

УДК 37.091.26-021.4-047.37

Сілін Є.С., Чапни К.Е.

¹ кандидат фіз.-мат. н., доцент кафедри математики та інформатики, ДВНЗ «ДДПУ»

e-mail: silin-evgen@meta.ua,

ORCID 0000-0003-2470-2704

² здобувач магістерського РВО фізико-математичного факультету, ДВНЗ «ДДПУ»

e-mail: ksushachapny@gmail.com,

ORCID 0009-0001-9836-8473

СТАТИСТИЧНІ ДОСЛІДЖЕННЯ ЯКОСТІ ТЕСТУ

У роботі викладено результати дослідження критеріїв якості – валідності та надійності тестів фахового вступного випробування з математики для вступу абітурієнтів до фізико-математичного факультету Донбаського державного педагогічного університету. Встановлені відповідні числові характеристики (коефіцієнт валідності й коефіцієнт надійності), надана їх якісна оцінка та доведена статистична значущість.

Ключові слова: педагогічне вимірювання, тестування, тестові завдання, валідність тесту, надійність тесту, кореляційний зв'язок, коефіцієнт рангової кореляції Спірмена, коефіцієнт рангової кореляції Кендала.

Вступ

Постановка проблеми. В останнє десятиліття зростає значущість використання різних методів та інструментів оцінювання знань, навичок і вмінь у професійних та наукових сферах, зокрема, в освітній. Один із ключових аспектів такого оцінювання – використання процедури тестування, яка дозволяє об'єктивно, швидко та систематизовано оцінити рівень компетентності здобувачів освіти.

Тестування є актуальним й корисним інструментом у сучасному освітньому, науковому та професійному середовищі. Завдяки тестуванню ми отримуємо можливість об'єктивно оцінити знання, навички та компетенції здобувачів, що дозволяє адаптувати навчальні плани, вдосконалювати методи навчання та відповідати потребам ринку праці. Крім того, тестування створює можливість для здійснення об'єктивних порівнянь, аналізу розвитку та рівня володіння компетенціями в часі. Відчутний попит на тестові інструменти відображає те, як вони сприяють підвищенню якості освіти, ефективності навчання та відбору кращих кандидатів у професійних галузях.

Добре відомо, що досягнення точних та надійних результатів педагогічних вимірювань часто ускладнюється дією різноманітних факторів, які можуть суттєво впливати на процес й результати оцінювання. Отже, природним чином, постає вимога щодо забезпечення розробки та впровадження якісних інструментів тестування, починаючи з постановки цілей педагогічного вимірювання, створення специфікації тесту та тестових завдань, шкалювання, дотримання процедури проведення тестування, аналізу й статистичної обробки результатів.

Одним із суттєвих чинників забезпечення високої якості тестових інструментів є оцінка й аналіз валідності та надійності тесту в цілому,

показників диференційної здатності окремих тестових завдань та інших критеріїв якості. У цьому контексті зазвичай використовують методи математичної статистики, які дозволяють об'єктивно та достовірно оцінити результати педагогічного вимірювання. Проведення аналізу якості тестування підвищує точність й довіру до отриманих результатів педагогічного вимірювання.

Метою дослідження є проведення аналізу та оцінки якості тестів вступного фахового випробування для абітурієнтів за бакалаврським рівнем вищої освіти зі скороченим терміном навчання, спеціальність 014 Середня освіта (Математика), зокрема його валідності та надійності:

1. Встановлення коефіцієнтів рангової кореляції (за Спірменом та Кендалом);
2. Перевірка статистичної значущості коефіцієнтів рангової кореляції;
3. Надання якісної інтерпретації числових значень коефіцієнтів кореляції.

1. Основні поняття та попередні відомості

Валідність методу – це комплексна оцінка, яка заснована на характеристиках інструментів і процедур вимірювання, а також на властивостях властивостей явища, що досліджується. Таким чином, валідність методу відображає відповідність того, що вимірюється за допомогою даного методу, тому, що він повинен вимірювати. Ця характеристика встановлює межі діапазону, для якого метод забезпечує статистично ймовірні результати. [1].

