

В. М. Кухаренко,

кандидат технічних наук,
професор, науковий керівник проблемної лабораторії дистанційного навчання,
Національний технічний університет "Харківський політехнічний інститут",
E-mail: kukharenkovn@gmail.com;

Л. П. Перхун,

кандидат педагогічних наук, доцент,
завідувач сектору дистанційного навчання,
E-mail: lperkhun@gmail.com;

Н. М. Товмаченко,

кандидат технічних наук, старший науковий співробітник,
заступник завідувача сектору дистанційного навчання,
E-mail: avt9tnn@gmail.com;
Національна академія статистики, обліку та аудиту

Методика комплексного оцінювання якості тестів. Частина 1

Розпочато виклад методики комплексного оцінювання якості тестів, що поєднує класичну теорію, методи Data Mining та Item Response Theory. Проведено факторний та кореляційно-регресійний аналіз. Використано показники описової статистики, точково-бісеріальний коефіцієнт кореляції, коефіцієнт кореляції Пірсона для бінарних змінних. Показана відмінність у тлумаченні результатів розрахунків для нормативно орієнтованих і критеріально орієнтованих тестів.

Ключові слова: дистанційне навчання, тестове завдання, тест, валідність, кореляція, факторний аналіз.

Інформатизація сучасного суспільства привела до широкого і стрімкого впровадження дистанційних технологій навчання практично в усіх типах навчальних закладів нашої країни [1]. Дослідженню різноманітних аспектів використання цифрових технологій в освітніх процесах приділяється значна увага як в Україні, так і в усьому світі. Не залишається без уваги і тестовий контроль знань.

Теоретичними та практичними аспектами впровадження й організації дистанційного навчання у середній і вищій школах, технологією розробки дистанційних курсів займалися такі вітчизняні вчені, як В. Биков, Ю. Богачков, В. Єршова, Д. Ільницький, Н. Краснопольська, І. Кулага, В. Кухаренко, А. Матвійчук, С. Стрельник, Б. Шуневич тощо. У роботах Г. Кравцова та В. Шарка висвітлюються питання стандартизації навчальних технологій, реалізованих на навчальній платформі Moodle, у тому числі відповідність структури та параметрів тестів міжнародним стандартам IMS та SCORM. Публікації О. Бурлакова, О. Галицького, П. Микитенка, І. Мушеника присвячені дослідженню вбудованих у середовище дистанційного навчання Moodle статистичних показників якості тестових завдань. Аналіз та опис програмних засобів для оцінювання якості тестових завдань наведено у роботах В. Фетісова. Особливостями застосування комп'ютерних технологій тестування та теоретичними аспектами статистичного оцінювання якості тестів займа-

лися Л. Крокер, Л. Кухар, М. Сіницький, П. Федорук. Проте роботи, присвячені комплексному аналізу якості тесту, в тому числі із застосуванням методів Data Mining, практично відсутні.

Метою дослідження є розробка комплексної методики оцінки якості окремих тестових завдань і тесту загалом.

Наразі у педагогічній практиці застосовуються два класи тестів: нормативно орієнтовані (НО) та критеріально орієнтовані (КО). НО тест використовують для порівняння навчальних досягнень студентів одного з одним. КО тест дозволяє оцінити кількість засвоєного навчального матеріалу. За його результатами приймається рішення щодо можливості / неможливості студента продовжувати навчання на основі попереднього завдання, за яким можна визначити мінімальну компетентність. Для НО тесту бажано, щоб результати тестування відповідали нормальному закону розподілу та мали значну дисперсію, тоді як для КО тестів спостерігаються значні відхилення від нормального розподілу та невелика дисперсія [2]. Зважаючи на це, по-різному мають інтерпретуватися результати статистичної обробки масиву тестових відповідей.

