

Удосконалення методів автоматизованого оцінювання якості тестових завдань в освітніх вимірюваннях

Вінницький національний технічний університет

Анотація

У роботі обґрунтовано актуальність розробки інформаційної технології для оцінювання якості тестових завдань в освітніх вимірюваннях. Проаналізовано основні показники якості тестів, такі як складність, дискримінативність та надійність, а також вивчено можливості їхнього удосконалення за допомогою інтеграції аналізу часу виконання завдань. Встановлено, що аналіз часу виконання завдань не лише дозволяє покращити точність та розширити критерії оцінювання, але й виявляти аномальні поведінкові патерни, що в цілому сприяє підвищенню ефективності тестування.

Ключові слова: оцінка якості тестових завдань, складність тестів, дискримінативність, надійність, аналіз часу, освітні вимірювання.

Abstract

The paper justifies the relevance of developing an information technology for evaluating the quality of test tasks in educational measurements. The main quality indicators of tests, such as difficulty, discrimination, and reliability, are analyzed, and the possibilities for their improvement through the integration of time analysis of task completion are explored. It is established that the analysis of completion time not only improves the accuracy and expands the evaluation criteria but also reveals anomalous behavioral patterns, thereby enhancing the overall effectiveness of testing.

Keywords: test task quality evaluation, test difficulty, discrimination, reliability, time analysis, educational measurements.

Вступ

Автоматизована оцінка якості тестових завдань є актуальним напрямом досліджень у галузі освітніх технологій, що набуває особливого значення в умовах глобальної цифровізації освіти. Вимірювання навчальних досягнень за допомогою комп'ютерних технологій поступово витісняє традиційні методи оцінювання знань і наразі широко впроваджується як у сучасних електронних освітніх платформах, так і в державних закладах освіти. У зв'язку з цим зростає потреба в системному, об'єктивному та ефективному аналізі якості тестових матеріалів з метою забезпечення їхньої валідності, надійності та дидактичної доцільності. Одним із ключових аспектів такого аналізу є автоматизація процесів оцінювання якості тестових завдань, зокрема за критеріями складності, дискримінативності та ефективності дистракторів [1].

Важливою передумовою якісної розробки та інтерпретації результатів тестування є розмежування двох функціонально різних напрямків: контрольних тестів, призначених для об'єктивного вимірювання рівня сформованості знань, та навчальних тестів, що виконують формувальну функцію і спрямовані на активізацію пізнавальної діяльності та удосконалення процесу засвоєння навчального матеріалу.

Ці два напрями тестування – контрольний та навчальний – відрізняються як за метою, так і за функціональним призначенням, а відтак і за характером завдань та вимогами до їхньої побудови.

Навчальні тести спрямовані на підтримку та стимулювання процесу засвоєння матеріалу, забезпечення можливостей для самоперевірки та формування зворотного зв'язку між здобувачем освіти й освітнім середовищем. Їхнє використання передбачає більшу гнучкість у побудові завдань, можливість включення пояснень, підказок, елементів інтерактивності та гейміфікації, що робить такі тести ефективним інструментом який не завжди потребує впровадження формального оцінювання.

Натомість контрольні тести орієнтовані на об'єктивне оцінювання рівня знань, умінь і навичок, а тому вимагають чітко сформульованих завдань із однозначно визначеними правильними відповідями, стандартизованих умов проведення та високого рівня психометричних характеристик. Отже, саме для контрольних тестів, на відміну від навчальних, оцінка якості завдань є вирішальною, оскільки їхні результати використовуються для формування загального рейтингу, порівняння навчальних досягнень, відбору кандидатів або загального моніторингу якості освіти.

Результати дослідження

Відповідно до класичної теорії, основними показниками оцінювання якості тестів є складність завдання, дискримінативність та надійність тесту [2].

Складність завдання (P_k), визначає частку правильних відповідей на конкретне завдання, що дозволяє оцінити, наскільки воно є легким чи складним для студентів:

$$P_k = \frac{n_i}{n}, \quad (1)$$

де n_i – кількість учасників тестування, що надали правильну відповідь на k -те питання; n – загальна кількість учасників тестування.

При цьому оптимальна складність завдання ($a < P_k < b$) може варіюватися та визначатися специфічними особливостями відповідного тесту. Проте зазвичай використовують двосторонню 10% область для якої оптимальні межі становлять $0,1 < P_k < 0,9$. Якщо ж $P_k > 0,9$ – завдання вважається занадто легким, а при $P_k < 0,1$, навпаки, завдання є занадто складним. Виявлені складні та прості завдання з тесту виключаються і результат тестування після цього перераховується.

