

РОЗРОБКА І ВИКОРИСТАННЯ МЕТОДУ ОЦІНКИ ТА КОНТРОЛЮ ЗНАНЬ СТУДЕНТІВ ТЕХНІЧНОГО УНІВЕРСИТЕТУ

* Павленко П. М., д-р техн. наук, доц., ** Нестеренко Ю. Г., ** Серков Є. О., ** Водяник В. В.

* Національний авіаційний університет, м. Київ

** Черкаський державний технологічний університет, м. Черкаси

nesterenko_chdtu@mail.ru

У статті розглянуто методику оцінки рівня засвоєння викладеного матеріалу студентами за допомогою математичного методу тестового контролю. За допомогою спеціалізованих комп'ютерних програм тестування запропоновано визначення коефіцієнтів надійності тестування з конкретної інженерної дисципліни для поліпшення якості навчання та контролю.

Ключові слова: програми тестування, тестовий контроль, математичні методи оцінювання знань, надійність виконання тестів.

In the article the method of estimation of level of mastering of the expounded material by students by means mathematical method of test control is examined. By means the specialized computer programs of testing determination of coefficients of reliability of testing in relation to concrete engineering discipline with the purpose of improvement of quality of teaching and control is offered.

Keywords: testing programs, test control, mathematical methods of evaluation of knowledges, reliability of implementation of tests.

Вступ

Процес навчання інформаційним технологіям стає ефективним за рахунок упровадження в навчальний процес програм тестування, які дають змогу за обмежений час контролю визначити рівень тестованих знань.

Розробка основних вимог до організації навчання можлива за умови використання під час розробки програми тестування адекватних математичних моделей процесу навчання і тестового контролю.

Мета та задачі дослідження

Метою дослідження є аналіз математичних методів оцінювання тестового контролю у технічному університеті, статистична обробка результатів до і після проведення тестування в академічних групах з підготовки інженерів. Одним з основних напрямів підвищення якості навчання є розв'язання таких задач:

- вибір математичного апарата для об'єктивної оцінки тестованих знань;
- зменшення витрат навчального часу при проведенні тестування;
- поліпшення контролю в системі навчання;
- оцінка рівня якості викладання конкретної технічної дисципліни.

Виклад результатів досліджень

Кожен модуль дисципліни характеризується вхідним, проміжним і вихідним контролем, який доцільно проводити за допомогою різних програм тестування. Контроль знань являє собою структуру вигляду

$$K_{\zeta} = K_{\text{вх}} \cup K_{\text{вих}}$$

де $K_{\text{вх}}$ — множина вхідних і проміжних контролів; $K_{\text{вих}}$ — множина вихідних контролів [1].

Вхідний та проміжний контроль характеризуються ідентифікатором дисципліни, показником приналежності до модуля, «забудькуватістю» матеріалу, пов'язаною зі складністю.

Вихідний контроль у своїй структурі передбачає ідентифікатор дисципліни, показник приналежності до модуля, показник якості розуміння матеріалу. Модель взаємодії різних видів контролю визначають за допомогою виразу

$$K_{\text{вих}} \Rightarrow K_{\text{вх}}$$

тобто множина вихідних контролів посиляється на множину вхідних і проміжних.

Для кожного вихідного контролю має визначатись попередній контроль $K_{\text{вх}} \Rightarrow K_{\text{вих}}$.

Під час проведення тестового контролю існує поняття складності виконання завдання

$$q_i = \frac{N_i}{N_{\delta}}$$

де N_i — кількість неправильних відповідей; N_{δ} — кількість тестованих людей.

У теорії тестів і на практиці тестування, як правило, розглядають два основних критерії якості тесту — надійність і валідність. Надійність тестів асоціюють за точністю вимірювання тестованих знань, а валідність — за адекватністю, обґрунтованістю і професіоналізмом питань тесту. Під час вимірювання надійності тестів передбачається лінійна модель, у якій дисперсія оцінки дорівнює сумі істинного та помилкового значень. Дисперсія випадкової величини — це міра розкиду випадкової величини, тобто її відхилен-

ня від математичного сподівання, яке, у свою чергу, є мірою середнього значення випадкової величини в теорії вірогідності.

