

РІДНА

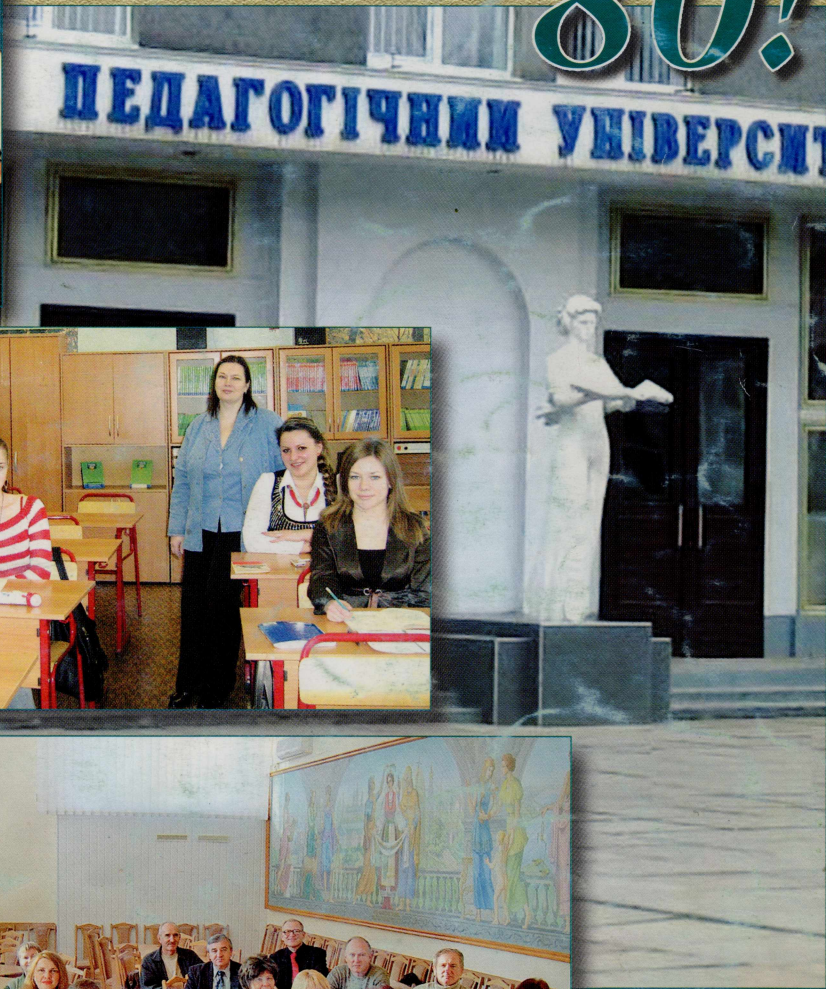
Щомісячний науково-педагогічний журнал
ISSN 0131-6788

ЛИПЕНЬ
СЕРПЕНЬ

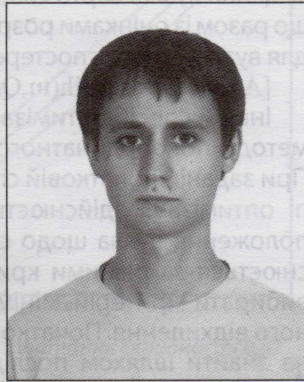
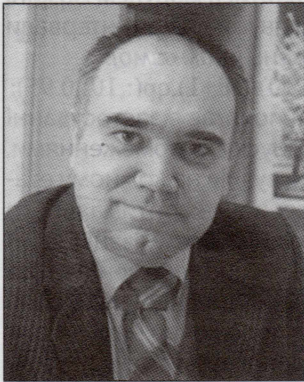
ШКОЛА

Криворізькому державному педагогічному університету –

80!



Слайн-моделі в тестових вимірюваннях знань



Ігор ШЕЛЕВИЦЬКИЙ,

доктор технічних наук, проректор з інформаційних технологій,

Роман ДУБАН,

інженер-програміст

Тестовий контроль знань набуває дедалі більшого поширення в сучасній українській освіті. Його перевагою є об'єктивність, простота перевірки, низька вартість, можливість дистанційного проведення. Однак створення якісних тестів є складним завданням, що потребує значних ресурсів. Основні вимоги до тестів і процедуру їх створення та перевірки викладено в нормативних документах МОН України [2], монографіях [1] і методичних розробках [5]. На жаль, під час тестування здебільшого застосовують тести, що не пройшли належної перевірки і мають невизначені властивості як інструмент вимірювання знань. Імовірною причиною цього є відсутність практично орієнтованих методів, алгоритмів і програмного забезпечення для оцінювання тестів.