У навчальній діяльності Національної академії статистики, обліку та аудиту використовуються, в основному, критеріально орієнтовані тести. Вони реалізовані на базі системи дистанційного навчання (СДН) Moodle, яка дозволяє створювати тестові питання різних типів та неодноразово

включати їх у різні набори тестових завдань [3; 4]. Середовище Moodle пропонує розрахунок деяких статистичних показників щодо виконаного тесту та окремих його завдань: середня оцінка та медіана, стандартне відхилення, асиметрія, ексцес, коефіцієнт внутрішньої згоди, стандартна похибка тощо. Але наведених характеристик для обґрунтованого прийняття результатів тестування недостатньо [5; 6].

Авторами запропоновано комплексну методику оцінювання якості окремих тестових завдань і тесту загалом (рис. 1).

Схарактеризуємо окремі кроки запропованої методики. Для ілюстрації розрахунків скористуємося матрицею X результатів тестування 10-ти студентів за 10-ма тестовими завданнями [7]. Виконання окремого тестового завдання оцінювалося за бінарною шкалою:

$$x_{ij} = \begin{cases} 1, & \text{якщо студент } i \text{ правильно виконав завдання } j; \\ 0, & \text{якщо студент } i \text{ неправильно виконав завдання } j, \end{cases}$$

де i – номер студента, $i = 1, 10$, j – номер тестового завдання, $j = 1, 10$.

Матриця X має вигляд:

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad (1)$$

Зауважимо, що більшість математичних формул для обчислення статистичних характеристик вибірки ми опускаємо, оскільки вони загальновідомі. Всі розрахунки, що описані у статті, виконувались у пакеті SPSS [8].

Крок 1. Аналіз тесту загалом за допомогою описових статистик

На цьому кроці аналізується розподіл результатів тестування та значення основних описових статистик. Аналізу передують розрахунок індивідуальних тестових балів студентів. Індивідуальний тестовий бал i -го студента розраховується як сума елементів i -го рядка матриці X за формулою:

$$IS_i = \sum_{j=1}^{10} x_{ij}. \quad (2)$$

Якісний НО тест повинен ілюструвати нормальний розподіл індивідуальних балів студентів, його середнє значення має збігатися з модою, а більшість IS_i (близько 68%) групується навколо середнього. Додатне значення асиметрії свідчить про складність запропонованого тесту для конкретної групи студентів, від'ємне – про легкість.

Для КО тесту, навпаки, розподіл індивідуальних балів тесту є довільним. Як правило, відбувається зсув у бік високих балів.

У табл. 1 наведено результат розрахунку описових статистик за даними матриці (1). На рис. 2 зображено гістограму розподілу індивідуальних тестових балів студентів та її порівняння з кривою нормального розподілу.

Таблиця 1

Описові статистики результатів тестування за даними матриці (1)

Середнє	Медіана	Мода	Дисперсія	Стандартне відхилення	Асиметрія	Ексцес
5,0	4,5	4,0	6,888889	2,624669	0,276533	-0,41173

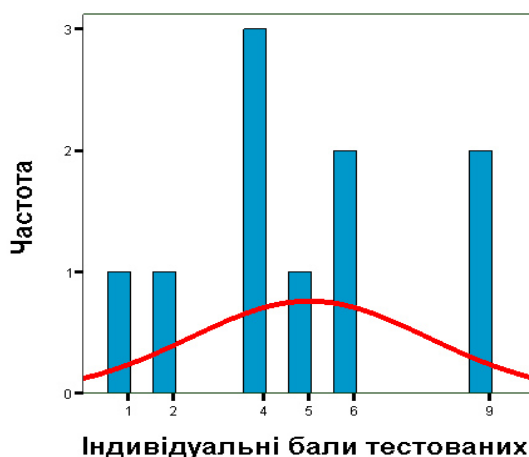


Рис. 2. Гістограма розподілу індивідуальних тестових балів студентів за даними матриці (1)

З аналізу табл. 1 та рис. 2 у випадку нормативно орієнтованого тесту можна зробити такі висновки:

1. Середній бал студента становить 5 (половину з 10 максимально можливих), тобто кількість студентів, які отримали більшу за середню оцінку, приблизно дорівнює кількості студентів, які отримали оцінку, меншу за середню. Для цього класу тестів це є прийнятним;

2. Найбільша кількість студентів набрала 4 тестових бали ($Mo = 4$);

3. Правостороння асиметрія свідчить про те, що тест у цілому був занадто складним для аналізованої групи студентів ($A > 0$ $\bar{X} > Me > Mo$);

4. Від'ємний ексцес ($E < 0$) свідчить про деяку якісну неоднорідність групи студентів, що проходили тестування.

