

Боголюбов В. М., Замостьян В. П.,  
Середа Л. І.

## АНАЛІЗ ТЕСТОВИХ ЗАВДАНЬ ПРИ ФІНАЛЬНОМУ КОНТРОЛІ РІВНЯ ЗНАНЬ З ЕКОЛОГІЧНИХ ДИСЦИПЛІН

*Проведені дослідження якісних характеристик тестових завдань при фінальному контролі рівня знань з екологічних дисциплін дозволили визначити ступінь надійності та об'єктивності тестових завдань і скоригувати рівень валідності тестів і завдань деяких навчальних дисциплін, які викладаються на кафедрі екології НаУКМА.*

У більшості розвинених країн світу основною формою визначення рівня знань на всіх етапах навчання є технологія тестового контролю [1].

Тестування визнано найбільш об'єктивним і прогресивним методом оцінювання рівня знань з повністю сформованим термінологічним і понятійним апаратом. Проект державного стандарту вищої освіти України основним засобом діагностики рівня освітньо-професійної підготовки студентів також визначає стандартизований тестовий контроль.

У педагогічних дослідженнях під *тестуванням* мають на увазі метод педагогічної діагностики, за допомогою якого оцінюють якісні характеристики навчального процесу і який максимально відповідає принципам “зіставлення”, об'єктивності, надійності та валідності вимірів [2].

Одним з етапів тестового контролю є психометричний аналіз тестів та тестових завдань. Проект державного стандарту вищої освіти України визначає технологію психометричного аналізу тестів через визначення таких статистичних параметрів [3]:

1. Середній арифметичний бал —  $X$ ;
2. Середньоквадратичне відхилення —  $\sigma$ .
3. Коефіцієнт надійності —  $R$ .
4. Стандартна помилка вимірювання —  $E$ .
5. Коефіцієнт валідності —  $V$ .

Однією з основних проблем тестування є питання про те, виконаний чи невиконаний іспит. Найчастіше використовується *абсолютний критерій*: оцінка визначається сумою правильних відповідей на тестові завдання — іспит вважається складеним, якщо правильні відповіді становлять не менше 70 % від загальної кількості тестових завдань. Недоліком такого підходу є

те, що при цьому практично неможливо забезпечити рівноцінність всіх тестових завдань за складністю. З метою усунення такого недоліку сучасні технології оцінювання результатів тестування використовують *відносний критерій*, який дозволяє оцінювати результати кожного студента на підставі аналізу результатів відповідей всіх інших студентів, які брали участь у тестуванні. При цьому, чим чисельніша група студентів, тим надійніший результат. Вважається, що статистичні розрахунки будуть достовірними, якщо в групі нараховується не менше 20 осіб, а кількість тестових завдань становить 80—120 (помилка вимірювання при цьому не перевищує 10 %).

При користуванні відносним критерієм необхідно проводити попередній якісний аналіз, який полягає у визначенні індексу складності та індексу диференціюючої здатності для кожного тестового завдання. Алгоритм проведення якісного аналізу складається з 6 етапів (рис. 1). Кількість балів  $N_i$  кожного студента в абсолютних числах визначається числом правильних відповідей. Після ранжування всіх результатів від найбільшого до найменшого визначають “сильну” і “слабку” групи, використовуючи принцип “33 %”: до “сильної” групи відносять 33 % студентів з найвищим балом, а до “слабкої” — 33 % студентів з найнижчим балом. Індекси складності  $I_{cj}$  та диференціюючої здатності  $I_{dj}$  для кожного тестового завдання визначаються як відносні коефіцієнти суми і різниці правильних відповідей на це завдання у “сильній” і “слабкій” групах за формулами:

$$I_{cj} = (H + L) / n, \quad (1)$$

$$I_{dj} = 2(H - L) / n, \quad (2)$$

де  $H$  — кількість правильних відповідей у “сильній” групі;

$L$  — кількість правильних відповідей у “слабкій” групі;

$n$  — загальна кількість осіб у двох групах;

$j$  — кількість тестових завдань в екзаменаційному тесті.

