

ВАЛІДНІСТЬ ТЕСТОВИХ ЗАВДАНЬ ІЗ ВИЗНАЧЕНИМИ КОЕФІЦІЄНТАМИ СКЛАДНОСТІ

Національний авіаційний університет

Досліджено технологію визначення валідності тестового завдання в залежності від коефіцієнтів складності питань та їх форми. Доведено, що тестові завдання комплексної форми валідні та більш адекватно характеризують рівень знань людини, що навчається

Вступ

Модернізація сучасної освіти заключається у підвищенні якості знань тих, хто навчається, що, в свою чергу, залежить не тільки від учбової роботи педагогів, але й від якості тестових завдань (ТЗ), необхідних для перевірки знань. В той же час, у зв'язку з вступом багатьох провідних вищих навчальних закладів країни до Болонської системи освіти, проведення контролю, без застосування тестування, взагалі неможливе. Тому, при оцінці знань досить широко використовують таку форму контролю, як тестування [1].

Для створення тестів необхідна математична обробка результатів тестування та визначення якості систем діагностики знань [2]. Оскільки, тест являє собою вимірвальний інструмент, то необхідно звернути увагу на те, що процес вимірювання в інформаційних системах ускладнений в зв'язку з відсутністю еталонних одиниць. Тому, необхідна максимальна відповідність даного тесту основним критеріям, що визначають якість методу: надійність, валідність та точність.

Переважає більшість існуючих систем контролю знань проводить лише оцінку знань і не проводить жодних оцінок якості самого тесту, в них відсутня можливість корегування тестових завдань (ТЗ), не враховується таке поняття, як складність питань ТЗ. Відомо [1], що безпідставне збільшення складності ТЗ призводить до збільшення часу на проведення тестування, що в свою чергу, призводить до швидкого стомлення студентів, а

це збільшує кількість помилок і в значній мірі знижує об'єктивність оцінювання.

Таким чином завдання створення та дослідження надійності ТЗ, на основі яких стане можливою побудова сучасної автоматизованої системи контролю знань є доволі актуальним.

Аналіз досліджень і публікацій

Аналіз літератури [2, 4, 6] дозволяє виділити дві найбільш поширених форми ТЗ: закрыта та відкрита.

ТЗ закрытої форми складається із питання та варіантів відповіді, до складу якої входить правильна (еталонна). На даний час поширення набуває ТЗ відкритої форми, що являє собою сукупність завдань із довільно сконструйованою відповіддю. В цих системах найчастіше потрібне повне співпадіння із еталоном, або проведення семантичного аналізу відповіді. Як правило, студент отримує 1 за правильне виконання завдання, а за неправильне 0. Такі системи оцінювання не може надати повної інформації о знаннях студента, тому що, в такому випадку, необхідно велика кількість питань, що збільшує кількість помилок. Отже, при розробці ТЗ необхідно враховувати їх складність питань. При цьому найкраще використовувати обидві форми питань.

Таким чином, дослідження надійності ТЗ із попередньо розрахованими коефіцієнтами складності питань представляє науковий інтерес.

Постановка завдання

Дослідити валідність ТЗ із попередньо розрахованими коефіцієнтами складності.

Дослідження надійності ТЗ

Розглянемо методику розрахунку вагових коефіцієнтів (індексів складності) w_j , де s – кількість питань, $i = \overline{1, s}$, що визначають достовірність результатів контролю знань, яка запропонована в [3]. За допомогою вагових коефіцієнтів враховується ступінь складності ТЗ по відношенню до інших, задається максимальна кількість балів, яку може отримати студент при правильному виконанні завдання. А це у свою чергу визначає кількісну та якісну характеристику знань. Для покращення достовірності результатів тестування при проектному розрахунку пропонується визначати вагові коефіцієнти питань у ході попереднього експериментального тестування.

Проведемо аналіз ТЗ із попередньо розрахованими коефіцієнтами складності питань комбінованої форми.

Тестування [6] – науково обґрунтований процес вимірювання якості властивостей особистості. При цьому тест є інструментом, що складається з кваліметрично вивереної системи ТЗ, стандартизованої процедури проведення, із заздальгідь спроектованої технології обробки та аналізу результатів. ТЗ повинні відповідати певним критеріям якості (рис. 1).

