

ОПТИМІЗАЦІЯ ЗАВДАНЬ В ТЕСТОВІЙ ФОРМІ НА ОСНОВІ ЙМОВІРНІСНОЇ МОДЕЛІ

Визначено педагогічні умови оптимізації завдань у тестовій формі, здійснено аналіз результатів тестування. Була використана однопараметрична ймовірнісна модель Раша, за допомогою якої було визначено надійність системи завдань у цілому, а також недоліки деяких окремих завдань. Проведена робота дає можливість виявити некоректно побудовані завдання, з'ясувати пробіли в засвоєнні матеріалу і, як подальший результат, побудувати об'єктивний тест.

Одним з важливих заходів для забезпечення процесу функціонування та розвитку освіти в Україні є впровадження сучасних систем контролю якості знань студентів та моніторингу якості освіти [1]. Специфікою тесту є те, що він являє собою інструмент оцінки якості та контролю рівня знань студентів і забезпечує технологізацію цього процесу. Інша функція завдань у тестовій формі – навчальна, адже правильно складені завдання спонукають студентів до творчого мислення та до застосування вивченого матеріалу при розв'язанні нових задач. Для того, щоб бути дійсно ефективним і надійним інструментом, як кожне завдання, так і весь тест як система завдань потребують аналізу. Мета такого аналізу – створити якісний і об'єктивний тест. Серед задач такого аналізу – виявити, наскільки коректно завдання виявляють якість знань з вивченої дисципліни, наскільки повно перевіряє тест пройдений розділ.

Тестування – це більше, ніж просто опитування студентів чи учнів за допомогою декількох питань із запропонованим вибором варіанту відповіді. По-перше, існують різні типи завдань у тестовій формі; по-друге, це тільки частина тестового процесу, який включає також інші задачі, що для їх вирішення, окрім усього, потрібна математична і статистична теорія тестів, оскільки тест містить не тільки завдання, але й результати та аналіз цих результатів. Важливість такої роботи полягає в тому, що лише після такої перевірки системи завдань можна говорити про неї як про тест, і в подальшому використовувати її для надійного оцінювання і отримання вимірів при моніторингу якості навчання.

У наведеній роботі пропонуються результати аналізу завдань з фізики. Цей аналіз дозволяє створити якісний тест і зробити об'єктивною оцінку рівня підготовленості студентів. Проведене тестування у студентів радіофізичного факультету дало оцінки рівня студентів, які ввійшли до поточних результатів роботи студентів.

Математична модель

Для кількісного опису властивостей завдання, і тесту в цілому, потрібно використовувати певну модель поведінки очікуваних результатів. До таких властивостей належить трудність завдання, здатність розділяти студентів за рівнем компетенції, ефективність дистракторів. Розглядаючи два параметри – трудність завдання і рівень знань студента, – маємо модель Раша [2].

Модель передбачає конкретну залежність ймовірності відповіді на дане завдання. Чим вищий рівень знань студента, тим більшим є значення ймовірності правильно виконати завдання. Звичайно, трудність завдання теж визначає цю ймовірність. Поставивши у відповідність рівню студента безрозмірний параметр β , а рівню трудності завдання параметр δ , маємо логістичну залежність для ймовірності відповіді. Графічно ця залежність виглядає так, як зображено на рис.1.

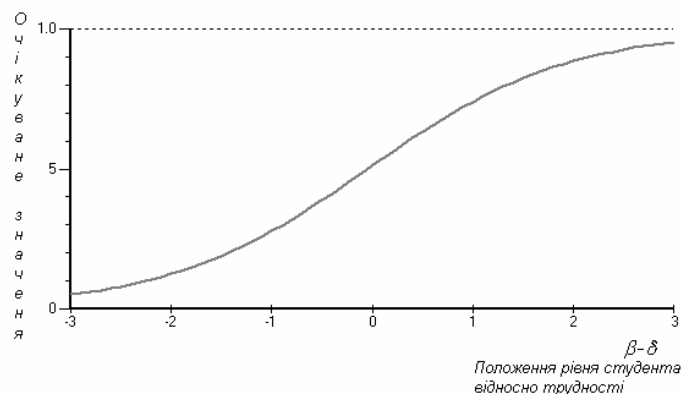


Рис. 1. Характеристична крива для опису взаємодії студента із завданням.

Вже з самої діаграми можна зробити ряд висновків, наприклад, крутизна кривої говорить про те, наскільки сильно дане завдання розділяє студентів на тих, хто знає добре дане питання від тих, у кого компетентності недостатньо. Точка, в якій ця крива перетинає значення ймовірності 0,5, показує величину трудності завдання.

Ще більше інформації можна одержати при співставленні теоретичної кривої з отриманими результатами, а також розглядаючи характеристики всієї системи завдань. Для ілюстрації такого аналізу розглянемо результати тестування з фізики. Завдання, описані нижче, використовувалися для перевірки засвоєння студентами матеріалу з курсу фізики у першому семестрі.