Якщо ж тест критеріально орієнтований, то висновки зовсім інші:

1. Середній бал студента становить 5, що свідчить про низький рівень засвоєння студентами матеріалу, за яким відбулося тестування. Для цього класу тестів, як правило, встановлюється поріг не менше 70% від максимально можливої кількості балів, а отже, більше половини студентів цей поріг не подолали;

2. Правостороння асиметрія свідчить про те, що більшість студентів погано засвоїли матеріал, за яким складала тест, або більшість тестових питань викладено неякісно (відсутність чіткого формулювання завдання, двоїстість тлумачення тощо). Вершина розподілу індивідуальних балів студентів для цього класу тестів має бути зміщена вправо, середнє значення має бути близьким до порогового, а коливання навколо нього (стандартне відхилення) – незначним.

Крок 2. Оцінка валідності тестових завдань

У науково-педагогічній літературі приділяється значна увага тлумаченню поняття “валідність”, яке є досить багатогранним. Так, за визначенням А. Анастасі, валідність тесту у найпростішому розумінні – це уявлення про те, що саме тест вимірює і наскільки якісно він дозволяє це виміряти [9]. Більшість дослідників зазначають, що валідність не вимірюється, її можна тільки якісно оцінити: задовільна, недостатня, припустима тощо. Однак для висновків застосовуються кількісні величини – коефіцієнти оцінки кореляційного зв'язку.

В авторській методиці комплексного оцінювання якості тесту пропонується використовувати:

– точково-бісеріальний коефіцієнт кореляції; він дозволяє оцінити ступінь зв'язку між величинами, вимірними у різних шкалах: відповідями на конкретне тестове завдання (бінарна шкала) та індивідуальними результатами по всьому тесту в цілому (інтервальна шкала);

– коефіцієнт кореляції Пірсона для бінарних змінних (φ -коефіцієнт спряженості) для оцінки внутрішньої узгодженості завдань.

Точково-бісеріальний коефіцієнт кореляції розраховується за формулою:

$$r_{pbis} = \frac{(\bar{X}_1) - (\bar{X}_0)}{S_x} \sqrt{\frac{(N_1) \cdot (N_0)}{N \cdot (N-1)}}, \quad (3)$$

де $(\bar{X}_1)_j$ – середнє значення індивідуальних балів студентів, які правильно виконали j -те тестове завдання; $(\bar{X}_0)_j$ – середнє значення індивідуальних балів студентів, які неправильно виконали j -те тестове завдання; N_1 – кількість студентів, які правильно виконали j -те тестове завдання; N_0 – кількість студентів, які неправильно виконали j -те тестове завдання; N – загальна кількість студентів, які брали участь у тестуванні ($N = N_1 + N_0$); S_x – стандартне відхилення всіх індивідуальних балів студентів.

Точково-бісеріальний коефіцієнт кореляції може змінюватися від -1 до $+1$. Чим ближчий він за модулем до 1 , тим зв'язок тісніший. Якщо $|r_{pbis}| \leq 0,5$ зв'язок вважається досить слабким і таке завдання бажано переглянути на предмет редагування або заміни.