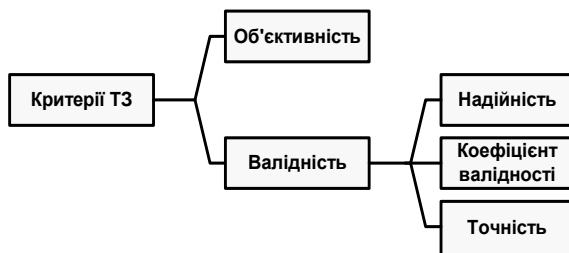


Рис. 1. Критерії якості ТЗ

Серед критеріїв виділяють найважливішу властивість: валідність, що характеризує ступінь відповідності тесту своєму призначенню вимірювати те, для чого він створений, а також дійсна здатність вимірювати ту характеристику, для якої він заявлений [6]. Валідність визначається за трьома критеріями.

Надійність тесту розглядається як характеристика, що відображає точність вимірювання та стійкість результатів до

дії сторонніх факторів [4]. Тобто, під надійністю розуміють узгодженість результатів проведення тесту на одній і тій самій дослідній групі при різних педагогічних умовах.

Існує кілька понять надійності та методів її визначення:

- надійність паралельних форм;
- ретестова надійність;
- надійність розщеплених частин тесту.

Надійність паралельних форм визначається за допомогою проведення тестування на одній групі студентів із використанням ідентичних варіантів ТЗ. Але, як правило, розробка ідентичних варіантів ТЗ та повторне тестування із однаковими умовами практично не можливе бо виникає ефект впізнавання.

Визначення надійності при повторному тестування через певний проміжок часу називається ретестовою надійністю. Та така надійність не дає об'єктивної оцінки.

Тому найбільш поширеним методом визначення надійності є аналіз розщеплених частин ТЗ. Тести, що задовольняють умовам надійності називаються аутентичними.

Проведемо аналіз надійності ТЗ. Нехай, числова оцінка $ball_{ij}$ успішності виконання j -го завдання i -им студентом, де $j = \overline{1, n}$, $i = \overline{1, s}$, n та s – кількість студентів та питань відповідно. Представимо результати тестування у вигляді матриці $\{ball_{ij}\}$ з n рядками та s стовпцями. Матриця результатів показує виконання всіх завдань тими, хто навчається.

Розглянемо методику розрахунку критеріїв валідності ТЗ, яка полягає в статистичній обробці результатів та розрахуємо надійність ТЗ.

Знайдемо індивідуальні бали кожного студента $ball_i^{ind}$ (1).

$$ball_i^{ind} = \sum_{j=1}^s ball_{ij}. \quad (1)$$

Знайдемо середнє значення сумарних балів \overline{ball} (2) учасників тестування та середній результат студентів за кожне завдання (3):

$$\overline{ball} = \frac{\sum_{i=1}^n ball_i^{ind}}{n}, \quad (2)$$

$$\overline{ball}_j^{ind} = \frac{\sum_{i=1}^n ball_{ij}}{n}. \quad (3)$$

Для розрахунку надійності тестового завдання необхідно знайти стандартне відхилення d_{ball} (4) сумарних балів учасників тестування та величину d_j^2 – дисперсію результатів студентів по j -ому завданню (5), де $j = \overline{1, s}$.

$$d_{ball} = \sqrt{\frac{\sum_{i=1}^n (ball_i^{ind} - \overline{ball})^2}{n-1}}, \quad (4)$$

$$d_j^2 = \frac{\sum_{i=1}^n (ball_{ij} - \overline{ball}_j^{ind})^2}{n-1}. \quad (5)$$

Визначимо коефіцієнт кореляції Пірсона (6) для тих номерів питань, коефіцієнти складності яких однакові:

$$R_j = \frac{\sum_{i=1}^n (ball_{ij} \times ball_i^{ind})}{n} - \overline{ball}_j^{ind} \times \overline{ball} \times \frac{n}{n-1} \times \frac{1}{d_j \times d_{ball}}. \quad (6)$$

Надійність ТЗ тим вища, чим вища погодженість результатів тестування окремих частин одного й того ж ТЗ. Погодженість питань із однаковими коефіцієнтами складності визначається коефіцієнтом Кьюдера-Річардсона (7):

$$\beta = \frac{d_{ball}^2 - \sum_{j=1}^s d_j^2}{2d_{ball}^2} + \sqrt{\left(\frac{d_{ball}^2 - \sum_{j=1}^s d_j^2}{2d_{ball}^2} \right)^2 + \frac{\sum_{j=1}^s R_j^2 d_j^2}{2d_{ball}^2}}. \quad (7)$$