Ідеальними вважаються тестові завдання з індексом складності в межах від 50 до 60 %, а прийнятними — від 30 до 70 %.

Індекс диференціюючої здатності є якісною характеристикою тестового завдання — чим він

вищий, тим найточніша відповідь на це завдання виявляє найбільш знаних з даної дисципліни. При  $I_{dj} \leq 0,14$  тестове завдання вважається некоректним і його вилучають. Тестове завдання з  $0,15 \leq I_{dj} \leq 0,24$  вважається суперечливим, підлягає додатковому аналізу і коригуванню.

Тестові завдання з незадовільними значеннями індексів вилучаються. Якщо значення індексу складності низьке, а індекс диференціюючої здатності задовільний, тестове завдання залишають для остаточних підрахунків.

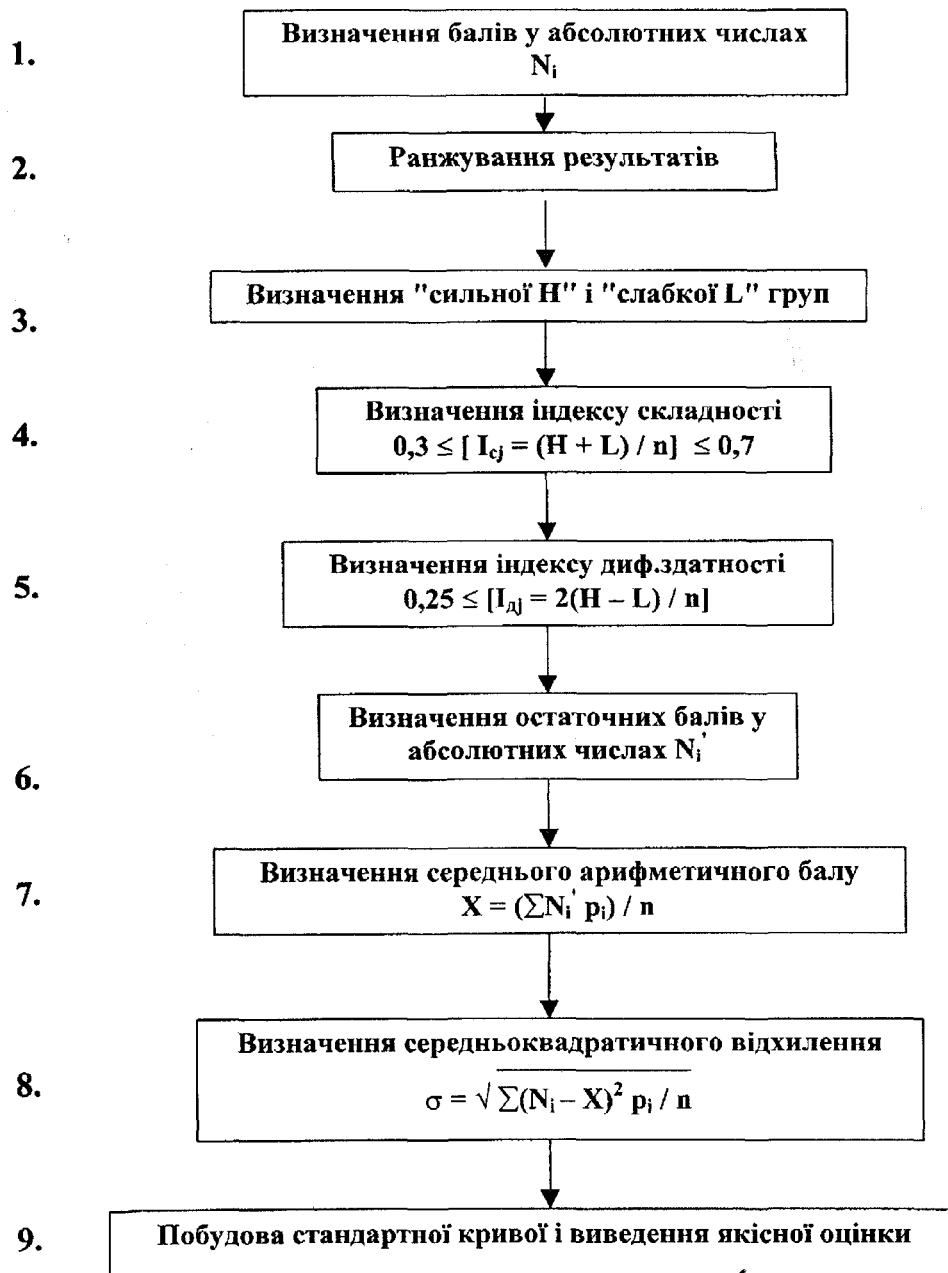


Рис. 1. Блок-схема алгоритму оцінювання за відносним критерієм:

$i$  — загальна кількість студентів на іспиті;  $n$  — загальна кількість осіб у “сильній” і “слабкій” групах;  $p$  — кількість студентів, які отримали даний бал  $X_i$ ; (решта позначень — у тексті)

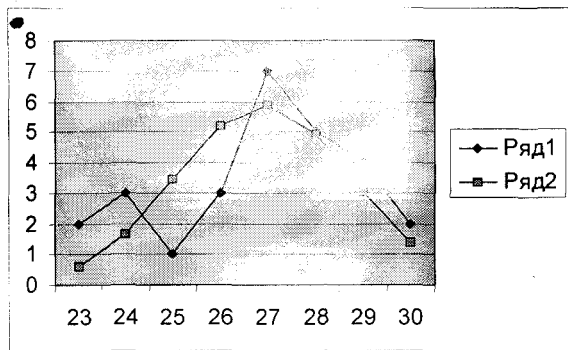


Рис. 2. Статистичні криві розподілу відповідей

Така процедура аналізу тестових завдань допомагає виключити вплив незадовільно складених завдань на результати іспиту.

