

ОГЛЯД КРИТЕРІЇВ НАДІЙНОСТІ ОСВІТНІХ ТЕСТІВ

Тернопільський національний педагогічний університет імені Володимира Гнатюка

Анотація

У статті проводиться огляд критеріїв надійності освітніх тестів. Зокрема, розглядається поняття надійності тестів та критерії надійності. Найчастіше в якості критерію надійності тестового завдання використовують коефіцієнт надійності Гутмана і коефіцієнт кореляції Спірман-Брауна, коефіцієнт надійності KR-20, а також їх деякі модифікації.

Ключові слова: надійність, якість, тест, коефіцієнт надійності Гутмана, коефіцієнт кореляції Спірман-Брауна, коефіцієнт надійності KR-20.

OVERVIEW OF CRITERIA OF RELIABILITY OF EDUCATIONAL TESTS

Abstract

The article deals with the quality criteria of educational tests. In particular, the concept of test reliability and reliability criteria are considered. In most cases, the Guttman's reliability coefficient and the Spearman-Brown correlation coefficient, reliability coefficient KR-20 and some of their modifications are used as a criterion for the reliability of the test. The criteria of validity of test tasks are also considered.

Keywords: reliability, quality, test, the Guttman's reliability coefficient, the Spearman-Brown correlation coefficient, KR-20 reliability coefficient.

Оцінювання знань в традиційному (вузькому) розумінні розглядається як визначення підсумкового рівня підготовки того, хто навчається у рамках деякої предметної галузі після проходження курсу навчання. Ключовим моментом цього твердження є нерозривність процесу контролю і процесу навчання. Це твердження правильне для традиційних форм навчання, альтернативних заочних і для нових форм, що розвиваються, таких як дистанційна освіта із застосуванням інформаційних технологій.

Існують завдання оцінювання знань і поза учбовим процесом, які використовуються у виробничому середовищі. Основними з таких завдань є: атестація персоналу, проведення сертифікаційних іспитів, добір кандидатів на конкретні вакансії тощо. Результатом оцінки знань, як правило, являється визначення деякої величини у рамках шкали оцінювання. Наслідки оцінювання можуть мати найрізноманітніші результати - від чисто морального ефекту, до визначення критичних висновків і ухвалення доленосного рішення.

У зв'язку з надзвичайною важливістю точності оцінювання знань необхідно підходити до цього явища як до процесу об'єктивного вимірювання і результати такого вимірювання опрацьовувати за стандартними математичними методами і супроводжувати стандартними характеристиками точності.

Увесь процес оцінювання знань (тестування) повинен виконуватися у рамках науково-обґрунтованої методології, з практичною апробацією, що підтверджує правильність теорії [5]. До основних критеріїв якості освітніх тестів відносять надійність, валідність, складність, дискримінативність, роздільна здатність.

Критерій надійності тестів. Надійністю тесту називається ступінь збігу його результатів при повторному тестуванні одних і тих же учнів в однакових або близьких умовах. Відомо значне число критеріїв надійності тесту [3]. Наприклад, в якості такого критерію можна використовувати коефіцієнт кореляції Пірсона між двома паралельними тестами на одній і тій же вибірці учнів. Однак повторна перевірка знань з одного й того ж питання пов'язана із зайвим психологічним навантаженням учнів і їх перевтомою. Крім того, створити істинно паралельні тести практично нереально.

Як критерій надійності можна використовувати також коефіцієнт кореляції результатів тестування і результатів експертних оцінок. До зайвого психологічного навантаження учнів і їх перевтоми в цьому випадку додається необхідність організації групи експертів і, тим самим, збільшення

навантаження на вчителів.

Найчастіше в якості критерію надійності тестового завдання використовують коефіцієнт надійності Гутмана і коефіцієнт кореляції Спірман-Брауна, а також їх деякі модифікації.

Далі нам потрібні будуть такі позначення:

$P = (p_1, p_2, \dots, p_N) = (p_i, i \in [1: N])$ - множина учнів, де i - номер учня, N - їх загальна кількість;

$T = (t_1, t_2, \dots, t_M) = (t_j, j \in [1: M])$ - набір тестових завдань тесту T , де j - номер тестового завдання в тесті, M - загальне число завдань;

$X = (x_{ij}, i \in [1: N], j \in [1: M])$ - матриця результатів тестування, де x_{ij} - оцінка i -го учня за виконання j -го завдання.