Відповідно до П. Клайна [2] валідність тесту можна поділити на наступні типи:

очевидна валідність – тест вважається валідним, якщо у опитуваних виникає враження, що він дійсно вимірює те, що має виміряти;

конкурентна валідність – оцінюється через кореляцію результатів даного тесту з результатами інших тестів, які вимірюють ту саму характеристику;

прогностична валідність – вивчаються кореляції між показниками тесту та певним критерієм, що характеризує вимірювану властивість, але це відбувається через певний проміжок часу;

змістовна валідність – тест вважається змістовно валідним, якщо можна довести, що завдання тесту відображають всі аспекти області, що досліджується, і при цьому інструкції до завдань чітко сформульовані;

конструктивна валідність – цей підхід включає всі аспекти визначення валідності, перераховані вище.

У контексті проведення вступних випробувань до університету, особливо корисною є критеріальна валідність, яка дозволяє співставити результати тестування абітурієнтів із успішністю вже студентів під час навчання у ЗВО.

Забезпечення надійності педагогічних вимірювань є критично важливим для отримання точних та надійних результатів, на основі яких можна приймати об'єктивні рішення в освіті. Це сприяє вдосконаленню навчальних програм, визначенню успішності здобувачів.

Надійність методу вимірювання визначається ступенем стійкості результатів [3]. Перевірка надійності методу стосується насамперед співставлення результатів при повторних вимірах. Надійність методу залежить від: об'єктивності методу (об'єктивності процедур тестування); параметрів інструменту вимірювання (якості тесту); стабільності характеристики, що вимірюється. Надійність тесту також залежить від кількості тестових завдань.

За визначенням П. Клайна, тест вважається надійним, якщо він є внутрішньо узгодженим, тобто, різні частини тесту вимірюють одну і ту ж характеристику. Також тест вважається надійним, якщо результати залишаються стійкими для одного і того ж випробуваного під час повторного тестування (за умови, що стан випробуваного залишився незмінним). [4].

Найбільш поширеними методами перевірки надійності тесту є наступні.

Метод поділу на половини – тест розділяється на дві рівні половини, результати відповідей на обидві половини порівнюються між собою за допомогою коефіцієнта кореляції.

Метод повторного тестування (метод ретесту) – порівнюються результати тесту, який проводиться двічі через деякий проміжок часу, зі збереженням ідентичних умов.

Метод альтернативних форм – створюються дві альтернативні форми тесту, які мають ідентичні рівні складності та вимірюють ті ж самі концепції або навички. Учасники випробування виконують обидві форми тесту в різний час.

2. Основна частина

У 2018 — 2021 роках правила прийому на навчання в ДДПУ за ОКР «Бакалавр» передбачали вступ до університету на основі раніше здобутого освітньо-кваліфікаційного рівня молодшого спеціаліста, освітньо-професійного ступеня фахового молодшого бакалавра, освітнього ступеня молодшого бакалавра шляхом складання фахового вступного випробування.

В якості інструменту такого випробування для абітурієнтів фізико-математичного факультету було обрано тестування. Ці тести відносяться до гетерогенних, оскільки охоплювали основні розділи алгебри, аналітичної геометрії, математичного аналізу, теорії ймовірностей та математичної статистики. Кожен варіант тесту містить 25 завдань формату А (закритого типу, з вибором однієї правильної відповіді), серед яких 8 теоретичних та 17 практичних завдань; перші двадцять завдань оцінюються по 2 бали, а п'ять останніх, підвищеної складності – по 4 бали. Максимальний бал, який

можна отримати за тест – 60 балів. За 4 роки використання тесту, участь у тестуванні прийняв 31 абітурієнт.

Первинні дані дослідження – результати фахового вступного випробування абітурієнтів до ДДПУ та семестрові оцінки за 100-бальною шкалою з окремих навчальних дисциплін отримано шляхом опрацювання особових справ студентів фізико-математичного факультету, які навчалися за скороченим терміном (джерело – архів та відділ кадрів), документації деканату фізико-математичного факультету.

2.1. Для знаходження критеріальної валідності скористаємося результатами подальшого навчання вступників на фізико-математичному факультеті за навчальними дисциплінами, які відповідають тематиці вступного тесту: елементарна математика (5 семестр), проективна геометрія (6 семестр), теорія ймовірностей та математична статистика (8 семестр), диференціальна геометрія і топологія (6 семестр), комплексний аналіз (8 семестр). Індивідуальні семестрові бали кожного здобувача за названими курсами підсумовуємо та проводимо ранжування. Також визначаємо ранги учасників фахового випробування за результатами виконання тестових завдань.

