

ОБЧИСЛЮВАЛЬНА ТЕХНОЛОГІЯ ВИЗНАЧЕННЯ СТАНДАРТУ ОЦІНЮВАННЯ ДЛЯ КРИТЕРІАЛЬНО-ОРІЄНТОВАНИХ ТЕСТІВ

У роботі запропоновано обчислювальну технологію знаходження критеріального балу для критеріально-орієнтованих тестів на основі методу множинного порівняльного аналізу експертних оцінок складності тестових завдань з використанням коефіцієнту конкордації. Метод вибору тестового стандарту базуються на експертних оцінках складності завдання. У випадку залучення групи експертів для оцінки змісту тестових завдань пропонується перевіряти узгодженість їх рішень за допомогою методу множинного порівняльного аналізу з використанням рангового коефіцієнту кореляції Кендалла.

Ключові слова: оцінка навчальних досягнень, критеріально-орієнтований тест, стандарт оцінювання, експертна оцінка якості тестових завдань, множинний порівняльний аналіз, коефіцієнт конкордації.

Тестування навчальних досягнень є важливою складовою частиною навчального процесу. Педагогічне тестування широко застосовується для контролю знань в різних цілях. За цілями застосування педагогічні тести поділяються на два класи – нормативно-орієнтовані і критеріально-орієнтовані.

Нормативно-орієнтований педагогічний тест (norm-referenced test) дозволяє порівнювати навчальні досягнення (рівень підготовки, рівень професійних знань і умінь) окремих випробовуваних один з одним. Критеріально-орієнтований педагогічний тест (criterion-referenced test) дозволяє оцінювати, якою мірою випробовувані оволоділи необхідним навчальним матеріалом.

Нормативно-орієнтовані педагогічні тести використовуються для того, щоб отримати надійні і нормально розподілені бали для порівняння тестованих. Критеріально-орієнтовані педагогічні тести застосовуються для того, щоб інтерпретувати результати тестування відповідно рівнем навченості випробовуваних. Істотне значення набуває використання критеріально-орієнтованих тестів в системі вищої освіти для поточної, проміжної і підсумкової атестації студентів. Викладачі навчальних закладів можуть використовувати результати критеріально-орієнтованих тестів, щоб правильно оцінити рівень підготовки студентів, а також визначити якість тих чи інших програм навчання.

Задача обробки даних педагогічного тестування та контролю знань за критеріально-орієнтованою методикою в роботах [1–4] розглядається як задача підтримки прийняття рішень при класифікації об'єктів навчання (тестованих) за рівнем підготовки, для вирішення якої використовуються методи теорії статистичних рішень (критерій Байеса, Неймана-Пірсона, мінімакса, Вальда). Застосування даних методів спрямоване на створення систем статистичного контролю знань за альтернативною ознакою.

На сьогодні розрізняють два основних види критеріально-орієнтованих тестів, що відрізняються один від одного окремими характеристиками – це тести орієнтовані на визначення рівня індивідуальних досягнень студентів та кваліфікаційні тести.

Кваліфікаційні тести (mastery tests – тести на майстерність) – використовуються для визначення в групі осіб, що проходять тестування, кваліфікованих або некваліфікованих у відповідній області [5]. Цей вид тестування базується на заздалегідь вибраній нормі оцінювання – критеріальному балі. Найчастіше такі тести використовуються для розділення групи студентів на підгрупи (дві або більше) в залежності від рівня засвоєння необхідного об'єму знань у відповідній змістовній області. Наприклад, розділення на підгрупи: тих, хто повністю оволодів відповідними знаннями, тих, яким належить засвоєнні знання додатково поглибити або розширити, тих, кому необхідно повторити вивчення даної змістовної області. Таким чином, для проходження кваліфікаційного тесту студент має досягти необхідного мінімального критерію (стандарту) оцінювання. На підставі цього, встановленого розробниками тесту критерію, приймається рішення відносно кожного з тих, хто тестується. В тесті даний критерій, за звичай, відображається визначеною кількістю правильно виконаних завдань.

Методика встановлення критерію оцінювання кваліфікації за результатами тестування, методика оцінки рівня помилки цього критерію та порядок прийняття рішення щодо кваліфікації або недостатній підготовленості студента є найважливішими елементами, що мають прийматися до уваги під час розроблення кваліфікаційних тестів.

