляції, а також аналіз кореляційної матриці. Аналіз дистракторів не виявив невдалий дистрактор або помилку в тексті завдання. Отже, завдання №12 необхідно замінити іншим зі складністю $\beta \in (-2; -1) \cup (1; 2)$.

На неправильність політомічного завдання №5 вказують значення коефіцієнта Пірсона, аналіз кореляційної матриці та сумісний аналіз характеристичних кривих підрівнів та інформаційної кривої завдання. Отже, дане завдання також необхідно замінити іншим зі складністю $\beta \in (-2; -1) \cup (1; 2)$.

На неправильність політомічного завдання №2 вказує аналіз кореляційної матриці. На основі сумісного аналізу характеристичних кривих підрівнів та інформаційної кривої завдання прийнято рішення зменшити кількість рівнів завдання №2 до трьох видален-

ням в тексті завдання з першої множини варіант «приріст аргументу Δx ».

Завдання множинного вибору №6 дещо виділялось на етапах аналізу кореляційної матриці, значень індексу розрізняльної здатності та коефіцієнту бісеріальної кореляції. Аналіз дистракторів виявив невдалий дистрактор а). Отже, в завданні № 6 дистрактор а) необхідно замінити.

На основі аналізу дистракторів завдань множинного вибору в завданнях №3 і №8 прийнято рішення замінити відповідно дистрактори с) і b).

Висновки. Проведений аналіз підтвердив практичну значущість методики аналізу якості тестів з вищої математики. Він показав, що:

- застосування розробленої методики дозволяє не лише виявити невдалий тесту, але і вказати на причини такої «невдачливості»;
- на основі такого аналізу у подальшому можливе здійснення покращення якості як окремих тестових завдань, так і тесту у цілому;
- взаємозв'язок КТТ та IRT яскраво виражений при аналізі неправильно складених завдань. Прийняття рішень для подальшого покращення тесту неможливе без сумісного врахування як показників КТТ, так і IRT;
- розроблена методика дозволяє проводити швидкий, зручний та наглядний аналіз якості завдань та тестів, але не може замінити ретельний аналіз тесту укладачем.

Література та джерела

1. Диховичний О.О. Комплексна методика аналізу якості тестів з вищої математики / О.О.Диховичний, А.Ф.Дудко // Комп'ютерно-орієнтовані системи навчання: Зб. наук. праць / Ребрада. – К.: НПУ імені М.П.Драгоманова. [Подано до друку]
2. Linden W. Handbook of Modern Item Response Theory / W.Linden, R.Hambleton. – NY: Springer-Verlag, 1997. – 510 p.
3. Аванесов В.С. Item Response Theory: Основные понятия и положения / Вадим Сергеевич Аванесов // Педагогические измерения. – 2007. – №2. – С.3-28
4. Ефремова Н.Ф. Тестовый контроль в образовании: Учеб. пособие / Надежда Федоровна Ефремова. – М.: Университетская книга, Логос, 2005. – 368 с.
5. Ким В.С. Тестирование учебных достижений. Монография / Владимир Сергеевич Ким. – Уссурийск: Издательство УГПИ, 2007. – 214 с.

6. Лісова Т.В. Моделі та методи сучасної теорії тестів / Т.В.Лісова. – Ніжин: Видавець ПП Лисенко М.М., 2012. - 112 с.
7. Чельшкова М.Б. Теория и практика конструирования педагогических тестов / Марина Борисовна Чельшкова. – М.: Логос, 2002. – 431 с.
8. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов. / под общей ред. В.И.Звонникова и М.Б.Чельшковой. – М.: Логос. - 2012. – 668 с.
9. Кухар Л.О. Конструювання тестів. Курс лекцій: навч. посіб. /Л.О.Кухар, В.П.Сергієнко. – Луцьк, 2010. – 182 с.

В статье исследован один из актуальных вопросов в области компьютерного тестирования – вопрос качества тестов. Авторами на основании методов классической теории тестов (КТТ) и Item Response Theory (IRT) в статье [1] была предложена методика такого анализа. В данной работе рассмотрено применение вышеупомянутой методики. В качестве контрольного примера использован компьютерный тест на тему «Дифференциальное исчисление функций одной переменной», который был проведен для студентов первого курса НТУУ «КПИ». Данная методика в дальнейшем может быть использована для совершенствования системы тестирования по высшей математике в НТУУ «КПИ», а также, возможно, в других учебных заведениях.

Ключевые слова: тестирование по высшей математике, анализ тестовых заданий, анализ теста в целом, латентные параметры, характеристические кривые заданий, характеристические кривые уровней заданий, IRT-модели.

The present paper studies one of the pressing questions in the area of computerized testing is question of tests quality. Authors put forward the methodology of such analysis on the basis of the methods of the classical test theory (КТТ) and Item Response Theory (IRT) in the paper [1]. In this paper the application of the above methodology is considered. As a case in point computerized test on "Differential calculus of functions of one variable" is used. The test was administered for the first-year students of Faculty of Radio Engineering and Institute of Telecommunication Systems of NTU "KPI". Later on this methodology can be used to improve the system of testing in higher mathematics at NTU "KPI" and possibly at other educational institutions.

Key words: testing in higher mathematics, analysis of test items, analysis of the test, latent parameters, item characteristic curves, characteristic curves of item levels, IRT-models.