Розглянемо декілька методів оцінки надійності виконання тестів.

Формула Д. Армора визначає коефіцієнт «збудькуватості» матеріалу, пов'язаний зі складністю тесту

$$\theta = \frac{n}{n-1} \left(1 - \frac{1}{\lambda_1} \right),$$

де λ_1 — найбільше значення кореня характеристичного рівняння $|R - \lambda_1| = 0$.

Д. Армор використовував відомий факт кореляції тестових висловів між собою і став розглядати її як аргумент, статистичною функцією якого є надійність тесту. Якщо всі вислови вимірюють одну й ту ж ознаку або властивість, то для фіксованого їх числа чим більше кореляція (статистичний взаємозв'язок двох або кількох випадкових величин) між ними, тим надійніший тест. З іншого боку, висока кореляція забезпечує об'єктивну оцінку кореляційної матриці R і, отже, є підставою виділення такого чинника, який зможе пояснити зв'язок більшої частини дисперсії в R . Отже, надійність тестів повинна бути пов'язана з результатом факторного аналізу.

До недоліків цього методу належать те, що коли тестові завдання розрізняються за складністю, важко отримати об'єктивний результат оцінювання тестованих знань.

Більш формалізованими є тести на основі *Item Response Theory* (IRT), у яких передбачається існування аналітичних виразів для задання вірогідності правильних відповідей тестованих. Тести IRT засновані на математико-статистичних моделях. Вони мають ряд переваг порівняно з іншими методами оцінки надійності тестів, а саме — об'єктивність оцінювання рівня знань тестованих і параметра складності завдання за однією й тією ж шкалою. У тестах IRT розглядають умовну вірогідність правильного виконання тестованих із певним рівнем знань різних за складністю завдань тесту. Основою математичної моделі IRT є однопараметрична логістична функція Р. Раша, яку називають характеристичною кривою. Вона має вигляд

$$P_j = \frac{1}{1 + \exp(\delta_j - \theta)},$$

де P_j — вірогідність правильної відповіді тестованих з будь-яким рівнем підготовки при певному рівні складності виконання завдання; δ_j — рівень складності конкретного j -го завдання тесту; θ — рівень знань, латентна змінна.

Латентними якостями називають позитивні й негативні якості особи, які не піддаються безпосередньому вимірюванню. Прикладом латентних якостей є «підготовка студентів», «знання навчальної дисципліни», «здатність розуміти», «інтелектуальний розвиток» та ін.

Датський математик Г. Раш увів два поняття — «логіт рівня знань» та «логіт рівня складності завдання». Перше поняття він визначив як натуральний логарифм частки правильних відповідей, тестованих на всі завдання тесту до частки неправильних відповідей; а друге поняття як натуральний логарифм іншого відношення, а саме, частки неправильних відповідей завдання тесту до частки правильних відповідей на ті ж завдання.

Формула Кудера–Річардсона визначає коефіцієнт надійності при однократному тестуванні

$$K_{kr} = \frac{n}{n-1} \left(1 - \frac{\bar{x}(n - \bar{x})}{n(S^2)} \right),$$

де K_{kr} — коефіцієнт достовірності тесту; n — кількість завдань тесту; \bar{x} — середня кількість правильно виконаних завдань; S^2 — розбіжності в правильно виконаних завданнях.

Коефіцієнт достовірності тесту K_{kr} можна представити як мультиплікативну похибку правильності застосування тесту для визначення певної групи тестованих знань. Зменшення цього коефіцієнту показує на стабільність результатів тестування, а також на зменшення кількості випадкових відповідей під час тестування.