Таким чином під час ідентифікації рівня знань учасників навчального процесу з використанням методу послідовного аналізу для визначення ймовірності невиконання завдань p_i , $i = \overline{1, n}$ для певного класу (1..n) постає задача вибору стандарту оцінювання q_i (ймовірності вірно виконаних завдань для кожного класу).

Методи вибору тестового стандарту базуються на експертних оцінках змісту тестового завдання. На основі дослідження даних методів, описаних у роботах [6–8], для чотирьохальтернативної задачі класифікації тестованих за рівнями навченості: «початковий» (I клас), «середній» (II), «достатній» (III) та «високий» (IV) обрано метод Ангофф, який заснований на послідовних експертних оцінках змісту тестових завдань. Спочатку виконується вибір стандарту оцінювання для I класу. Експерт-викладач для кожного завдання тесту встановлює ймовірність того, що мінімально компетентний студент дасть на нього вірну відповідь. Для однозначності експерту (або групі експертів) пропонується обрати значення ймовірності P_i зі значень 0,9, 0,8, ..., 0,1. Визначивши суму значень даних

ймовірностей отримаємо критеріальний бал: $K = \sum_{i=1}^n P_i$, де n – загальна кількість завдань у тесті. Ста-

ндарт оцінювання визначається за формулою: $q = \frac{K}{n}$. Тоді ймовірність невиконання завдань $p = 1 - q$.

Після вибору критеріального балу для «початкового» рівня навченості, експерт проводить оцінку кожного тестового завдання вже на більш високий стандарт «середній», «достатній» та «високий». Для цього експерт повторює для кожного завдання описану вище процедуру, але вже присвоює ймовірності 0,9, 0,8, ..., 0,1 для мінімально компетентних «трієчників», «хорошистів», «відмінників». Знаходження критеріального балу для «хорошистів» можна виконати шляхом визначення напівсуми критерійних балів для «відмінників» та «трієчників».

У випадку залучення групи експертів для оцінки змісту тестових завдань пропонується перевіряти узгодженість їх рішень за допомогою методу множинного порівняльного аналізу експертних оцінок з використанням коефіцієнту конкордації. Відповідно до робіт [8–10] формується обчислювальна процедура підтримки прийняття узгодженого рішення, яка містить наступні етапи:

1. Задаються значення змінних $t = 1, k_j^0 = \frac{1}{m}, j = \overline{1, m}$, де k_j^0 – початкові значення компетентності експертів; m – кількість експертів.

2. Обчислюються значення x_i^t та записують у матричній формі:

$$x_i^t = \sum_{j=1}^m x_{ij} k_j^{t-1}, \quad i = \overline{1, n}, \quad (1)$$

де x_{ij} – значення оцінки j -м експертом i -го запитання тесту; n – кількість завдань в тесті.

3. Обчислюються значення λ^t за формулою:

$$\lambda^t = \sum_{i=1}^n \sum_{j=1}^m x_{ij} x_i^t, \quad t = 1, 2, \dots, \quad (2)$$

4. Збільшується $t: t = t + 1$. Обчислюються значення k_j^t за формулою:

$$k_j^t = \frac{1}{\lambda^t} \sum_{i=1}^n x_{ij} x_i^t, \quad \sum_{j=1}^m k_j^t = 1, \quad j = \overline{1, m}, \quad (3)$$

5. Повторюються пункти 2-4, поки процес не зійдеться з деякою заданою точністю ε ($\varepsilon = 10^{-6}$). Отримують вектор групової оцінки тестових завдань: $X^t = (x_1, x_2, \dots, x_n)'$ та вектор вагових

коефіцієнтів компетентності експертів: $K^t = (k_1, k_2, \dots, k_m)'$. Враховуючи, що обчислювальний процес рекурентний, досліджується його збіжність. Виключаючи з рівнянь (1–3) змінні k_j^{t-1} та x_i^t , отримуємо:

$$x' = \frac{1}{\lambda^{t-1}} Bx^{t-1}, \quad k' = \frac{1}{\lambda^{t-1}} Ck^{t-1}, \quad t = 1, 2, \dots, \quad (4)$$

де B та C – матриці розмірності $(n \times n)$ та $(m \times m)$ відповідно:

$$B = XX', \quad C = X'X, \quad X = \|x_{i,j}\|.$$

З теореми Перрона-Фробеніуса [11] виходить: якщо матриці B та C додатньо визначені, то при $t \rightarrow \infty$, вектори X, K сходяться до власних векторів матриць B та C , які відповідають максимальним власним числам цих матриць. Граничні значення векторів X, K обчислюються при рішенні наступних рівнянь:

$$BX = \lambda_B X, \quad \sum_{i=1}^n x_i = 1, \quad |B - \lambda_B X| = 0,$$

$$CK = \lambda_C K, \quad \sum_{j=1}^m k_j = 1, \quad |C - \lambda_C K| = 0,$$

де λ_B, λ_C – максимальні власні числа матриць B та C , які приймаються як групові оцінки компетентності експертів.