Коефіцієнт надійності Кьюдера-Річардсона [4] дає об'єктивний результат у випадку, коли припущення, що питання із однаковими коефіцієнтами складності ідентичні, є правдивим. В іншому випадку розрахуємо коефіцієнт надійності α за формулою (8):

$$\alpha = \frac{s}{s-1} \left[1 - \frac{\sum_{j=1}^s \sigma_j^2}{\sigma_{ball}^2} \right], \quad (8)$$

де s - кількість питань ТЗ, σ_{ball}^2 - квадрат стандартного відхилення для всього ТЗ (9), σ_j^2 - квадрат стандартного відхилення результатів тестування для кожного питання (12).

$$\sigma_{ball}^2 = \frac{1}{n-1} \sum_{i=1}^n (ball_i' - ball^s)^2, \quad (9)$$

де $ball_i'$ - результат i -го студента (10), $ball^s$ - середнє значення за все ТЗ (11), n - кількість студентів.

$$ball_i' = \frac{\sum_{j=1}^s ball_{ij}'}{s}, \quad (10)$$

$$ball^s = \frac{\sum_{i=1}^n ball_i'}{n}. \quad (11)$$

Розрахуємо квадрат стандартного відхилення для всього ТЗ:

$$\sigma_i^2 = \frac{1}{n-1} \sum_{i=1}^n (ball_{ij} - ball_j^s)^2. \quad (12)$$

Для того, щоб об'єктивно оцінити знання людини, що навчається, необхідна максимальна точність, що визначається як мінімально припустима помилка методу: $\varepsilon = \sigma / \sqrt{NM}$, де σ – середньоквадратичне відхилення результатів тестування для всіх студентів, N – обсяг вибірки, M – математичне очікування.

Порівняльний аналіз валідності ТЗ

Проведемо аналіз системи діагностики знань. Під час тестування студентам були запропоновані питання із відкритою формою відповіді, тобто необхідно ввести свою відповідь. Якщо відповідь студента невірна, то йому надавалася можливість обирати одну із запропонованих варіантів відповідей. При цьому в системі вже визначені коефіцієнти згідно методики [3], але паралельно розраховувалися результати

тестування у дихотометричній системі, а саме, студент отримував 0 або 1 в залежності від відповіді (рис. 2). Під час розрахунків кількості балів $ball$ за кожне питання враховувалися коефіцієнти складності типів питань w' та w'' відповідно для відкритої та закритої форми питань. Як правило $w' < w''$ та $w' = 1$, $w'' = 0.8$, але можуть бути визначені експертом даної області знань.

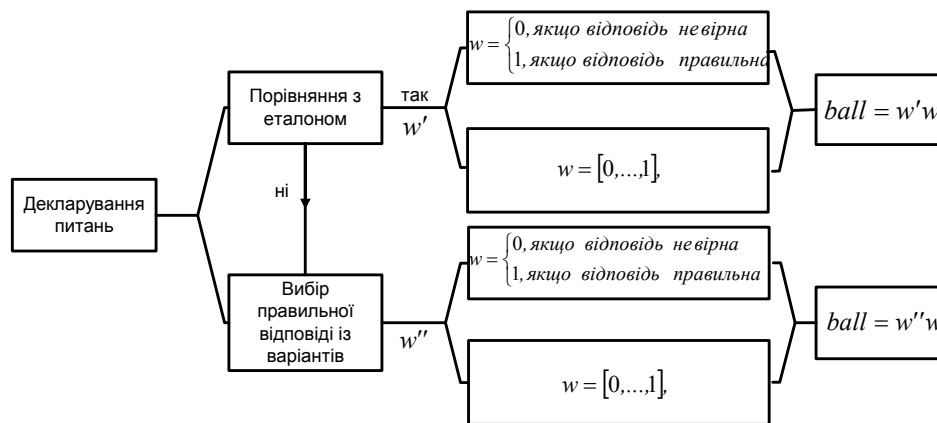


Рис. 2. Схема декларування питання при тестуванні

Розглянута методика діагностики знань студентів за допомогою ТЗ із визначеними ваговими коефіцієнтами використовує експериментальні дані, а отже є емпіричною системою, для корегування якої необхідно невелика кількість дослідних тестувань. Для дослідження адекватності методики було проведено тестування кількох груп студентів із різним рівнем підготовки. На практиці при проведенні тестування значення результатів тестування, а саме, кількості балів отриманих за відповіді, змінювалися наступним чином (рис. 3). Тобто, I та II – результат ТЗ закритої та відкритої форми питання дихотометричній системі відповідно. Як видно графік III для $w = [0, \dots, 1]$ у комплексній формі ТЗ має більш адекватні дані.