Ознайомившись з різними програмами тестування, які успішно застосовуються у ВНЗ України та Росії, на факультеті комп'ютеризованих технологій машинобудування ЧДТУ було розроблено декілька тестів по інженерним дисциплінах:

1. Програма тестування з дисципліни «Взаємозамінність, стандартизація і технічні вимірювання» (ВСТВ) (рис. 1). Середовище розробки тесту — IDE Borland C++ 6.0. Мова програмування — C++.

Питання та відповіді в програмі тестування сформовано у вигляді картинок (файлів із розширенням *.jpg), що скорочує час редагування і доповнення бази питань та відповідей. Довідкова інформація до тестів (за необхідності) додається у вигляді сканованих сторінок довідників. Результати тестування подано у вигляді таблиці з можливістю відстеження кількості балів і дати проходження тестів. Крім того, у цій програмі тестування реалізовано режим пошуку тестованих знань. Для кожного питання сформовано чотири відповіді, одна з яких є правильною. Тест складається з 10 варіантів, у кожному з яких 10 питань.

Час проходження кожного питання обмежений, правильність відповіді на питання підтверджується індикатором зеленого кольору.

Максимальна кількість балів при відповіді на всі питання дорівнює 100 балів. Після проходження тестування в програмі передбачено ре-

жим перегляду та аналізу неправильних відповідей.

2. Програма тестування з дисципліни «Метрологічне забезпечення наукових досліджень» (МЗНД) (рис. 2). Середовище розробки тесту — Lazarus. Мова програмування — Free Pascal

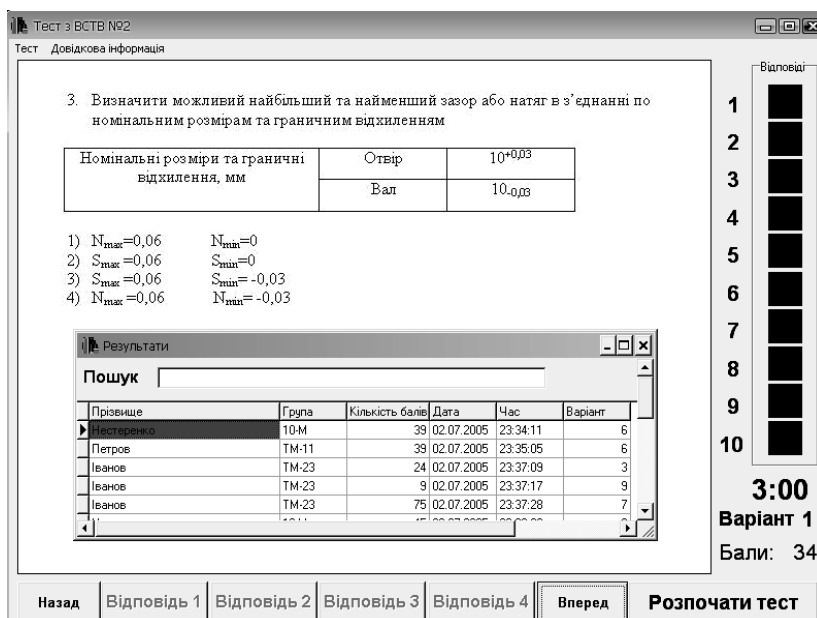


Рис. 1. Інтерфейс програми тестування з дисципліни «Взаємозамінність, стандартизація і технічні вимірювання (ВСТВ)»

