

УДК 004, 681.518, 378.146

Статистичний та змістовий аналіз результатів комп'ютерного сеансу тестування з фізики

Данилов А. Б., ст. викл. каф. ІМПФ

Національний університет «Львівська політехніка»
(вул. С. Бандери, 12, м. Львів, 79013, Україна)

На сьогодні розроблено низку автоматизованих систем потокового тестування знань. Однією з таких систем є розробка ХНУРЕ Opentest. Вона дає змогу оцінити знання студентів за традиційною, а також квантильною шкалою, провести аналіз засвоєння студентами матеріалу за окремими темами, оцінити профіль знань студентів. Також ця система забезпечує багатий статистичний матеріал, на базі якого викладач може вилучити з бази даних неспроможні питання, питання, які є зайвими в тестовому сеансі, а також питання, що містять некоректності, а тому їх умову треба змінити.

Покажемо результати застосування Opentest на прикладі проведеного сеансу тестування в академічній групі ПФ-21 з дисципліни «Оптика». На першому етапі створювався масив питань тестового типу. Масив складався з 98 питань, розбитих на 7 тем, по 14 питань в кожній. Питання охоплювали матеріал, що читався студентам в другій половині семестру в рамках курсу «Оптика», і стосувалися розділів «Хвильова оптика» і «Квантова оптика». Тестові питання належали до так званих питань 1 типу, коли студенту потрібно обрати 1 або кілька правильних відповідей з запропонованих варіантів, і 2 типу, коли потрібно знайти відповідність між двома колонками фізичних понять, формул, або формулювань, з пропорцією приблизно 10 до 1. Середня альтернативність варіантів відповідей для різних тем коливалася в межах 4–5.

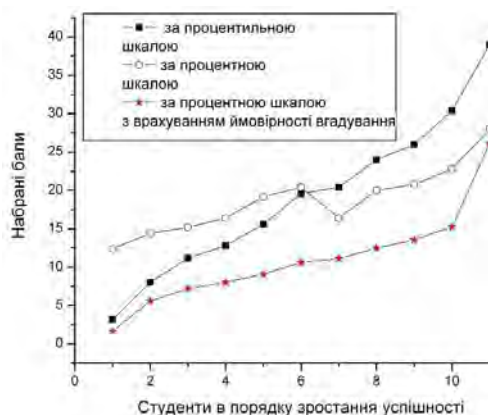


Рис. 1 Порівняльні графіки результатів тестування (максимальний бал 40)

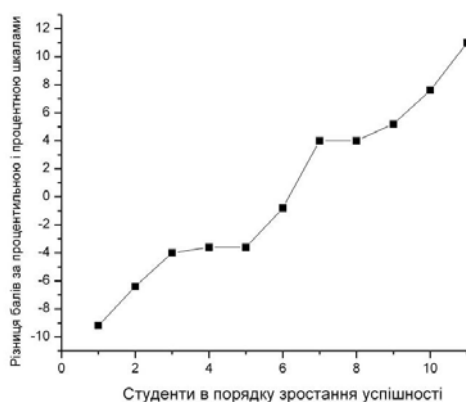


Рис. 2 Різниця в балах за процентильною та процентною шкалами

Сеанс тестування складався з 21 запитання. Таким чином, на кожную тему припадало по три питання. Тривалість сеансу – 80 хвилин. Питання поділялися за складністю на 2 типи: 1 і 2 рівня складності, а правильні відповіді оцінювалися, відповідно 1 і 2 балами. Якщо правильних варіантів було кілька, оцінка питання визначалася як $k \frac{a}{b}$, де k – рівень складності питання, a – кількість даних студентом правильних відповідей, b – загальна кількість правильних відповідей.

Частотний аналіз з кроком 10 % показує, що розподіл результатів студентів тестованої групи близький до нормального з незначним відхиленням у бік нижчого рівня. За даного частотного розподілу результатів оцінка за процентильною шкалою вища за первинну оцінку для студентів з вищою успішністю і нижча для слабших студентів. Ця ж оцінка вища за кориговану оцінку з врахуванням ймовірності вгадування для всіх студентів.

Тематичний аналіз тесту свідчить про значну різницю трудностей питань за різними темами. Тема «Задачі III рівня складності» передбачає високий рівень трудності. Проте теми «Дифракція Фраунгофера» і «Подвійне променезаломлення» за рівнем трудності мали б бути близькими до 4 інших теоретичних тем. З іншого боку високий коефіцієнт Пірсона теми «Дифракція Фраунгофера» свідчить про достатню надійність результатів тестування. Аналіз результатів для теми «Подвійне променезаломлення», зокрема низький коефіцієнт Пірсона, вказав на необхідність детального аналізу окремих питань теми. 8 з 14 питань теми дали внесок у кореляційний коефіцієнт і мають бути виключені з тесту, або належним чином переформульовані. Для більшості з них характерний високий рівень випадковості відповідей, що унеможливорює визначення профілю знань студента.

За результатами тестування і статистичного аналізу будувалася коригована матриця результатів тестування. Вона містить лише 20 питань, які повністю задовольняють вимоги, щодо питань, які мають включатися в сеанс тестування. Вибиралися питання, частота випадань яких, враховуючи однотипні, становила ≥ 3 . Вилучалися питання, частота випадання яких була недостатньою для статистичного аналізу. Також з таблиці виключалися всі прості ($T > 0,8$) і складні ($T < 0,2$) питання та питання з низькою кореляцією (від'ємний коефіцієнт Пірсона). При цьому проводилося об'єднання питань, що мали схожі умови. Це дало змогу збільшити кількість завдань, що увійшли в кориговану матрицю. Найбільшої достовірності результатів вдається досягнути для тих питань, які фігурують у тесті щонайменше 10 разів. Тому достатню точність результатів можна отримати для потоків з кількістю студентів більше 50.

Незважаючи на малу кількість тестованих, результати тестування добре корелюють з середніми оцінками студентів протягом навчання. Як наслідок: проведення тестового сеансу за допомогою програми Orentest дало змогу оцінити рівень засвоєння студентами матеріалу за різними темами та визначити коло питань, які необхідно з тих чи інших причин виключити з масиву тестових питань. Для повної достовірності аналізу масиву тестових питань потрібно щонайменше 200-300 тестованих. Питання переведення оцінки за тестування у семестрову або модульну оцінку залишається відкритим і потребує подальших досліджень. Очевидно безпосереднє переведення оцінки тестування у підсумкову оцінку є недоцільним. Застосування програми в рамках оцінювання залишкових знань можливе лише для великих потоків за умови вироблення спільних підходів до шкалювання.