За результатами тестування були розраховані надійність β (8) та α (9) ТЗ, загальна складність \bar{w} ТЗ та кількість s питань.

Таким чином, як видно із табл. 1 за статистичними параметрами використання обох форм тестових питань із визначеними коефіцієнтами складності за запропонованою методикою є валідним. Для визначення достовірного рівня знань необхідна менша кількість питань із визначеними коефіцієнтами.

Проведемо експериментальну перевірку підпорядкованості результатів тесту нормальному закону. Для цього за результатами тестування побудуємо графік щільності розподілу $\rho(ball_{ind})$ осіб, що отримали відповідну кількість балів $ball_{ind}$ (рис. 4).

Розглянемо ступінь відхилення результатів тестування від симетричного виду відносно середнього значення. У випадку, якщо $w = [0, \dots, 1]$ у комплексній формі ТЗ ступінь зсуву графіка розподілу частот від симетричного виду $As \rightarrow 0$ при мірі плосковерхості та гостроверхості $Ex \rightarrow 0$ (I). А це означає, що ТЗ задово-

льняє умові валідності. В той же час, для закритої форми ТЗ при тому ж коефіцієнті складності k питання $As < 0$ та $Ex < 0$ (II). Тобто, тест надто простий.

Для відкритої форми питань ступінь зсуву графіка розподілу частот від симетричного виду $As > 0$ при мірі плосковерхості та гостроверхості $Ex > 0$, тобто тест надто складний (III).

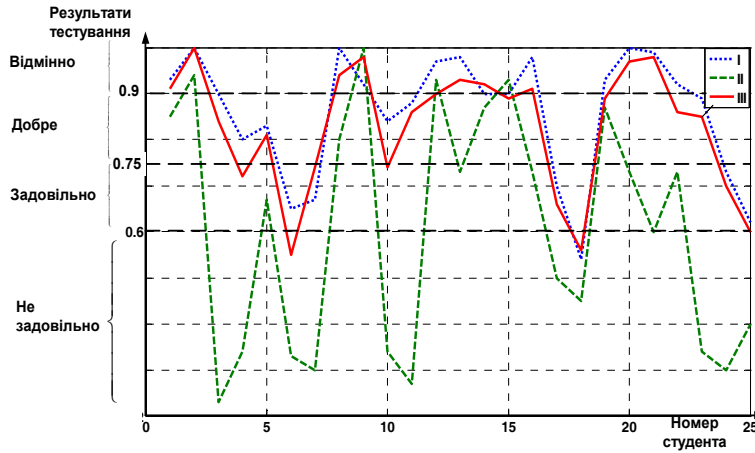


Рис. 3. Результати тестування групи студентів

Таблиця 1.

Вид ТЗ	Коефіцієнт складності w	Надійність β ТЗ	Надійність α ТЗ	Загальна складність \bar{w}	Кількість s питань
Відкрита форма	$w = \begin{cases} 0, \text{ якщо відповідь невірна} \\ 1, \text{ якщо відповідь правильна} \end{cases}$	0.7	0.68	0.82	15
Закрита форма		0.68	0.65	0.5	15
Комплексна форма	$w = [0, \dots, 1]$	0.89	0.84	0.7	15
Нормовані значення		> 0.8	> 0.7	$0.35 \leq w \leq 0.84$	≤ 30