Коефіцієнт кореляції Пірсона (φ -коефіцієнт спряженості) обчислюється як:

$$\varphi_{ij} = \frac{P_{ij} - P_i P_j}{\sqrt{(p_i q_i)(p_j q_j)}}, \quad (4)$$

де P_{ij} – частка студентів, які дали правильні відповіді (отримали по одному балу) за обидва завдання з номерами i та j ; P_i – частка студентів, що дали правильну відповідь (отримали 1 бал) на i -те завдання; P_j – частка студентів, що дали правильну відповідь (отримали 1 бал) на j -те завдання; q_i – частка студентів, що дали неправильну відповідь (отримали 0 балів) на i -те завдання; q_j – частка студентів, що дали неправильну відповідь (отримали 0 балів) на j -те завдання; $q_i = 1 - p_i$; $q_j = 1 - p_j$.

Коефіцієнт кореляції Пірсона також набуває значень з відрізка $[-1; 1]$. Якщо деякі пари тестових завдань матимуть від'ємний коефіцієнт, це означитиме, що між ними існує обернений зв'язок і для покращення валідності тесту одне з них краще переробити або замінити іншим. Небажаною також є наявність тестових завдань з коефіцієнтом кореляції більшим за $0,7$. Такий тісний зв'язок свідчить про те, що обидва тестових завдання справляють однаковий вплив на загальний результат виконання тесту. Для критеріально орієнтованих тестів це є нормальним, а у випадку нормативно орієнтованого тесту одне з завдань слід замінити або переробити. В обох випадках незадовільних значень коефіцієнта кореляції слід

вилучати завдання, що має найменше значення
точково-бісеріального коефіцієнта.

Значення коефіцієнтів точково-бісеріальної ко-
реляції за даними матриці X узагальнено у табл. 2.

Таблиця 2

Коефіцієнти точково-бісеріальної кореляції за даними матриці X

Номер тестового завдання	1	2	3	4	5	6	7	8	9	10
Значення ($r_{pbis}j$)	0,535	0,5	0,263	0,738	0,723	0,643	0,789	0,246	0,803	0,535

Аналіз отриманих коефіцієнтів точково-бісеріальної кореляції свідчить про відсутність зв'язку між тестовими завданнями 2, 3 та 8 і результатами виконання тесту в цілому. Отже, ці тестові завдання потребують ретельного аналізу та коригування

або заміни. Інші тестові завдання вважатимемо такими, що забезпечують необхідну валідність тесту.

Підсумки обчислення парних коефіцієнтів кореляції Пірсона між результатами виконання окремих тестових завдань за даними матриці X узагальнено у табл. 3.

Таблиця 3

Коефіцієнти кореляції Пірсона за даними матриці X

Номер тестового завдання	Номер тестового завдання									
	1	2	3	4	5	6	7	8	9	10
1	1,000									
2	0,667	1,000								
3	0,509	0,218	1,000							
4	0,408	0,612	0,356	1,000						
5	0,333	0,000	0,218	0,408	1,000					
6	0,333	0,000	0,218	0,408	0,600	1,000				
7	0,218	0,327	-0,048	0,535	0,655	0,218	1,000			
8	-0,408	-0,102	-0,356	-0,167	0,000	0,000	0,356	1,000		
9	0,167	0,250	-0,218	0,408	0,500	0,500	0,764	0,612	1,000	
10	0,111	0,167	-0,509	0,272	0,333	0,333	0,509	0,408	0,667	1,000

З аналізу табл. 3 можна зробити такі висновки: тестове завдання 8 має найбільшу кількість від'ємних парних коефіцієнтів кореляції – з тестовими завданнями 1, 2, 3 та 4. Тестове завдання 3 також має чотири від'ємні парні коефіцієнти з завданнями 7–10. Якщо тестові завдання 3 і 8 переглянути (відкоригувати або замінити), всі від'ємні числа з таблиці 3 будуть вилучені й тест матиме внутрішню узгодженість. У випадку нормативно орієнтованого тесту слід звернути увагу на завдання 7 і 9, між якими існує досить щільний зв'язок (коефіцієнт кореляції 0,764). Точково-бісеріальні коефіцієнти кореляції дорівнюють 0,789 і 0,803 відповідно. Оскільки для тестового завдання 7 цей коефіцієнт менший, воно підлягає коригуванню або заміні. Врахування описаних зауважень буде забезпечувати необхідну валідність.