Після визначення середньоарифметичного бала та середньоквадратичного відхилення будується крива нормального розподілу (рис. 2) шляхом нанесення на вісь абсцис середньоарифметичного бала  $X$  і значень  $(X + \sigma)$ ,  $(X + 2\sigma)$ ,  $(X + 3\sigma)$  та  $(X - \sigma)$ ,  $(X - 2\sigma)$ ,  $(X - 3\sigma)$ , а на вісь ординат — число студентів, які отримали відповідну кількість балів (в даному варіанті тестового іспиту від 1 до 7).

Нами було проаналізовано тестові завдання з курсів “Моніторинг навколишнього середовища” та “Агроекологія”, які викладаються для студентів-екологів. Тести мали тридцять тестових завдань. У тестуванні брали участь 25 студентів.

Процес тестування відбувався у такій послідовності: всі студенти розміщувалися в аудиторії по одному за столом у шаховому порядку і одночасно отримували тестові завдання; через 45 хвилин завдання збирали — за відсутності відповіді на якесь тестове завдання вважали, що відповідь неправильна.

Якісний аналіз тестових завдань після іспиту навесні 1999 р. показав, що більшість тестових завдань було складено некоректно: за індексом складності задовільними були 5 завдань, а за індексом диференціюючої здатності — 10. Обом індексам задовольняли тільки 3 завдання.

Після відповідного коригування тестових завдань результати іспиту, проведеного з новою групою в кількості 27 студентів, були значно кращими — за індексом складності задовільними були 25 завдань, а за індексом диференціюючої здатності — 28 (рис. 3). Обом індексам задовольняли 23 завдання.

