

УДК 004:519.22/.25:61:371.275

DOI

АЛГОРИТМ СТАТИСТИЧНОЇ ОБРОБКИ ТА АНАЛІЗУ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ ДЛЯ ОЦІНКИ ЯКОСТІ ТЕСТУ

П. Р. Сельський

*ДВНЗ “Тернопільський державний медичний університет
імені І. Я. Горбачевського МОЗ України”*

ALGORITHM OF THE STATISTICAL PROCESSING AND ANALYSIS OF THE TESTING RESULTS FOR THE EVALUATION OF THE QUALITY OF THE TEST

P. R. Selskyi

SHEI “Ternopil State Medical University by I. Ya. Horbachevsky of MPH of Ukraine”

У статті розкрито досвід впровадження новітніх методик оцінювання у Тернопільському державному медичному університеті імені І. Я. Горбачевського. Обґрунтовано алгоритм процедури статистичної обробки результатів тестування із застосуванням поетапного та комплексного підходу.

The article deals with an experience in innovative assessment methods, implementing by I. Ya. Horbachevsky Ternopil State Medical University. The algorithm of the statistical analysis procedure of the independent testing results with the use of step-by-step and complex approaches is grounded.

Вступ. Розвиток сучасної медичної освіти безпосередньо пов'язаний із впровадженням новітніх інформаційних технологій, що дає змогу реалізації інноваційних навчальних та оцінювальних методик [1–4]. У центрі уваги багатьох науковців продовжує залишатися дослідження моделі оцінювання із використанням тестових запитань з багатьма варіантами відповідей [5–9]. Проте не до кінця вирішеною залишається розробка механізмів аналізу об'єктивності різних інструментів оцінювання та їх перевірки в роботі. Подолати окремі проблеми застосування тестового контролю знань у медичній освіті можна за умови системного підходу та комплексного дослідження ефективності тестових методик із акцентуванням уваги на питаннях автоматизації процедури підготовки тестових завдань та підвищення валідності і надійності всієї методики контролю знань.

Мета – обґрунтувати процедуру статистичної обробки результатів тестування для оцінки якості тесту на основі аналізу комплексного семестрового тестового іспиту із застосуванням поетапного та комплексного підходу.

Основна частина. Для проведення модульного контролю використовувалися запитання з багатьма

варіантами відповідей у формі бланкових тестів та наступною автоматизованою обробкою результатів. Після завершення тестування проводили аналіз рівня якості всіх тестових завдань, які використовували у тесті. Процедури статистичної обробки результатів тестування і методи оцінки якості тесту, які впроваджено відповідно до класичної теорії тестування [10–12], проводили за наступною схемою.

1 етап. Первинний аналіз гістограм успішності відповідей на кожне запитання. Для цього усіх учасників тестування ранжували відповідно до загального результату іспиту. Гістограму успішності відповідей на запитання будували на основі результатів оцінювання досліджуваних, проранжованих у 5 груп. Форма гістограми вказувала на валідність завдання. Для валідних тестових завдань спостерігається монотонне спадання гістограми від групи найуспішніших студентів до найслабших (рис. 1).

Для невалідних тестових завдань, які підлягали корекції чи заміні, гістограма коливається біля певної горизонтальної лінії.

2 етап. Статистична обробка матриці результатів тестування. Числову оцінку успішності виконання j -го завдання, виконаного i -досліджуваним, позначали через x_{ij} . Результати тестування представляли у вигляді матриці $\{x_{ij}\}$ з n рядками і m стовпцями

© П. Р. Сельський



Рис. 1. Гістограма успішності відповідей на валідне запитання.

($i=1, \dots, n; j=1, \dots, m$). Використано дихотомічну шкалу оцінювання результатів тестування, при якій можливі оцінки склалися з двох елементів $\{0; 1\}$: 0 – завдання не виконане, 1 – завдання виконане правильно. Процес статистичної обробки матриці результатів відбувався по кроках.