Крок 3. Оцінка факторної валідності тесту

Факторна валідність тесту оцінюється за допомогою інструментів факторного аналізу і дозволяє строго статистично проаналізувати структуру тестових завдань, виявити приховані закономірності їх взаємозв'язку та груповий вплив на досліджувану ознаку (індивідуальний тестовий бал).

Спочатку розглянемо графік кам'янистого спаду, який дозволяє визначити необхідну кількість факторів (рис. 3).

З аналізу рис. 3 видно, що тільки три фактори мають власні значення, більші за 1, отже, доречним буде виділення саме трьох факторів. Кумулятивний відсоток поясненої дисперсії при цьому становить 76,59% (табл. 4).

За даними табл. 5, де подано факторні навантаження на окремі тестові завдання, до фактора 1 увійшли тестові завдання з номерами 4, 5, 7, 9, які є домінуючими, тобто такими, що найбільше корелюють з індивідуальним тестовим балом студента (табл. 2). Фактор 2 об'єднав завдання 3 і 8, які практично не корелюють з результативною ознакою (табл. 2), крім того, від'ємно корелюють з рядом інших тестових питань (табл. 3). Фактор 3 містить одне завдання з номером 2, кореляція якого з індивідуальним тестовим балом становить 0,5. Тестові завдання 1, 6 та 10 взагалі не потрапили в жодний фактор, їх кореляція з результативною ознакою дещо більша за 0,5, але менша за 0,7.

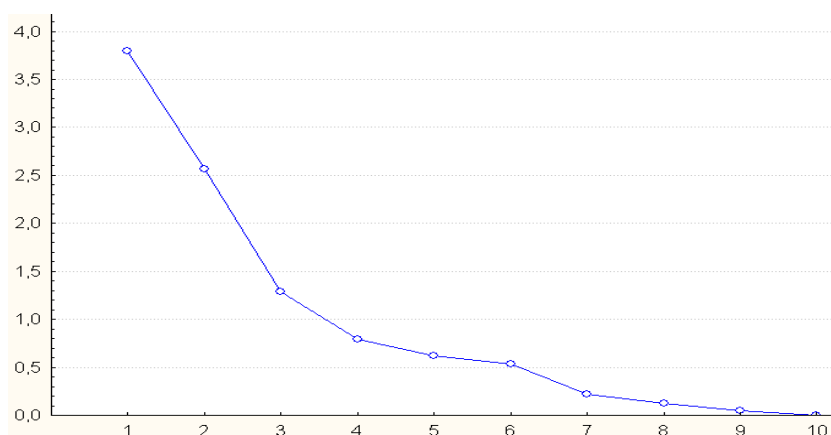


Рис. 3. Графік кам'янистого спаду власних значень

Таблиця 4

Власні значення факторів та дисперсія, що ними пояснюється

Номер фактора	Власні значення	Частка загальної дисперсії	Кумулятивні власні значення	Кумулятивна частка загальної дисперсії
1	3,800049	38,00049	3,800049	38,00049
2	2,569403	25,69403	6,369542	63,69452
3	1,290006	12,90006	7,659458	76,59458

Таблиця 5

Факторні навантаження на тестові завдання

Номер тестового завдання	Фактор 1	Фактор 2	Фактор 3
1	0,505	0,683	0,163
2	0,487	0,441	0,716
3	0,098	0,805	-0,243
4	0,719	0,394	0,164
5	0,726	0,073	-0,534
6	0,620	0,114	-0,580
7	0,828	-0,207	0,095
8	0,237	-0,777	0,107
9	0,843	-0,427	0,026
10	0,647	-0,509	0,151

З отриманих результатів можна зробити такі висновки:

1. Для будь-якого класу тестів завдання фактора 2 слід відкоригувати або замінити;

2. Для нормативно орієнтованих тестів слід уважно переглянути завдання 1, 2 та 10. На перші два з них за даними матриці X правильні відповіді дали практично всі студенти, тобто ці завдання виявилися занадто легкими. На останнє завдання правильна відповідь тільки одна. Це може свідчити про його складність або про те, що студенти не засвоїли зміст відповідного навчального елемента.