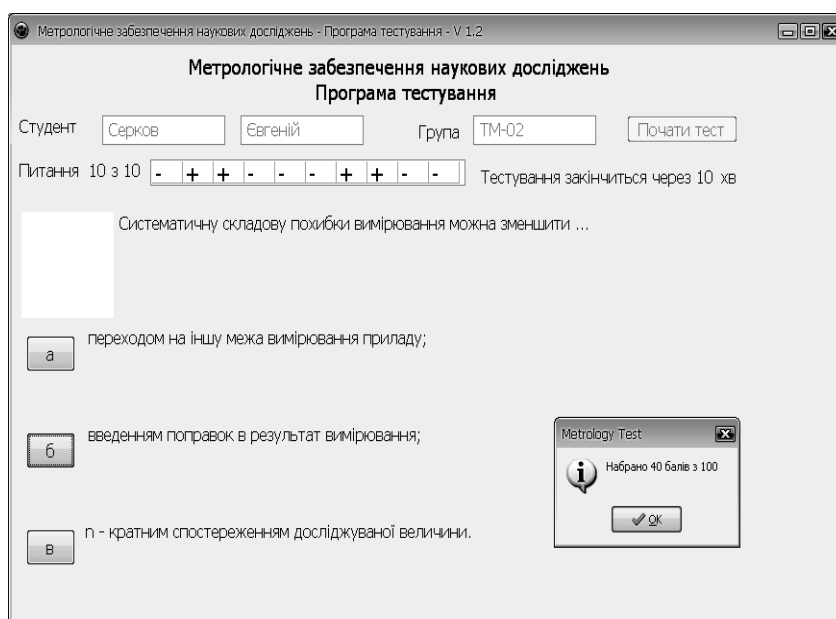


Рис. 2. Інтерфейс програми тестування з дисципліни «Метрологічне забезпечення наукових досліджень (МЗНД)»

Питання та відповіді сформовано у вигляді тексту, за необхідності до питань додаються картини. 10 питань вибирають випадковим чином із 150 можливих питань із метрології [2].

Правильна відповідь на питання підтверджується знаком «+», а неправильна «-».

Загальний час проходження тесту 10 хв.

Для кожного питання передбачено три відповіді, одна з яких є правильною. Результати тестування оцінюють за стобальною шкалою.

До переваг цього тесту належать: випадковий вибір питань при кожному запуску програми те-

стування; простота оформлення, відсутність будь-яких кольорових або динамічних об'єктів, що дає можливість зосередитися на пошуку правильних відповідей; відсутність можливості пропускати питання та повертатися до них через певний час.

Аналіз результатів досліджень

Для збору інформації про засвоєння викладачем матеріалу з дисципліни «МЗНД» було використано розроблену програму.

Тестування проводилося в групі з 20 студентів, яких вибрали з кількох академічних груп. Вхідний контроль знань проводився зі студентами, які не вивчали дисципліну «МЗНД», але мали базовий, несистематизований рівень знань, отриманий на попередніх курсах та в школі: з

фізики, електротехніки, основ вимірювання фізичних величин.

Вихідний контроль знань проводився з тими самими студентами після вивчення дисципліни «МЗНД». Результати тестування вхідного та вихідного контролю наведено в табл. 1.

Частка кількості набраних балів у процесі вивчення дисципліни подано в табл. 2.

Середній набраний бал по вхідному контролю знань становить 23 бали, а по вихідному — 67 балів. За результатами досліджень було побудовано графік (рис. 3), з якого видно, що екстремуми функції вихідного контролю змістилися в бік покращення результатів тесту на 30–40 %, що свідчить про краще засвоєння навчального матеріалу.

Таблиця 1

Результати тестування

Вхідний контроль знань																				
Студент	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Бали	30	30	0	10	40	40	20	30	50	10	10	50	0	0	10	40	40	0	40	10
Вихідний контроль знань																				
Студент	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Бали	70	40	70	50	90	50	100	40	60	60	100	70	80	60	70	60	80	80	70	40

Таблиця 2

Частка набраних балів

Кількість балів		0	10	20	30	40	50	60	70	80	90	100
Частка, %	Вхідний контроль знань	20	25	5	15	25	10	0	0	0	0	0
	Вихідний контроль знань	0	0	0	0	15	10	20	25	15	5	10