1 крок. Обчислювали індивідуальні бали досліджуваних $y_i (i=1, \dots, n)$, які показують результат виконання тесту кожним студентом:

$$y_i = \sum_{j=1}^m x_{ij}.$$

2 крок. Обчислювали середні результати сумарних \bar{y} балів досліджуваних:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}.$$

3 крок. Обчислювали середні результати \bar{x}_j досліджуваних за кожним завданням:

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}.$$

Для дихотомічних даних зазначену величину називають індексом складності (p). Індекс p_j визначає міру складності завдання $j (j=1, 2, \dots, m)$:

$$p_j = \frac{\sum_{i=1}^n x_{ij}}{n}.$$

Чим більша величина p_j , тим більша частина успішно справляється із завданням j [13].

Тестове завдання кваліфікували як завдання середньої складності при $0,36 < p_j \leq 0,84$ [7]. При цьому тестові запитання, які не входили у ці межі, підлягали корекції чи видаленню з бази.

4 крок. Обчислювали індекс $I_j (j=1, 2, \dots, m)$ дискримінації завдання, який показує, наскільки добре конкретне тестове завдання розрізняє екзаменованих з високим і низьким балами. Для цього із загальної сукупності досліджуваних виділяли дві підгрупи – тих, хто отримав найвищі сумарні бали, і тих, хто отримав найнижчі. Тоді індекс дискримінації може бути визначений як різниця між відносними чисельностями випробовуваних, які правильно виконали завдання j в цих двох підгрупах. Для цього упорядковану сукупність сумарних балів ділили на три частини і порівнювали результати виконання кожного завдання j першою та останньою третинами досліджуваних. У цьому випадку для дихотомічних даних зазначений індекс обчислювали за такою формулою:

$$I_j = \frac{\sum_{i=1}^{n/3} x_{ij} - \sum_{i=2n/3+1}^n x_{ij}}{n/3}.$$

Брали за основу той факт, що тестове завдання має достатню розподільну здатність при $I_j \geq 0,2$ [7]. Відповідно, тестові завдання із нижчим індексом дискримінації вилучали з бази.

3 етап. Дослідження відхилення від нормального розподілу. Проводилося дослідження розподілу частот, оскільки для перевірки статистичних гіпотез, які застосовуються в класичній теорії тестів, використовують припущення про нормальний розподіл сумарних балів випробовуваних. Для характеристики відхилення від нормального розподілу обчислювали середньоквадратичне відхилення (Q) і центральний момент (μ) із визначенням коефіцієнтів асиметрії (A_s) та ексцесу (E_x).

$$A_s = \frac{\mu_3}{Q^3}, \text{ де,}$$

$$\mu_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \frac{1}{n} \sum_{j=1}^n x_j)^3,$$

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \frac{1}{n} \sum_{j=1}^n x_j)^2}.$$

$$E_x = \frac{\mu_4}{Q^4} - 3, \text{ де:}$$

$$\mu_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \frac{1}{n} \sum x_j)^4 .$$

При цьому процес вимірювання вважали об'єктивним при $A_s = 0$ та $E_x = 0$ [7].

4 етап. Визначення надійності тесту. У зв'язку із значною кількістю досліджуваних студентів та відсутністю можливості повторної перевірки знань за допомогою ідентичного тесту чи еквівалентної його форми (паралельного тесту) дослідження надійності базувалося на одноразовому використанні єдиної форми тесту. Для цього застосовано метод розщеплення, який полягає у розділенні тестової матриці на дві половини, що складаються із завдань з парними та непарними номерами. Оцінку надійності тесту проводили з використанням коефіцієнта кореляції $r_{1/2}$, за формулою Спірмана–Брауна:

$$\rho = \frac{2r_{1/2}}{1 + r_{1/2}} .$$

Тест вважали достатньо надійним, якщо значення коефіцієнта ρ потрапляло в інтервал 0,6–0,89. Надійність вважали дуже високою при значенні $\rho \geq 0,9$ [13].