На завдання 6 студенти дали таку саму кількість правильних відповідей, як і на завдання 5, точково-бісеріальний коефіцієнт кореляції для нього (№ 6) досить високий — 0,643, з іншими тестовими завданнями кореляція додатна, отже, це завдання можна залишити без змін;

3. Для критеріально орієнтованих тестів завдання 1, 2 та 6 не вимагають перегляду та коригування, адже правильні відповіді студентів означають лише те, що вони гарно засвоїли відповідний навчальний елемент. Завдання 10 слід проаналізувати, оскільки мала кількість правильних відпові-

дей на нього може свідчити як про некоректне формулювання, так і про те, що студенти практично не засвоїли зміст відповідного елемента.

У випадку нормативно орієнтованого тесту

бажано також дослідити вплив виділених факторів на результативну ознаку – індивідуальний тестовий бал студента. Для цього обчислюють відповідні парні коефіцієнти кореляції (табл. 6).

Таблиця 6

Парні коефіцієнти кореляції між факторами та індивідуальним тестовим балом

Змінні	Індивідуальний тестовий бал	Фактор 1	Фактор 2	Фактор 3
Індивідуальний тестовий бал	1,00			
Фактор 1	-0,98	1,00		
Фактор 2	0,10	0,00	1,00	
Фактор 3	0,04	0,00	0,00	1,00

Коефіцієнти кореляції, наведені у табл. 6, свідчать про суттєвий вплив на індивідуальний бал студента тільки фактора 1. Це підтверджується і результатами регресійного аналізу, проведеного методом покрокового включення змінних у модель. Регресійне рівняння $IS = 5 - 2,575 \cdot F1$ виявилось єдиним статистично значущим рівнянням у цілому разом зі статистичною значущістю його параметрів. Тобто якщо б у тесті залишилися тільки завдання фактора 1 (тобто завдання з номерами 4, 5, 7 та 9), індивідуальні бали студентів були б такими

самими, як і за умови оцінювання за всіма завданнями тесту.

Результати проведеного кореляційно-регресійного аналізу підтверджують важливість коригування або заміни тестових завдань 1, 2, 3, 8, та 10 у випадку нормативно орієнтованого тесту.

Опис подальших кроків методики комплексного оцінювання якості тесту буде продовжено у наступних публікаціях. Також цікавим для дослідження є питання порівняння результатів класичної теорії, методів Data Mining та Item Response Theory.

Список використаних джерел

1. Світовий досвід організації та розвитку університетської системи дистанційного навчання / І. В. Кулага та ін. Київ: Інститут вищої освіти КНЕУ імені Вадима Гетьмана, 2013. URL: https://kneu.edu.ua/userfiles/education2_0/13-4713_verstka%281%29.pdf
2. Сіницький М. Є. Статистичні інструменти вимірювання якості освіти. Частина 1. Класичний підхід // Науковий вісник НАСОО. 2014. № 4. С. 58–69.
3. The Models of Distance Forms of Learning in National Academy of Statistics, Accounting and Audit / L. V. Deryhlyazov et al. // Науковий вісник Національної академії статистики, обліку та аудиту. 2017. № 3. С. 79–90.
4. Статистичні методи аналізу результатів тестового контролю знань / Л. В. Дериглазов та ін. // MoodleMoot Ukraine 2018. Теорія і практика використання системи управління навчанням Moodle: тези доп. Шостої міжнар. наук.-практ. конф. (25 травня 2018 р., м. Київ). Київ: КНУБА, 2018. С. 28.
5. Мокрієв М. В. Аналіз тестових завдань засобами Moodle // MoodleMoot Ukraine 2018. Теорія і практика використання системи управління навчанням Moodle: тези доп. П'ятої міжнар. наук.-практ. конф. (26–27 травня 2017 р., м. Київ). Київ: КНУБА, 2017. С. 18.
6. Quiz statistics report. MoodleDocs. URL: https://docs.moodle.org/32/en/Quiz_statistics_report
7. Кухар Л. О. Сергієнко В. П. Конструювання тестів. Курс лекцій: навч. посіб. Луцьк, 2010. 182 с.
8. Товмаченко Н. М. Основні характеристики пакета статистичних програм Statistical Package for Social Science та його застосування в прикладній статистиці // Науковий вісник Державної академії статистики, обліку та аудиту. 2007. № 2. с. 11–16.
9. Анастаси А., Урбина С. Психологическое тестирование / пер. с англ. Санкт-Петербург: Питер, 2001. 688 с.