№ студ. № завд.	Сильна група									Н	Слабка група									L	Індекс І <sub>с</sub>	Індекс І <sub>д</sub>
	1	2	3	4	5	6	7	8	9		1	2	3	4	5	6	7	8	9			
1	1	1	1	1	1	1	1	1	1	9	1	1	1	1	0	1	1	1	8	94,4	0,11	
2	1	1	1	1	1	1	1	1	1	9	1	0	1	1	0	1	1	0	1	6	83,3	0,33
3	1	1	1	1	1	1	1	1	1	9	1	1	0	1	1	1	1	1	1	8	94,4	0,11
4	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
5	1	1	1	1	1	1	1	1	1	9	1	1	1	1	0	0	0	0	1	6	83,3	0,33
6	1	1	1	1	1	1	1	1	1	9	1	1	1	1	0	0	0	0	1	5	77,8	0,44
7	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
8	1	1	1	1	1	1	1	1	1	9	1	1	1	1	0	0	1	1	1	7	88,9	0,22
9	1	1	1	1	1	1	1	1	1	9	1	1	0	1	1	1	1	1	1	8	94,4	0,11
10	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	0	1	1	8	94,4	0,11	
11	1	1	1	0	0	1	0	1	1	6	1	1	1	0	0	0	1	0	0	4	55,6	0,22
12	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	0	8	94,4	0,11
13	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
14	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
15	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
16	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
17	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
18	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
19	1	0	1	1	1	1	1	0	1	7	0	0	1	0	1	1	1	1	0	5	66,7	0,22
20	1	1	1	1	1	1	1	1	1	9	0	1	1	1	1	1	1	1	1	8	94,4	0,11
21	1	1	1	1	1	1	1	1	1	9	0	1	1	1	1	1	1	1	1	8	94,4	0,11
22	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0
23	1	1	1	1	1	1	1	1	1	9	0	0	1	1	1	1	1	1	1	7	88,9	0,22
24	1	1	1	1	1	1	1	1	1	9	1	1	0	0	1	1	0	1	1	6	83,3	0,33
25	0	0	1	1	1	0	1	1	1	6	0	1	0	1	0	1	1	1	0	5	61,1	0,11
26	1	1	1	1	0	1	0	1	1	7	0	1	1	0	0	1	0	0	4	61,1	0,33	
27	1	1	1	1	1	1	1	1	1	9	1	0	1	1	1	1	1	1	1	8	94,4	0,11
28	1	1	1	1	1	1	1	1	1	9	1	0	1	1	1	1	1	1	1	8	94,4	0,11
29	1	1	1	1	1	1	1	1	1	9	1	1	1	0	0	0	1	0	0	4	72,2	0,56
30	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9	100	0

Рис. 3. Результати аналізу “якості” тестових питань

Після нанесення підсумкових балів кожного студента на стандартну криву розподілу, яка при достатньо великій кількості даних (при двох ітераціях — 52 студенти і 60 тестових завдань) близька до кривої нормального розподілу, визначається відносна оцінка студента за чотирибальною шкалою (“відмінно”, “добре”, “задовільно” і “незадовільно”). Якщо остаточний бал знаходиться в межах  $(X + \sigma)$ , то можна ставити оцінку “добре”, в межах  $(X + 2\sigma)$  і більше, то стави-

ться оцінка “відмінно”, в межах  $(X - \sigma)$  — ставиться оцінка “задовільно”, а в межах  $(X - 2\sigma)$  і нижче — “незадовільно”.

Таким чином, використання процедури якісної оцінки тестових завдань дозволяє шляхом кількох послідовних ітерацій суттєво удосконалити тестові завдання при фінальному контролі рівня знань студентів з кожної дисципліни.

1. Булах І. Методи контролю та оцінювання рівня знань // Сучасні системи вищої освіти: порівняння для України: Наукове видання. — К.: KM Academia, 1997. — С. 169—185.

2. Інгенкамп К. Педагогічна діагностика. — М.: Педагогіка, 1991.

3. Державний стандарт вищої освіти України за спеціальністю “Екологія” (проект), 1998 р.

*Bogolyubov V. M., Zamostian V. P., Sereda L. I.*

## **ANALYSIS OF THE PROBLEM TESTS IN THE FINAL CONTROL OF ENVIRONMENTAL SCIENCES KNOWLEDGE LEVEL**

Investigations of quality characteristics of the problem tests in the final control of knowledge level in the environmental sciences have allowed determining the degree of reliability and objectivity of the problem tests. In addition, these investigations have helped to correct validity level of the problem tests in some disciplines that are lectured at the ecological department in National University of “Kyiv-Mohyla Academy”.