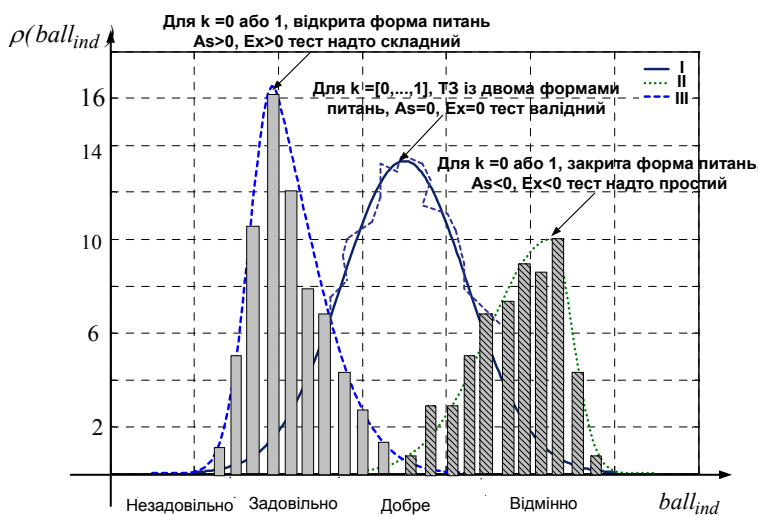


Рис. 4. Графік щільності розподілу $\rho(ball_{ind})$ осіб, що отримали відповідну кількість балів $ball_{ind}$

Для визначення оптимальної кількості питань необхідно враховувати складність кожного питання, складність тесту взагалі. Кількість питань необхідно обирати таким чином, щоб виконувалась умова

$$\varepsilon^H \cdot w_{em} \leq \bar{w} \leq \varepsilon^B \cdot w_{em},$$

де ε^B та ε^H – відповідно верхня та нижня довірчі границі вагових коефіцієнтів, розраховані в [3], \bar{w} - загальна складність тесту, w_{em} - складність форми ТЗ, тобто $w_{em} = \{w', w''\}$.

Структура множини питань повинна однозначно відповідати структурі множини дидактичних знань, а саме, повинна бути інструментом, що дозволяє оцінити рівень знань об'єктів навчання за основними поняттями.

В процесі дослідження результатів учбової групи знайдено, що при використанні ТЗ комплексної форми достатньо кількості s питань, яка дорівнює 15 (I), в той же час, питань закритої форми - $s = 23$ (II), а для відкритої - $s = 28$ (III).

Висновки

Таким чином, досліджено технологію визначення валідності ТЗ в залежності від попередньо розрахованих коефіцієнтів складності питань.

Проведено порівняльний аналіз валідності відкритої та закритої форм питання дихотометричної тестової системи, із комплексною формою ТЗ, з урахуванням коефіцієнтів складності питань згідно запропонованої методики.

Доведено, що ТЗ із урахуванням вагових коефіцієнтів питань валідні, а це в свою чергу дозволяє адекватно характеризувати складність тестових завдань та рівень підготовки об'єктів навчання. При цьому така методика не вимагає достатньо великої кількості експериментальних даних на відміну від інших відомих параметричних моделей.

Врахування коефіцієнтів складності у ТЗ із комплексною формою дозволяє зменшити кількість питань при дотри-

манні валідності у необхідних межах, а це, в свою чергу, зменшує час підготовки ТЗ та проведення тестування.

Список літератури

1. Аванесов В.С. Основы научной организации педагогического контроля в высшей школе /Аванесов В.С. – М.: МИ-СиС, 1989. – 168 с.
2. Олійник М.М. Тест як інструмент кількісної діагностики рівня знань в сучасних технологіях навчання /Олійник М.М. – Донецьк: ДонНУ, 2001. – 84 с.
3. Шибицька Н.М., Тимофієва Г.А. Інформаційна технологія визначення складності питань в тестових завданнях // Проблеми інформатизації та управління. Зб. нук. пр. – К.: НАУ, 2009. – Вип. 2 (26) – С. 152–156.
4. Федорчук П.І. Адаптивні тести: статистичні методи аналізу результатів тестового контролю знань / Федорчук П.І. // Математичні машини і системи. – К.: 2007. – №3,4. – С. 122–138.
5. Волков В.І., Алексєєв О.М., Алексєєв М.О. Тестовий контроль знань: Навчальний посібник / Волков В.І., Алексєєв О.М., Алексєєв М.О. – Суми: ВТД «Університетська книга», 2004. – 100 с.
6. Долинер Л.И. Компьютерные тесты успеваемости как средство оптимизации учебного процесса / Долинер Л.И. // Вестник Московского университета. – 2004. – №1.
7. Присняков В.Ф., Приснякова Л.М. Математическое моделирование переработки информации оператором человеко-машинных систем – М.: Машиностроение, 1990. – 248 с.
8. Шибицкая Н.Н. Метод идентификации объектов в эргатической системе управления процессом обучения // Кибернетика и вычислительная техника. – Вып. 121. – К.: 1999. – С. 52–58.

Подано до редакції 09.02.10