References

1. Kulaha, I. V., Ilnytskyi, D. O., Strelnyk, S. O., Matviichuk, A. V., Vasyilkova, N. V., & Turchaninova, V. M., et al. (2013). *Svitovyi dosvid organizatsii ta rozvytku universytetskoï systemy dystantsiinoho navchannia* [World experience of organization and development of university system of distance learning]. Kyiv: Instytut vyshchoi osvity KNEU imeni Vadyna Hetmana. Retrieved from https://kneu.edu.ua/userfiles/education2_0/13-4713_verstka%281%29.pdf [in Ukrainian].

2. Sinytskyi, M. Ye. (2015). Statystychni instrumenty vymiriuvannia yakosti osvity. Ch. 2. Klassychnyi pidkhid [Statistical tools for measuring the education quality. Part 2. Classical approach]. *Naukovyi visnyk Natsionalnoi akademii statystyky, obliku ta audytu*. – *Scientific bulletin of the National Academy of Statistics, Accounting and Audit*, 1, 75–86.

3. Deryhlazov L. V., Kukharenko, V. M., Perkhun, L. P., Tovmachenko, N. M. (2017). The Models of Distance Forms of Learning in National Academy of Statistics, Accounting and Audit. *Naukovyi visnyk Natsionalnoi akademii statystyky, obliku ta audytu* – *Scientific bulletin of the National Academy of Statistics, Accounting and Audit*, 3, 79–90 [in English].

4. Deryhlazov, L. V., Kukharenko, V. M., Perkhun, L. P., Tovmachenko, N. M. (2018). Statystychni metody analizu rezultativ testovoho kontroliu znan [Statistical methods for analyzing the results of knowledge test control]. Proceedings from MoodleMoot Ukraine 2018. Theory and Practice of Using the Moodle Learning Management System: *Shosta mizhnarodna naukovo-praktychna konferentsiia (25 travnia 2018 hoda)* – *Sixth International Scientific and Practical Conference*. (p. 28). Kyiv, KNUBA [in Ukrainian].

5. Mokriiev M.V. (2017). Analiz testovykh zavdan zasobamy Moodle [Analysis of test tasks by means of Moodle]. Proceedings from MoodleMoot Ukraine 2017. Theory and Practice of Using the Moodle Learning Management System: *Shosta mizhnarodna naukovo-praktychna konferentsiia (26–27 travnia 2017 hoda)* – *Sixth International Scientific and Practical Conference*. (p. 18). Kyiv, KNUBA [in Ukrainian].

6. Quiz statistics report. MoodleDocs. docs.moodle.org. Retrieved from https://docs.moodle.org/32/en/Quiz_statistics_report

7. Kukhar, L. O., & Serhiienko, V. P. (2010). *Konstruiuvannia testiv [Create a quality test]*. Lutsk [in Ukrainian].

8. Tovmachenko, N. M. (2007). Osnovni kharakterystyky paketa statystychnykh prohram Statistical Package for Social Science ta yoho zastosuvannia v prykladnii statystytsi [The basic properties of Statistical Package for the Social Science and its use in the applied statistics] // *Naukovyi visnyk Derzhavnoi akademii statystyky, obliku ta audytu* – *Scientific bulletin of the State Academy of Statistics, Accounting and Audit*, 2, 11–16 [in Ukrainian].