5 етап. Визначення валідності тесту. За коефіцієнт валідності приймали коефіцієнт кореляції (v) результатів тестових вимірювань і незалежного зовнішнього критерію – експертної оцінки. Якщо експертна оцінка знань досліджуваних, отримана

незалежно від процедури тестування, представлена числовою послідовністю Y_1, Y_2, \dots, Y_n , то коефіцієнт валідності розраховували за формулою:

$$V = \frac{\sum_{i=1}^n (Y_i \times y_i)}{s_Y \times s_y} - \bar{Y} \times \bar{y} \times \frac{n}{n-1} ,$$

де \bar{Y} – середнє арифметичне експертних оцінок, s_Y – стандартне відхилення цих оцінок:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} ,$$

$$s_Y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2} .$$

Таким чином, для визначення валідності необхідна наявність незалежного зовнішнього критерію, тобто оцінка експерта з кожного запитання, що неможливо у зв'язку із значним обсягом тестових завдань, які використовуються на КСТІ. Тому впроваджено визначення валідності тесту на основі індексу складності тесту (p_{ij}) в якості експертної оцінки. При цьому у зазначеній формулі $Y_i = y_{\max} \times p_{ij}$, де y_{\max} – максимально можлива кількість балів за тест. Тест вважали валідним при коефіцієнті кореляції (v) більшому 0,6 [7, 14].

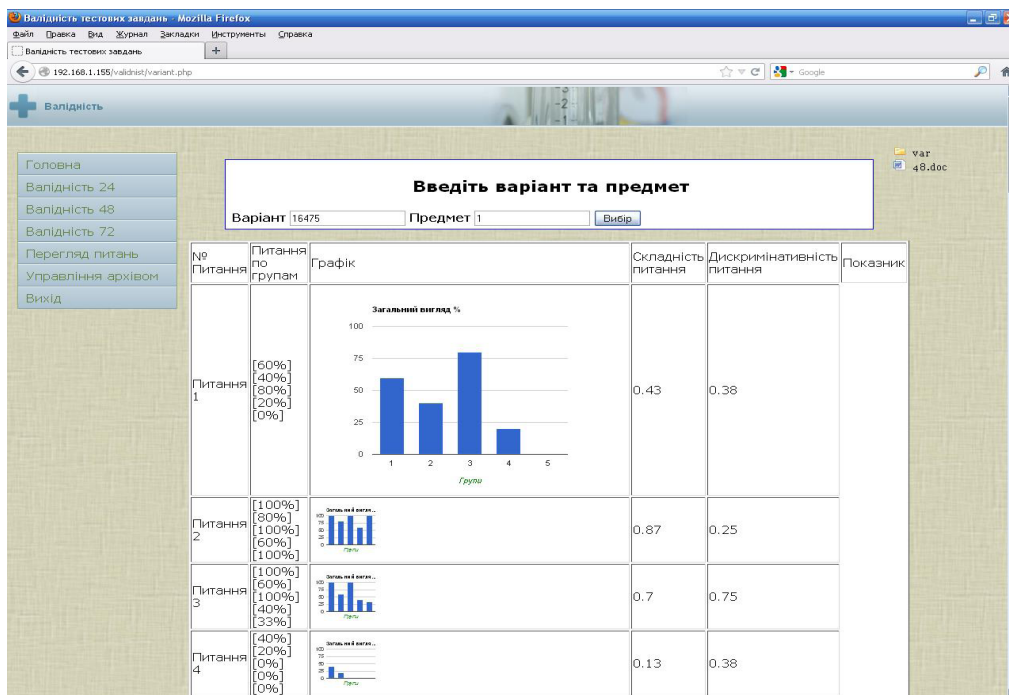


Рис. 2. Результати перевірки на валідність тестових завдань, які використовувались на КСТІ.

6 етап. Визначення якості тесту на основі аналізу результатів дослідження на попередніх етапах. Заключний висновок щодо якості тесту здійснювали на основі комплексного аналізу відхилення від нормального розподілу, надійності, валідності, а також результатів дослідження гістограм успішності, визначення індексів складності та дискримінації кожного тестового завдання із визначенням частки невалідних запитань.

Для забезпечення зворотного зв'язку з авторами тестів інформацію щодо перевірки тестових завдань на валідність розміщували у відповідному розділі внутрішньоуніверситетської мережі Інтранет (рис. 2). Увівши номер варіанта та порядковий номер модуля у буклеті, можна отримати результати перевірки на валідність кожного тестового завдання.

Для програмного забезпечення процедури статистичної обробки результатів тестування нами розроблено комп'ютерну програму "Визначення якості оцінювання знань та практичної компетентності

Список літератури

1. Волосовець О. П. Питання якості освіти в контексті впровадження засад Болонської декларації у вищій медичній школі / О. П. Волосовець // Досвід впровадження основних засад Болонського процесу в систему вищої медичної (фармацевтичної) освіти : матеріали науково-методичної конференції. – Т., 2005. – С. 8–16.