6. Виконується оцінка узгодженості рішень експертів. Початкова матриця спостережень $\|x_{ij}\|$, $i = \overline{1, n}$, $j = \overline{1, m}$ переформовується у матрицю рангів $\|R_{ij}\|$, на основі якої обчислюються характеристики:

$$r_i = \sum_{j=1}^m R_{ij}, \quad i = \overline{1, n}; \quad D = \frac{1}{n-1} \sum_{i=1}^n (r_i - \bar{r})^2, \quad \bar{r} = \frac{1}{n} \sum_{i=1}^n r_i; \quad W = \frac{D}{D_{\max}}, \quad 0 < W \leq 1,$$

де W – дисперсійний коефіцієнт конкордації (множинний коефіцієнт кореляції рангів Кендалла).

Якщо має місце відсутність пов'язаних рангів, то D_{\max} розраховується наступним чином:

$$\sum_{i=1}^n R_{ij} = \frac{1}{2} n(n+1), \quad j = \overline{1, m}, \quad \bar{r} = \frac{1}{2} (n+1)m,$$

$$S = \sum_{i=1}^n \left(\sum_{j=1}^m r_{ij} - \bar{r} \right)^2, \quad D = \frac{1}{n-1} \sum_{i=1}^n (r_i - \bar{r})^2 = \frac{1}{n-1} S, \quad D_{\max} = \frac{m^2(n^3 - n)}{12(n-1)},$$

На основі даних виразів обчислюються коефіцієнт конкордації:

$$W = \frac{12S}{m^2(n^3 - n)}.$$

Якщо мають місце зв'язані ранги, то:

$$W = \frac{12S}{m^2(n^3 - n) - m \sum_{j=1}^m T_j},$$

де $T_j = \sum_{k=1}^{H_j} (h_k^3 - h_k)$ – показник зв’язаних рангів у j -му ранжуванні; H_j – кількість груп рівнозначних рангів у j -му ранжуванні; h_k – кількість рівнозначних рангів у k -й групі зв’язаних рангів при ранжуванні за j .

Коефіцієнт конкордації дорівнює 1 у тих випадках, коли судження експертів за всіма об’єктами повністю співпадають, та дорівнює 0, коли всі ранжировки різні. В інших випадках його значення відповідає нерівності $0 \leq W \leq 1$. Враховуючи те, що коефіцієнт координації являє собою випадкову величину, виникає необхідність у перевірці його значущості.

Значущість W , а отже, перевірка гіпотези $H_0: W = 0$ при $n > 7$ відбувається на підставі статистичної характеристики для незв’язаних рангів:

$$\chi^2 = Wm(n-1) \quad (5)$$

або для зв’язаних рангів:

$$\chi^2 = \frac{12S}{mn(n+1) - \frac{1}{n-1} \sum_{j=1}^m T_j}, \quad (6)$$

які мають χ^2 – розподіл з $\nu = n - 1$ кількістю степенів вільності.

Знайдене значення χ^2 має бути більше табличного χ^2 , що визначається числом степенів вільності та рівнем довірчої ймовірності ($\chi^2 \geq \chi_{\text{табл}}^2$). Це підтверджує значущість коефіцієнта конкордації, дозволяє відкинути гіпотезу $H_0: W = 0$ та признати, що оцінки експертів узгоджені.

В статті розглянуто задачу визначення стандарту оцінювання для критеріально-орієнтованих тестів, для вирішення якої запропоновано обчислювальну технологію аналізу експертних оцінок складності тестових завдань з використанням коефіцієнту конкордації. Досягнуто підвищення ефективності процесу підтримки прийняття рішень під оцінки змісту тестових завдань за рахунок залучення групи експертів та перевірки узгодженості їх рішень за допомогою методу множинного порівняльного аналізу експертних оцінок.