Критерії надійності тесту зазвичай будуються на основі наступних величин:

$y_i = \sum_{j=1}^M x_{ij}$ - сумарний тестовий бал учня p_i за результатами виконання тесту T ;

$\tilde{r}_i = \frac{r_i}{M}, \tilde{w}_i = \frac{w_i}{M}$ - частки правильних і неправильних відповідей, де r_i, w_i - сумарні числа правильних і неправильних відповідей, отриманих учнем p_i за результатами виконання тесту T відповідно;

$\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$ - середній арифметичний бал за всіма учнями (тобто оцінка математичного очікування випадкової величини y_i);

$D = \frac{s^2}{N-1}$ - оцінки дисперсії тестових результатів усіх учнів, де $s^2 = \sum_{i=1}^N (y_i - \bar{y})^2$ - сума квадратів відхилень балів учнів від своїх середніх значень;

$\sigma = \sqrt{D}$ - відповідна оцінка середніх квадратичних відхилень.

Коефіцієнт надійності Гутмана. Одними з найбільш простих методів оцінки надійності тестів є методи, засновані на оцінці структурованості знань учня. Найбільш відомим методом цього класу є метод Гутмана (L. L. Guttman).

Припустимо, що завдання в тесті T розташовані в порядку зростання їх складності та матриця X є бінарною, тобто оцінки x_{ij} можуть приймати тільки значення нуль і одиниця. Рядок з номером i цієї матриці утворює, так званий, профіль учня p_i , що характеризує структуру його знань. При ідеальній структурі тесту (коли складність завдань дійсно зростає зі збільшенням їх номеру) та ідеальній структурі знань учня p_i «правильним» буде профіль, в якому спочатку йдуть тільки одиниці, а потім - тільки нулі. Відхилення оцінки x_{ij} від правильного профілю назвемо помилкою профілю і позначимо e_{ij} . Точніше кажучи, припустимо, що $e_{ij}=0$, якщо оцінка x_{ij} є «правильною», і $e_{ij}=1$ - в іншому випадку.

У разі, коли структура тесту є ідеальною, будь-яка помилка профілю означає або незнання учня, або його невдалу спробу вгадати правильну відповідь. При усередненні результатів тестування по всім учням останній ефект неминуче нівелюється та істотними виявляються лише інверсії (порушення вірної послідовності) в завданнях тесту. На цій підставі Гутманом була в якості критерію надійності тесту запропонована величина

$$\gamma = \gamma(P, T) = \frac{\sum_{i,j} e_{ij}}{NM} \quad (1)$$

звана коефіцієнтом надійності Гутмана. Тут підсумовування ведеться за всіма $i \in [1: N], j \in [1: M]$ і сума являє собою загальне число помилкових відповідей всіх учнів.

Іноді величину (1) називають коефіцієнтом структурованості тестових результатів.

У якості нижньої допустимої межі коефіцієнта надійності Гутмана зазвичай приймають величину 0,8.

На основі профілів учнів запропоновані й інші критерії якості тесту, наприклад, коефіцієнт правильності профілю [3].

Коефіцієнт кореляції Спірман-Брауна. Метод оцінки надійності тестів за допомогою коефіцієнта кореляції Спірман-Брауна (Spearman-Brown) заснований на ідеї оцінки стабільності результатів учнів. Метод відноситься до класу методів роздільної кореляції та є найбільш часто використовуваним методом цього класу [1].

Введемо ще такі позначення:

y_i^e, y_i^0 - сумарні тестові бали учня p_i за результатами виконання непарних і парних завдань тесту T відповідно, $y_i^e + y_i^0 = y_i$

$\bar{y}_e = \frac{\sum_{i=1}^N y_i^e}{N}$, $\bar{y}_o = \frac{\sum_{i=1}^N y_i^o}{N}$ - середні арифметичні бали по всім учням (тобто оцінки математичних очікувань випадкових величин y_i^e, y_i^o відповідно);

$D_e = \frac{s_e^2}{N-1}$, $D_o = \frac{s_o^2}{N-1}$ - оцінки дисперсій тестових результатів y_i^e, y_i^o відповідно, де $s_e^2 = \sum_{i=1}^N (s_i^e - \bar{y}_e)^2$, $s_o^2 = \sum_{i=1}^N (s_i^o - \bar{y}_o)^2$ - суми квадратів відхилень балів учня p_i від своїх середніх значень;

$\sigma_e = \sqrt{D_e}$, $\sigma_o = \sqrt{D_o}$ - оцінки середніх квадратичних відхилень величин y_i^e, y_i^o ;

$K_{eo} = \frac{S_{eo}}{N}$ - оцінка кореляційного моменту (моменту зв'язку) величин y_i^e, y_i^o , де $S_{eo} = \sum_{i=1}^N (y_i^e - \bar{y}_e)(y_i^o - \bar{y}_o)$ - сума добутків відхилень величин y_i^e, y_i^o від своїх середніх значень;

$r_{eo} = \frac{K_{eo}}{\sigma_e \sigma_o} \approx \frac{S_{eo}}{\sqrt{s_e^2 s_o^2}}$ - оцінка коефіцієнта кореляції величин y_i^e, y_i^o ;