2. Глибовець М. М. Деякі концепції реалізації навчальних взаємодій у системах дистанційної освіти / М. М. Глибовець, А. М. Глибовець, Ю. В. Рощенко // Вісник Київського університету. Серія: Фізико-математичні науки. – 2010. – Вип. № 1. – С. 84–93.

3. Мінцер О. П. Концептуальні узагальнення щодо структурної організації комп'ютерних мереж вищих медичних навчальних закладів / О. П. Мінцер, О. А. Рижов, В. П. Марценюк // Медична інформатика та інженерія. – 2013. – № 4. – С. 7–15.

4. Ковальчук Л. Я. Впровадження новітніх методик і систем навчального процесу в Тернопільському державному медичному університеті імені І. Я. Горбачевського / Л. Я. Ковальчук // Медична освіта. – 2009. – № 2. – С. 10–14.

5. Schuwirth L. W. T. Clinical review. ABC of learning and teaching in medicine. Written assessment / L. W. T. Schuwirth, C. P. M. van der Vleuten // British Medical Journal. – 2003. – № 326. – P. 643–645.

6. Epstein R. M. Assessment in Medical Education / R. M. Epstein // The New England Journal of Medicine. – 2007. – № 1. – P. 356–387.

7. Булах І. Є. Створюємо якісний тест : навч. посіб. / І. Є. Булах, М. Р. Мруга. – К. : Майстер-клас, 2006. – 160 с.

8. Марценюк В. П. Аналіз об'єктивності новітніх методик оцінювання при викладанні біологічної та біоорганічної хімії / В. П. Марценюк, П. Р. Сельський // Медична хімія. – 2012. – № 3. – С. 100–103.

в інформаційній системі управління якістю підготовки фахівців-медиків" [15].

Висновки. Представлено алгоритм процедури статистичної обробки результатів тестування із застосуванням поетапного та комплексного підходу.

Розроблено покроковий аналіз валідності тесту та всіх тестових завдань, включно із аналізом гістограм успішності, визначенням коефіцієнтів асиметрії та ексцесу, індексів складності та дискримінації.

Впроваджено визначення коефіцієнта надійності за умови одноразового використання єдиної форми тесту та запропоновано визначення валідності тесту на основі індексу складності в якості експертної оцінки.

Подальші дослідження з удосконалення процедури статистичної обробки та аналізу результатів тестування суттєво покращать якість освітніх методик при підготовці висококваліфікованих спеціалістів.

9. Сельський П. Р. Інформаційна система оцінювання знань в медичній освіті : монографія / П. Р. Сельський. – Тернопіль : ТДМУ, 2013. – 212 с.

10. Аванесов В. С. Теоретические основы разработки заданий в тестовой форме / В. С. Аванесов. – М. : Изд-во МГТА, 1995. – 168 с.

11. Thorndike R. L. Applied Psychometrics / R. L. Thorndike. – Boston : Houghton. Mifflin Co, 1982. – 390 p.

12. Аванесов В. С. Форма тестовых заданий : учебное пособие для учителей школ, лицеев, преподавателей вузов и колледжей. – 2-е изд., переработанное и расширенное / В. С. Аванесов. – М. : Центр тестирования, 2005. – 156 с.

13. Ащепкова Л. Я. Конструирование тестовых заданий и обработка результатов тестирования [Электронный ресурс] / Л. Я. Ащепкова // Библиотека кафедры прикладной информатики и информационных технологий Дальневосточного государственного университета, 2003. – Режим доступа : <http://kpmiit.wl.dvgu.ru/library/aschepkova/tests.php#statchar>

14. Norman G. The long case versus objective structured clinical examinations / G. Norman // BMJ. – 2002. – № 324. – P. 748–757.

15. Свід. про реєстр. автор. права на твір № 53202 Комп'ютерна програма "Визначення якості оцінювання знань та практичної компетентності в інформаційній системі управління якістю підготовки фахівців-медиків" ("ВЯОЗПКІСУЯПФ") / П. Р. Сельський, В. П. Марценюк, С. Б. Чеканов. – К. : Державний департамент інтелектуальної власності України. – Дата реєстрації 20.01.2014.

Отримано 18.06.15