Педагогічний аналіз тесту

Застосувавши алгоритм розрахунку параметрів завдань до результатів тестування, одержуємо числові значення параметрів тесту. Діаграми, на яких відкладено значення рівнів трудностей завдань та рівнів знань студентів, приведені на рис. 2 та рис. 3.

На рис. 2 зображений розподіл результатів для тесту № 1 у вигляді двох суміщених діаграм.

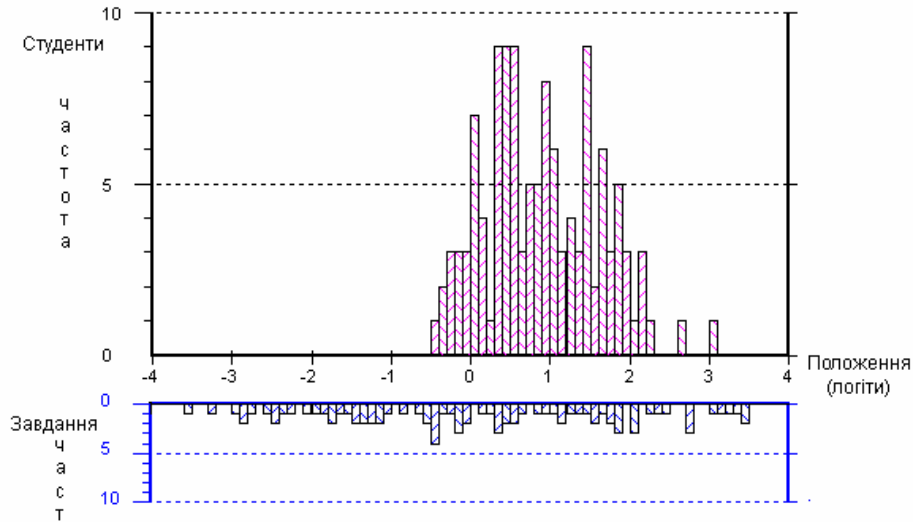


Рис. 2. Розподіл рівнів студентів та складностей завдань. Тест № 1

Верхня показує розподіл студентів за рівнем знань. На горизонтальній осі відкладено рівень у безрозмірних одиницях – логітах. На вертикальній осі відкладена частота, тобто кількість студентів з рівнями, що вкладаються в кожному з малих інтервалів – 0,1 логіта. Нижня частина містить розподіл трудностей тестових завдань. Нульове значення діаграми виставлене за середнім значенням труднощі. Шкали цих двох розподілів суміщені, що дає можливість аналізувати співвідношення між ними.

Рис. 3 містить таку ж інформацію, що й рис. 2. На ньому відкладені результати розрахунку параметрів для тесту 2.

Після порівняння двох діаграм на рис. 2 та рис. 3 розглянемо деякі особливості.

Порівнюючи розподіли рівнів студентів, можна побачити суттєву різницю, яка існує між першим та другим тестами. Про розподіл студентів при першому тестуванні не можна сказати, що він відповідає певному аналітичному закону. Це говорить про те, що кількість студентів як середнього, так і низького та вищого рівня є приблизно однаковою. У другому тесті ширина розподілу студентів лишається тією ж (приблизно 3,6 логіта), але з'являється певний характер розподілу.

Огинаюча для цієї діаграми близька до гаусіану [3]. Максимум розподілу відповідає точці в області, де знаходиться більшість студентів із середнім рівнем знань. На лівій та правій частині діаграми розташовані студенти з найнижчим та найвищим рівнями, і кількість таких студентів менша, ніж у першому тесті. Таку зміну в диференціації студентів протягом семестру можна пояснити. На початку, у першій половині семестру студенти ще використовують потенціал, який вони отримали під час підготовки до вступу в університет, вступний відбір виділив студентів з вищим рівнем. Протягом навчання матеріал дедалі ставав складнішим для опанування, що вимагало більше самостійної роботи та розумових зусиль. Таким чином, з'являється нова картина розподілу рівня знань. Можна сказати, що перший тест показує, якими учнями в школі були ті, хто проходили тестування. Другий тест вже дає певні висновки про те, якими студентами вони будуть у подальшому навчанні в університеті.