$\bar{\varepsilon} = \frac{\sum_{i=1}^N \varepsilon_i}{N}$ - середнє значення помилки ε_i (тобто оцінки математичного очікування випадкової величини ε_i), де $\varepsilon_i = y_i^e - y_i^o$ - помилка учня p_i за результатами виконання непарних і парних завдань тесту T ;

$D_\varepsilon = \frac{\sum_{i=1}^N (\varepsilon_i - \bar{\varepsilon})^2}{N-1}$ - оцінки дисперсії помилок ε_i

Коефіцієнтом кореляції Спірман- Брауна називається величина

$$\mu_{eo} = \frac{2r_{eo}}{1+r_{eo}} \quad (2)$$

Вважається, що тест досить надійний при $\mu_{eo} > 0,8$.

Варіантом формули (2) є формула

$$\widetilde{\mu}_{eo} = 1 - \frac{D_\varepsilon}{D}$$

де, нагадаємо, D_ε, D - оцінки дисперсії помилки тестування і дисперсії тестових результатів усіх учнів відповідно.

Деякі інші варіанти формули (2) розглянуті, наприклад, в роботі [3]. У цій же роботі розглянуті методи оцінки надійності тестів на основі використання результатів дисперсійного та факторного аналізу. Для гомогенних тестів, наприклад, за допомогою дисперсійного аналізу можна отримати індекс надійності тесту та індекс гомогенності тесту, як критерії його надійності.

Розглянутий коефіцієнт кореляції Спірман-Брауна отриманий за допомогою розщеплення тесту (split-half method). Іншими методами, які є набагато менш зручними та рідше використовуються є метод двох паралельних тестів (parallel-form reliability) та метод повторного тестування за допомогою одного і того ж тесту (test-retest reliability) [2].

Коефіцієнт надійності KR-20. Широке поширення на практиці набуло застосування для розрахунку надійності тесту так званої формули KR-20

$$K = \frac{M}{M-1} \left(1 - \frac{\sum \tilde{r}_i \tilde{w}_j}{D} \right)$$

де підсумовування проводиться за всіма $i \in [1: N]$, $j \in [1: M]$. Формула отримала свою назву на ім'я її творців Ф. Kuder і М. Richardson [5].

Вважається, що якщо величина коефіцієнта надійності K становить від 0,90 до 0,99, то тест має відмінну оцінку надійності, якщо від 0,80 до 0,89 - то хорошу, від 0,70 до 0,79 - задовільну, менш 0,69 - незадовільну надійність. Для поточного контролю знань рекомендується використовувати тест, який має коефіцієнта надійності не менше 0,80, а для підсумкової атестації - більш 0,90 [3]. Відзначимо, що за формулою KR-20 оцінюється надійність таких широко відомих тестів як SAT і TOEFL.

Висновки. У роботі проведено огляд основних критеріїв надійності освітніх тестів. Основний зміст роботи становить огляд критеріїв оцінки такого аспекту якості тесту, як надійність. Робота носить переважно оглядовий характер.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аванесов В. С. Тесты в социологическом исследовании. Москва : Изд-во «Наука», 1982. 199 с.
2. Сергієнко В. П., Малезик М. П., Сіткар Т. В. Комп'ютерні технології в тестуванні: навч. посіб. Луцьк : СПД Гадак Жанна Володимирівна, друкарня «Волиньполіграф»^{ТМ}, 2012. 290 с.

3. Олейник Н. М. Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения. Учебное пособие : Донецк, Донецкий Государственный Университет. (<http://opentest.com.ua/test-kak-instrument-izmereniya-urovnya-znaniy/>).

4. Kuder G. F., Richardson M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, v. 2, N 3. p.151–160.

5. Малежик М. П., Сіткар Т. В. Аналіз моделей та методів діагностики знань / Підготовка фахівців з освітніх вимірювань в Україні: [навчально-методичний комплекс]. Ніжин : Видавець ПП Лисенко М. М., 2012. Ч. 2. 398 с.

Сіткар Степан Вікторович – кандидат педагогічних наук, викладач кафедри машинознавства і транспорту, Тернопільський національний педагогічний університет імені Володимира Гнатюка, email: sitkars@gmail.com.

Сіткар Тарас Вікторович – кандидат педагогічних наук, доцент кафедри комп'ютерних технологій, Тернопільський національний педагогічний університет імені Володимира Гнатюка, email: sitkar@gmail.com.

Sitkar Stepan – PhD., Department of Mechanical Engineering and Transport, Ternopil Volodymyr Hnatiuk National Pedagogical University, email: sitkars@gmail.com.

Sitkar Taras – PhD., Department of Computer Technologies, Ternopil Volodymyr Hnatiuk National Pedagogical University, email: sitkar@gmail.com.