Спочатку використаємо формулу рангової кореляції Спірмена, у випадку наявності зв'язаних рангів [5]:

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d_i^2}{2 \times \sqrt{\sum x^2 \times \sum y^2}},$$

де $\sum x^2 = \frac{n^3 - n}{12} - T_x$; $\sum y^2 = \frac{n^3 - n}{12} - T_y$; $T_x = \frac{\sum(a^3 - a)}{12}$; $T_y = \frac{\sum(b^3 - b)}{12}$; d – різниця рангів; n – обсяг вибірки, $n = 31$; a – обсяг кожної групи однакових рангів у першому ранговому ряду X ; b – обсяг кожної групи однакових рангів у другому ранговому ряду Y .

Провівши відповідні розрахунки, ми отримали значення коефіцієнту Спірмена $r_s = 0,73$.

Перевіримо його статистичну значущість. Задамо ймовірність помилки першого роду $\alpha = 0,05$ та сформулюємо нульову й конкуруючу гіпотези: H_0 – тестові бали та сумарні семестрові бали з визначених дисциплін математичного циклу статистично не взаємозв'язані ($r_s = 0$); H_1 – тестові бали та сумарні семестрові бали значущо взаємозв'язані ($r_s \neq 0$). Далі обчислимо критичну точку за формулою:

$$T_{кр} = t_{кр} \sqrt{\frac{1 - p_\alpha}{n - 2}},$$

де n – об'єм вибірки, p_α – вибірковий коефіцієнт рангової кореляції Спірмена, $t_{кр}(\alpha; k)$ – критична точка двосторонньої критичної області

розподілу Стюдента. За допомогою таблиці критичних точок розподілу Стюдента знаходимо, що $t_{кр}(0,05;29) = 2,05$, отже, $T_{кр} = 0,26$. Оскільки $|r_s| > T_{кр}$ ($0,73 > 0,26$), то нульову гіпотезу відкидаємо.

Надамо якісну оцінку сили кореляційного зв'язку. Для цього скористаємося відомою шкалою Чедока [6]. Сила кореляційного зв'язку згідно класифікації Чедока – сильна ($r_s = 0,73$). Розмір стандартизованого ефекту згідно класифікації Коена [7] – великий.

Далі оцінимо валідність за допомогою рангової кореляції Кендала, у випадку наявності зв'язаних рангів [5]:

$$\tau = \frac{P - Q}{\left(\sqrt{\frac{n \times (n-1)}{2}} - T_x \right) \times \left(\sqrt{\frac{n \times (n-1)}{2}} - T_y \right)},$$

де $T_x = \frac{\sum(a_x - 1)}{2}$; $T_y = \frac{\sum(b_y - 1)}{2}$; P – кількість збігів (погоджені пари); Q – кількість інверсій (непогоджені пари); n – обсяг вибірки досліджуваних або кількість ознак; a_x – обсяг кожної групи однакових рангів у першому ранговому ряду X ; b_y – обсяг кожної групи однакових рангів у другому ранговому ряду Y . Після виконання відповідних обчислень, маємо: $\tau = 0,49$.

Перевіримо статистичну значущість отриманого значення коефіцієнта рангової кореляції Кендала. Критичну точку знаходимо за формулою:

$$T_{кр} = z_{кр} \sqrt{\frac{2(2n+5)}{9n(n-1)}},$$

де n – об'єм вибірки; $z_{кр}$ – критична точка двосторонньої критичної області, яку знаходять за таблицею значень функції Лапласа. Як і раніше, рівень значущості встановимо 0,05. Враховуючи, що $z_{кр} = 1,96$ і $n = 31$, отримаємо $T_{кр} = 0,25$. Оскільки $|\tau| > T_{кр}$ ($0,49 > 0,25$), то ранговий кореляційний зв'язок Кендала є статистично значущим.

Сила цього кореляційного зв'язку згідно класифікації за Чедоком – помірна ($\tau = 0,49$). Розмір стандартизованого ефекту згідно Коена – середній.

2.2. Для перевірки тесту на надійність, скористаємося методом розщеплення. Розділимо усі завдання тесту на парні та непарні номери й визначимо відповідні ранги щодо кожного абітурієнту.

Знову проведемо розрахунки коефіцієнту рангової кореляції Спірмена r_s (зв'язані ранги) й отримаємо, що $r_s = 0,93$. Це значення свідчить про наявність тісної прямої кореляції між балами за парні й непарні тестові завдання.