Використаний алгоритм обробки знаходить параметри для групи і тесту, тобто рівень знань студентів і значення складності завдань. Ці параметри незалежні один від одного, і саме завдяки тесту їх можна порівнювати. Порівняємо картину для складності завдань у першому та другому тестах. Особливість розподілу складностей завдань у першому тесті полягає в тому, що вона не має певного характеру, як у другому тесті, а розподілена рівномірніше. Розкид частотних стовпчиків інший, тобто тест більш рівномірний за складністю, ніж тест № 2 [4]. Межі, в яких лежать значення складності – від -3,5 до +3,5 логіта, приблизно вдвічі ширші за інтервал, у якому лежать значення студентських рівнів (від -0,5 до +3 логіта). Завдання зі складностями, що лежать за межами цього інтервалу, не є ефективними. Ті, які знаходяться ліворуч, занадто легкі і практично всі студенти можуть їх вирішити правильно, ті завдання, які знаходяться праворуч, занадто складні для розв'язання. "Хвіст" простих завдань дуже довгий у тесті № 1 (від -0,5 до 3,5 логіта), тобто майже такий, як весь діапазон рівнів студентів. До того ж інтервал завдань не суцільний, а має пробіли, що показує брак завдань середньої труднощі.

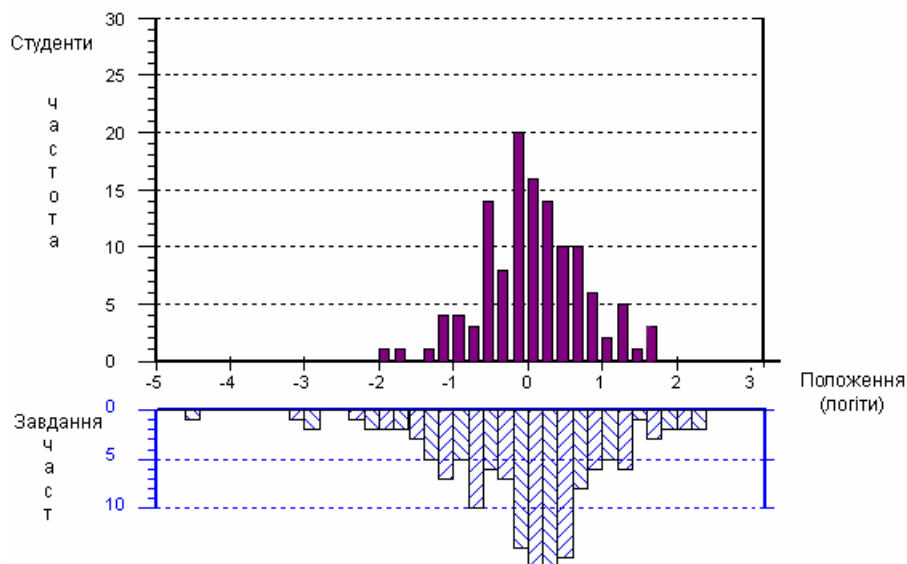


Рис. 3. Розподіл рівнів студентів та трудностей завдань. Тест № 2

У тесті № 2 інтервал трудностей завдань (від -2,5 до +2,5 логіта) практично співпадає з межами значень студентських рівнів підготовленості (рис. 2). Тобто тест № 2 добре збалансований за складностями. Інша його особливість полягає в тому, що він не такий однорідний, як тест № 1. Найбільш ймовірною причиною є те, що перший тест містить ряд завдань різної форми. Так, тут є завдання з перетягуванням графічних елементів, вказуванням відповідності, конструюванні формул з елементів, завдання з кількох логічних частин. Другий тест складається переважно із завдань з вибором однієї чи кількох правильних відповідей серед запропонованих. У цьому тесті частіше пропонується вибір з більшої кількості варіантів, ніж у першому тесті.

Незважаючи на вказані особливості в структурі тестів, вони обидва мають високі показники надійності, коефіцієнти розділення для них дорівнюють 0,852 і 0,962 відповідно. Такі високі значення говорять про те, що обидва тести виконують свою функцію.

Аналіз завдань та відповідей на них за допомогою моделі Раша дозволяє виявити завдання, які побудовані невдало та переробити їх. Для виявлення таких невдалих завдань необхідно провести аналіз відповідності теоретичній моделі. На графіку показані теоретична крива очікуваних значень і реальні отримані дані, що показують долю студентів, які виконали завдання (рис. 4 і рис. 5).

На рис. 4 показані точки для студентів з низьким (x), середнім (o) і високим (●) рівнем знань. Ці точки добре лягають на теоретичну криву; це показує, що завдання складене логічно коректно. Аналогічні точки відкладені на теоретичній кривій для іншого завдання на рис. 5. Видно, що точка, яка відповідає студентам з низьким рівнем (x) лежить набагато вище, а точка студентів з високим рівнем (●) лежить нижче кривої. Ситуація суперечлива, оскільки показує, що слабкіші студенти краще виконують завдання, ніж більш підготовлені. Це означає, що або завдання сформульоване невдало або ж тема була незрозумілою для всіх. Подальше обговорення разом зі студентами показало, що дана тематика ("Інерціальні системи") потребує додаткового розгляду та роз'яснення.