9. Anastazi, A. & Urbina, S. (2001). *Psikholohicheskoe testirovanie [Psychological Testing]*. Sankt-Peterburg: Piter [in English].

В. Н. Кухаренко,

кандидат технических наук,

профессор, научный руководитель

проблемной лаборатории дистанционного обучения,

Национальный технический университет “Харьковский политехнический институт”;

Л. П. Перхун,

кандидат педагогических наук, доцент,

заведующая сектором дистанционного обучения;

Н. Н. Товмаченко,

кандидат технических наук, старший научный сотрудник,

заместитель заведующей сектором дистанционного обучения;

Национальная академия статистики, учета и аудита

Методика комплексного оценивания качества тестов. Часть 1

Начато изложение методики комплексной оценки качества тестов, сочетающей классическую теорию, методы Data Mining и Item Response Theory. Проведен факторный и корреляционно-регрессионный анализ. Используются показатели описательной статистики, точно-бисериальный коэффициент корреляции, коэффициент корреляции Пирсона для бинарных переменных. Показана разница в толковании результатов расчетов для нормативно ориентированных и критериально ориентированных тестов.

Ключевые слова: дистанционное обучение, тестовое задание, тест, валидность, корреляция, факторный анализ.

V. M. Kukharenko,

PhD in Engineering,

*Professor, Scientific supervisor of Problem laboratory of distance learning,
National Technical University "Kharkiv Polytechnic Institute";*

L. P. Perkhun,

PhD in Pedagogy, Associate Professor,

Head of the Distance Learning Sector;

N. M. Tovmachenko,

PhD in Engineering, Senior Researcher,

Deputy Head of the Distance Learning Sector;

National Academy of Statistics, Accounting and Audit

The Method for Comprehensive Quality Evaluation of Tests. Part 1

Informatization of the modern society has led to the wide-scale and rapid introduction of distance training technologies in virtually all the categories of Ukrainian HEEs. Studies of the aspects related to applications of digital technologies in education processes are subject to close attention in Ukraine and beyond. An important component of the training process is test control of knowledge.

Education activities at the National Academy of Statistics, Accounting and Audit rely on criteria oriented tests. They are realized on the basis of distance training system Moodle that allows for creating test questions of various types and their repeated inclusion in various packages of test tasks. The Moodle environment allows for computation of selected statistical indicators on a fulfilled test and its individual tasks: average estimate and median, standard deviation, asymmetry, excess, internal agreement rate, standard error etc. However, these characteristics are not enough for justified acceptance of test results.

The article presents the first phase in elaborating a comprehensive method for quality evaluation of selected test tasks and the test as a whole. This method combining the classical theory, Data Mining and Item Response theory methods involves six steps. The first step, based on indicators of descriptive statistics, allows for evaluating the obtained distribution of test results. The second step involves evaluation of the validity of test tasks. The point-bead ratio is computed to derive the correlation between individual test task and individual test score of a student, with values higher than 0.5 considered satisfactory. Pearson correlation coefficient for binary variables shows the correlation between pairs of test tasks. The test tasks with negative correlations with the other test tasks are not considered as valid, and they have to be corrected or replaced. At the third step, the factor validity of the test is evaluated. The test tasks combined in groups using factor analysis methods are subject to further analysis to determine their impact on the final result, the individual test score of a student. All the above mentioned steps are illustrated by example. The computation is made by SPSS software package. The difference in interpretation of the computation results in each step for norm oriented and criteria oriented tests is demonstrated.

The description of further steps involves in the method for comprehensive quality evaluation of tests, which use Data Mining and Item Response Theory methods, will be continued in next publications.

Key words: *distance training, test task, test, validity, correlation, factor analysis.*

Бібліографічний опис для цитування:

Кухаренко В. М., Перхун Л. П., Товмаченко Н. М. Методика комплексного оцінювання якості тестів. Частина 1 // Статистика України. 2018. № 3. С. 40–48.