Перевіримо статистичну значущість отриманого результату для $\alpha = 0,05$. Зрозуміло, що $t_{кр}(0,05;29) = 2,05$, відповідно, $T_{кр} = 0,14$. Оскільки $|r_s| > T_{кр}$ ($0,93 > 0,14$), то коефіцієнт рангової кореляції Спірмена є значущим.

За шкалою Чедока сила кореляційного зв'язку є дуже сильна; розмір стандартизованого ефекту згідно класифікації Коена – великий.

Використовуючи формулу рангової кореляції Кендала у випадку наявності зв'язаних рангів, маємо: $\tau = 0,62$. Для рівня значущості 0,05 перевіримо нульову гіпотезу про рівність нулю генерального коефіцієнта рангової кореляції Кендала при конкуруючій гіпотезі $H_1: \tau \neq 0$. Оскільки $T_{кр} = 0,25$, то $|\tau| > T_{кр}$ ($0,62 > 0,25$) й нульову гіпотезу відкидаємо.

Сила кореляційного зв'язку згідно класифікації за шкалою Чедока – помірнa ($\tau = 0,62$). Розмір стандартизованого ефекту згідно Коена – середній.

Висновки

За результатами проведеного аналізу тесту з вищої математики для проведення фахового вступного випробування можемо констатувати, що цей метод педагогічних вимірювань є валідним та надійним й може використовуватися в подальшому. Коефіцієнти валідності й надійності, в основу яких покладено коефіцієнт рангової кореляції, мають сильний (кореляція за Спірменом) та середній (кореляція за Кендалом) рівні. Отримані результати є статистично значущими на рівні 0,05.

Перспективним є продовження дослідження окремих тестових завдань для перевірки їх дискримінативної здатності, складності та інших показників якості.

Література

1. Ландар І. П. Валідність результатів вимірювання. *Інформаційно-комунікаційні технології в освіті*. 2014. № 1.
URL: https://e-journals.npu.edu.ua/index.php/ikt/article/view/35/pdf_24 (20.07.23)
2. Ковальчук А. Тестові технології оцінювання якості. *Відкритий урок: розробки, технології, досвід*. 2009. № 3.
URL: [https://osvita.ua/school/method/5919/\(05.08.23\)](https://osvita.ua/school/method/5919/(05.08.23))
3. Булах І. Є., Мруга М. Р. Створюємо якісний тест: *Навчальний посібник*. К.: Майстер-клас. 2006. С. 15
4. Дяченко О. Ф. Організація тестового контролю знань студентів з курсу «Комп'ютерні мережі». *Науковий часопис НПУ ім. М. П. Драгоманова. Сер. 2 : Комп'ютерно-орієнтовані системи навчання*. 2011. №. 11. С. 2.
5. Авраменко В. І. Теорія ймовірностей і математична статистика : *навчальний посібник* / В. І. Авраменко, І. К. Карімов. 2-ге вид., перероб. і доп. Дніпродзержинськ : ДДТУ. 2013. С. 167

6. Боснюк В. Ф. Математичні методи в психології: курс лекцій. Харків. : НУЦЗУ, 2020. С. 52.
URL: [http://repositsc.nuczu.edu.ua/bitstream/123456789/11329/1/Математичні методи в психології \(Боснюк\).pdf](http://repositsc.nuczu.edu.ua/bitstream/123456789/11329/1/Математичні_методи_в_психології_(Боснюк).pdf)
7. Боснюк В. Ф. Математичні методи в психології: курс лекцій. Харків. : НУЦЗУ, 2020. С. 54.
URL: [http://repositsc.nuczu.edu.ua/bitstream/123456789/11329/1/Математичні методи в психології \(Боснюк\).pdf](http://repositsc.nuczu.edu.ua/bitstream/123456789/11329/1/Математичні_методи_в_психології_(Боснюк).pdf)

E.S. Silin, K.E. Chapny

Donbas State Pedagogical University, Sloviansk, Ukraine;

Statistical studies of test quality

The article presents the results of a study of quality criteria - the validity and reliability of the tests of the professional entrance examination in mathematics for the admission of applicants to the Donbas State Pedagogical University at the bachelor's level of higher education with a shortened period of study, specialty 014 Secondary education (Mathematics). Appropriate numerical characteristics (validity coefficient and reliability coefficient) were established, their qualitative assessment was provided, and statistical significance was proven.

Keywords: *pedagogical measurement, testing, test tasks, test validity, test reliability, correlation, the Spearman rank correlation coefficient, the Kendall rank correlation coefficient.*