Висновки:

Педагогічною умовою оптимізації процесу складання коректних та надійних завдань і тестів є аналіз результатів тестування, що також придатний для моніторингу навчання при вивченні різних дисциплін. Застосований до балів за тестування з фізики, аналіз дав результати, що показують особливості навчального процесу та свідчать про ефективність завдань у тестовій формі. З розподілу завдань за трудністю видно, що введення різноманітних форм завдань розширює діапазон трудностей завдань, робить розподіл завдань за трудністю більш рівномірним.

Результативність виконання деяких завдань не відповідає прогнозованим за теоретичною моделлю. Розбіжність у реальних і очікуваних результатах пояснюється недостатнім засвоєнням навчального матеріалу, що підлягав тестуванню. У зв'язку з цим виникає потреба у з'ясуванні причин недостатньої якості навчальною діяльності.

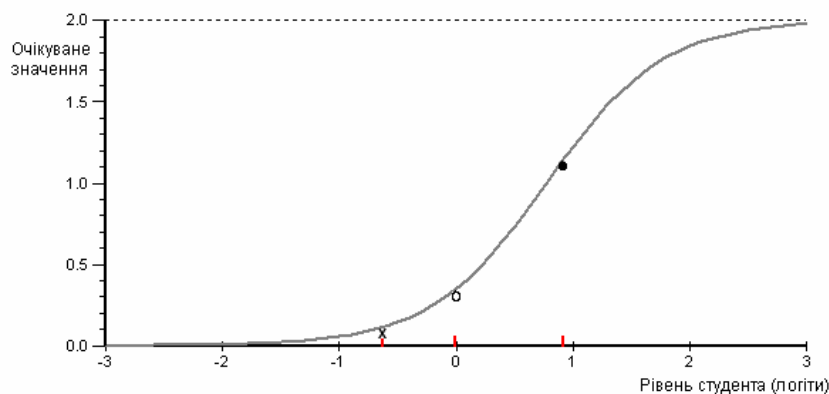


Рис. 4. Характеристична крива для завдання № 129

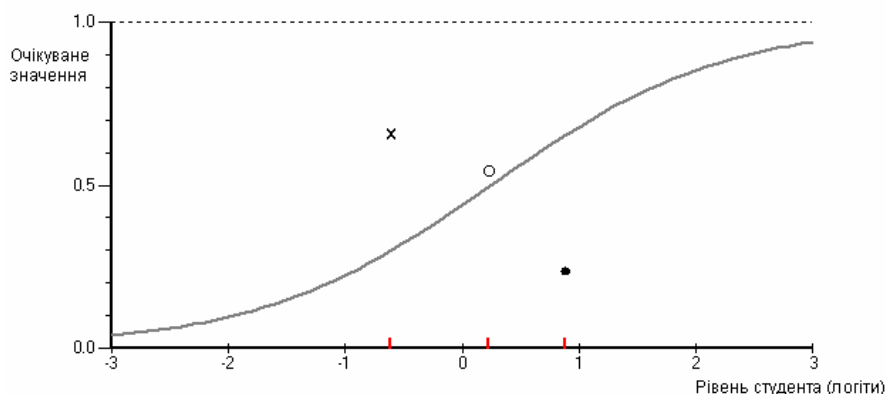


Рис. 5. Характеристична крива для завдання № 3

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ

1. Концептуальні засади розвитку педагогічної освіти України та її інтеграції в європейський освітній простір // Затверджено наказом МОН № 998 від 31.12.2004 р.
2. Embretston S.E. Test Design. – Orlando, San Diego, New York, London, Toronto, Montreal, Sydney, Tokio: Academic Press, Inc., 1979. – 222 p.
3. RUMM Laboratory Pty Ltd. RUMM2020. Retrieved from Rasch Unidimensional Measurement Model: <http://www.rummlab.com.au/> 2003.
4. Wainer H. Computerized Adaptive Testing. – Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. Publishers, 2000. – 335 p.

Матеріал надійшов до редакції 15.10. 2007 р.

Драган Е.В. Оптимизация заданий в тестовой форме на основе вероятностной модели.

Определены педагогические условия оптимизации заданий в тестовой форме, проведен статистический анализ результатов тестирования. Была использована однопараметрическая вероятностная модель Рашиа, с помощью которой определена надёжность системы заданий в целом, а также недостатки некоторых заданий. Проведённая работа дала возможность выявить некорректно построенные задания, найти пробелы в изученном материале и, как последующий результат, построить объективный тест.

Dragan Ye.V. The Optimization of Tasks in Test Form on the Basis of Probability Model.

Pedagogical conditions are specified for optimization of items in test form, statistical analysis of tests results was conducted. After Rasch unidimensional model applying reliability of the system of items was calculated, shortcomings of some items were revealed. The work conducted gives an opportunity to find items constructed inappropriately, reveal spaces in mastering of material and as the next stage to build an objective test.