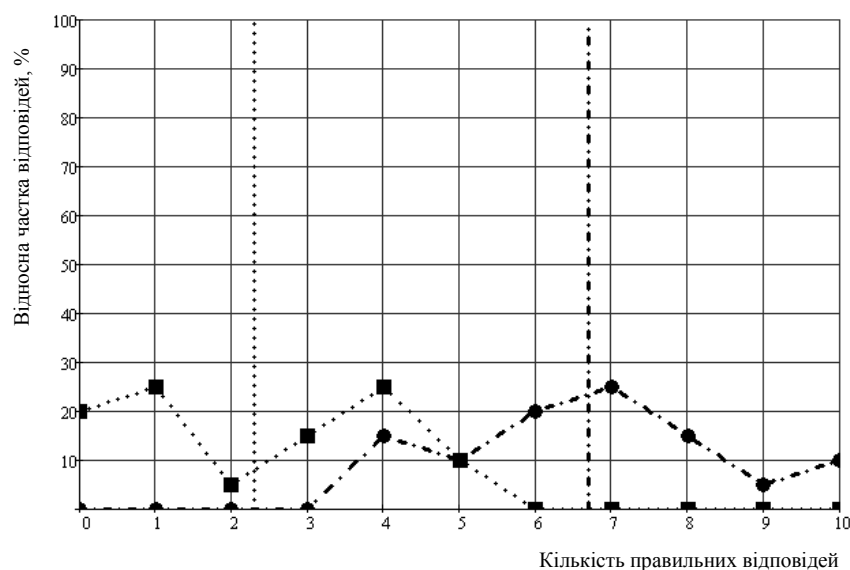


Рис. 3. Залежність кількості правильних відповідей від відносної частки відповідей:

- ■ — частка набраних балів при вхідному тестуванні, %;
- ● — частка набраних балів при вихідному тестуванні, %;
- ♦ ♦ ♦ ♦ — середній бал вхідного тестування; ♦ ♦ ♦ ♦ — середній бал вихідного тестування

Обробку результатів тестування було виконано в математичному пакеті *MathCAD*.

Вищевказані екстремуми виявили існування двох підгруп студентів, які розрізняються за ступенем засвоєння матеріалу. Величина першого максимуму стала відносно другого максимуму меншою, що свідчить про наполегливість у навчанні відстаючих студентів.

Середньоквадратичне відхилення кількості правильно виконаних завдань становить для вхідного тестування $S = 1,74$, а для вихідного — $S = 1,76$.

Коефіцієнт надійності для вхідного тестування становить $K_{kr} = 0,46$, а для вихідного — $K_{kr} = 0,32$, що свідчить про зменшення кількості випадкових відповідей приблизно в 1,5 разу.

Висновки

У результаті впровадження методу оцінки та контролю знань студентів технічного університету встановлено:

1. Якість навчання студентів значно підвищилась (приблизно в 3–4 рази) за рахунок впровадження тестового контролю в навчальний процес.

2. Найбільш ефективною для оцінки надійності тесту є формула Кудера–Річардсона, яка дала змогу обробити та проаналізувати результати перевірки рівня тестованих знань до надання теоретичної інформації з дисципліни «МЗНД» та після отримання інформації. Студенти після проходження дисципліни «МЗНД» значно підвищили відносну частку відповідей (з 23 до 67 %), що свідчить про правильність методики викладання дисципліни.

3. Вищевикладену методику оцінки тестованих знань пропонується впроваджувати в кожному кредитному модулі інших технічних дисциплін.

ЛІТЕРАТУРА

1. Павленко П. М. Модель процесу отримання рівня кваліфікації фахівців по впровадженню інформаційних технологій / Вісник Чернігівського державного технологічного університету: зб. наук. пр. / П. М. Павленко, Ю. В. Задонцев. — Чернігів : ЧДТУ, 2010. — № 45. — С. 139–144.

2. Ординарцева Н. П. Метрологія + стандартизація + сертифікація: практикум / под общ. ред. А. А. Данилова. — Пенза : Пензенский государственный университет, 2008. — 52 с.

Стаття надійшла до редакції 15.12.2010.