Список використаних джерел

1. Васильев В. И. Основы культуры адаптивного тестирования / В. И. Васильев, Т. Н. Тягунова. — М.: Издательство ИКАР, 2003. — 584 с.
2. Galeev I. A Learning Model in MONAP / I. Galeev, V. Ivanov, M. Akhmadullin // Human-Computer Interaction. The 6th International Conference. EWHCI'96. Moscow, 1996. — P. 320—323.
3. Автоматизация контроля обученности в процессе подготовки специалистов для систем безопасности / А. Н. Членов, И. Г. Дровникова, Т. А. Буцынская, П. А. Орлов // Научный информационный сборник «Проблемы безопасности и чрезвычайных ситуаций». — М.: Винити, 2009. — № 4. — С. 107—116.
4. Переверзев В. Ю. Критеріально-орієнтовані педагогічні тести для ітогової аттестації студентів / В. Ю. Переверзев. — М.: НМЦ СПО Міністерства освіти РФ, 1999. — 152 с.
5. Зінковський Ю. Ф., Мірських Г. О. Особливості педагогічних тестів // «Вісник НТУУ «КПІ». Серія — Радіотехніка. Радіоапаратобудування. — Вип. 42. — 2010. — С. 157—165.
6. Люсин Д. В. Основы разработки и применения критеріально-орієнтованих педагогічних тестів / Д. В. Люсин. — М.: Исследовательский центр, 1993. — 51 с.
7. Angoff W. H. Scales, norms, and equivalent scores / W. H. Angoff. — Princeton, NJ:ETS, 1984. — p. 153. — URL: <http://www.ets.org/Media/Research/pdf/Angoff.Scales.Norms.Equiv.Scores.pdf>.
8. Kaftandjjeva F. Methods for Setting Cut Scores in Criterion-referenced Achievement Tests / F. Kaftandjjeva. — Cito, Arnhem: EALTA. — 2010. — p. 170. — URL: http://www.ealta.eu.org/documents/resources/FK_second_doctorate.pdf.
9. Литвак Б. Г. Экспертная информация. Методы получения и анализа / Б. Г. Литвак. — М.: Радио и связь, 1982. — 118 с.
10. Принятие решений в системах мониторинга / Т. Г. Емельяненко, А. В. Зборовский, А. Ф. Приставка, Б. Е. Собко. — Днепропетровск: РИК НГУ, 2005. — 224 с.
11. Евланов Л. Г. Экспертные оценки в управлении / Л. Г. Евланов, В. А. Кутузов. — М.: Экономика, 1978. — 134 с.

Kateryna KUZMA
Mykolaiv

COMPUTER TECHNOLOGIES OF DEFINITION OF STANDART CRITERIA FOR CRITERION-REFERENCED TESTS

In the article we propound the computer technology for criterion score definition of criterion-referenced tests on the basis of the comparative analysis of multiple expert assessments of test complexity with the use of the rank correlation coefficient. The score definition is based on expert assessment of test quality. If a group of experts define test quality we propound to use the comparative analysis of test complexity with the use of the rank correlation coefficient.

Key words: assessment of educational achievement, criterion-referenced test, the assessment standard, expert assessment of test quality, multiple comparative analysis, rank correlation coefficient.

Екатерина КУЗЬМА
г. Николаев

ВЫЧИСЛИТЕЛЬНАЯ ТЕХНОЛОГИЯ ОПРЕДЕЛЕНИЯ СТАНДАРТА ОЦЕНКИ ДЛЯ КРИТЕРИАЛЬНО-ОРИЕНТИРОВАННЫХ ТЕСТОВ

В работе предложено вычислительную технологию нахождения критериального балла для критериально-ориентированных тестов на основе метода множественного сравнительного анализа экспертных оценок сложности тестовых заданий с использованием коэффициента конкордации. Метод выбора тестового стандарта основан на экспертных оценках сложности задачи. В случае привлечения группы экспертов для оценки содержания тестовых заданий предлагается проверять согласованность их решений с помощью метода множественного сравнительного анализа с использованием рангового коэффициента корреляции Кендалла.

Ключевые слова: оценка знаний, критериально-ориентированный тест, стандарт оценки, экспертная оценка качества тестовых заданий, множественный сравнительный анализ, коэффициент конкордации.

Стаття надійшла до редколегії 05.10.2016