Отже, цей показник є важливим для того, щоб уникнути як надмірно легких завдань, що не здатні розрізняти рівень знань, так і занадто складних, які можуть призвести до помилкових висновків про рівень підготовленості

Дискримінативність, або *індекс дискримінації*, характеризує здатність завдання відокремлювати студентів з високим рівнем знань від тих, хто має низький рівень. Це є критичним для забезпечення якості тесту, оскільки завдання повинно бути ефективним у розпізнаванні різниці між учасниками з різними рівнями підготовленості.

Методика розрахунку дискримінативності завдань тесту зазвичай здійснюється за допомогою двох основних підходів: індексу дискримінації (2) та точково-бісеріальної кореляції (3).

Для знаходження *індексу дискримінації* (D) відбувається поділ учасників тестування на дві групи: верхню та нижню, залежно від їх результатів на всьому тесті. Індекс дискримінації для кожного завдання обчислюється як різниця між часткою правильних відповідей у верхній групі (P_1) та часткою правильних відповідей у нижній групі (P_0):

$$D = \frac{P_1 - P_0}{P_{max}}, \quad (2)$$

де P_1 – кількість правильних відповідей у верхній групі, P_0 – кількість правильних відповідей у нижній групі, P_{max} – максимальна кількість правильних відповідей. Значення, близьке до 1, вказує на високу дискримінативність.

Точково-бісеріальна кореляція (ρ_k) знаходиться як кореляція між бінарною змінною (правильна чи неправильна відповідь на завдання) та неперервною змінною (результати тесту):

$$\rho_k = \frac{\bar{P}_1 - \bar{P}_0}{s} \sqrt{\frac{n_1 n_0}{n(n-1)}}, \quad (3)$$

Де $\bar{P}_1$ та $\bar{P}_0$ – середні значення результатів тесту для групи з правильними та неправильними відповідями відповідно, s – середнє квадратичне відхилення для всього набору даних, n_1 , n_0 – кількість спостережень у відповідних групах.

Наведені класичні підходи до оцінювання якості тестів, можуть бути удосконалені завдяки фіксації та аналізу часу виконання тесту. Такий аналіз став можливим завдяки впровадженню автоматизованих систем тестування, що дозволяє не лише точно фіксувати час, витрачений на виконання завдання, але й використовувати цей параметр для поглибленого оцінювання якості тесту.

Аналіз часу допоможе додатково виявляти завдання, які можуть бути занадто легкими або важкими, в залежності від того, скільки часу учасники витрачають на їхнє виконання. Зокрема, відхилення у кореляції між складністю завдання та часом його виконання можуть виявляти завдання з нечіткими або помилковими формулюваннями. Крім того, цей аналіз дозволяє виявляти аномальні поведінкові патерни, такі як вгадування чи шахрайство, що можуть бути помітні через нетипово швидкий чи

повільний час виконання завдання. Завдяки аналізу часу також можна вдосконалювати адаптивні стратегії тестування, змінюючи рівень складності завдань відповідно до швидкості та точності виконання попередніх завдань.

Таким чином, інтеграція аналізу часу виконання завдань дозволяє значно покращити точність оцінювання, забезпечуючи більш комплексний підхід до аналізу тестових матеріалів.

Висновки

Проведене дослідження показало, що класичні підходи, засновані на аналізі складності завдань, дискримінативності та надійності тестів, зберігають свою актуальність, але в умовах сучасних технологій їхній потенціал може бути значно розширений завдяки впровадженню автоматизованих систем тестування. Фіксація часу виконання завдань дає змогу побудувати інтегровану систему оцінки якості тестів, яка враховує не лише традиційні показники, такі як правильність відповідей, але й часові параметри, що дозволяє глибше аналізувати складність завдань, рівень когнітивного навантаження та поведінкові патерни учасників. Це відкриває нові можливості для вдосконалення адаптивних стратегій тестування, коригування завдань на основі виявлених аномалій у часі виконання та точнішого визначення рівня знань студентів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Кухаренко В. М., Перхун Л. П., Товмаченко Н. М. Методика комплексного оцінювання якості тестів. Частина 1 / В. М. Кухаренко, Л. П. Перхун, Н. М. Товмаченко // Статистика України. – 2018. – № 3. – С. 40–48.

2. Смідович Л. С., Артюх О. В., Косенко Н. В. Автоматизація аналізу якості тестових завдань та результатів тестування студентів ІТ-спеціальностей / Л. С. Смідович, О. В. Артюх, Н. В. Косенко // Сучасний стан наукових досліджень та технологій в промисловості. – 2020. – № 1 (11). – С. 19–27. DOI: <https://doi.org/10.30837/2522-9818.2020.11.019>.

Шевчук Олександр Федорович – доцент кафедри комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: shevchuk177@gmail.com

Христянчук Владислав Валерійович – студент групи КН24М факультету інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: ladislaokh@gmail.com

Shevchuk Oleksandr F. – Associate Professor of the Department of Computer Sciences, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: shevchuk177@gmail.com

Khrystianchuk Vladyslav V. – student, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: ladislaokh@gmail.com