Тестовий контроль дає змогу не лише визначити певний рівень знань респондента, а й спрогнозувати його можливість у вирішенні складних питань. Це особливо важливо у разі, коли від компетентності і вміння розв'язувати складні проблеми залежить життя і здоров'я людей, їхня безпека. Тому актуальним є розроблення математичних моделей, алгоритмів і програмного забезпечення, що вирішують ці завдання.

Сучасний тестологічний етап характеризується значною кількістю наукових досліджень, присвячених теорії тестування з погляду теорії ймовірності і статистики. Однією з центральних проблем досліджень є побудова адекватних моделей умовної ймовірності правильної відповіді на запитання. Найбільш відомими є однопараметрична функція Г.Раша, двопараметрична і трьохпараметрична функції А.Бірмабаума [7]. Важливою є побудова вказаних моделей за результатами тестування, однак

проблему становить нелінійна залежність зазначених моделей від параметрів.

Під час тестування респондентів з метою виявлення рівня їхніх знань отримана кількість балів залежить одночасно від складності тестових завдань і підготовки респондента. Узагальнено можна говорити про двовимірну функцію залежності ймовірності правильної відповіді від складності запитання й рівня знань респондента. Сумарний отриманий бал містить як деяку закономірну складову, так і випадкову. Вибір функціональної залежності ймовірності правильної відповіді від характеристик запитання й респондента довільний. Тому є значна кількість різноманітних моделей, які залежать від багатьох параметрів. Для опису ймовірності правильної відповіді i -го респондента на j -й запитання використовуємо функцію Г.Раша [8]:

$$P_{i,j} = \frac{e^{Dv_j(a_i - Q_j)}}{1 + e^{Dv_j(a_i - Q_j)}}$$

де: a_i – рівень знань респондента; Q_j – рівень складності запитання; v_j – коефіцієнт диференціації знань запитанням; D – числовий параметр, що нормує ймовірність.

Коефіцієнти a_i і Q_j є латентними ознаками, які безпосередньо не спостерігаються. Процедура їхнього визначення потребує об'ємних тестових експериментів з оцінюванням рівня знань і рівня складності запитань. Ці показники можна вимірювати у відповідних безумовних ймовірностях або в логітах. Якщо a_i ймовірність, то логіт одержують згідно з $\ln\left(\frac{a_i}{1-a_i}\right)$. Зворотне перетворення таке: $e^{a_i}/(1+e^{a_i})$. При статистичному опрацюванні результатів тестування краще користуватися ймовірностями, оскільки нелінійне перетворення на логіти зумовить ненормальність розподілу випадкової складової. Це ускладнить побудову довірчих інтервалів для оцінок. Нормальність випадкової складової в результатах тестування визначається центральною граничною теоремою.

Розглянута модель є основою, за допомогою якої досліджуються алгоритми оцінювання характеристик тесту. Для моделювання процесу тестування й дослідження алгоритмів опрацювання його результатів розроблено програмний модуль «tests» у середовищі MATLAB. Складність запитань і рівень знань респондентів задаються випадково і вважаються розподіленими за нормальним законом. Правильна відповідь формується як випадкова подія в окремому досліді, що має біноміальний закон розподілу. Опрацювання результатів полягає у визначенні середніх балів респондентів, які є

оцінками рівня знань. Середній відсоток правильних відповідей слугує для оцінювання складності запитань. Середній відсоток правильних відповідей на запитання респондентів з однаковими оцінками рівня знань слугує для оцінювання ймовірності правильної відповіді на окремі запитання тесту респондентів із відомими оцінками рівня знань. Останні точки утворюють профілі тестових запитань.

Отриманий емпіричний профіль містить значну шумову складову. У реальній ситуації істинні профілі невідомі. Можна побудувати лише їхню оцінку за окремими емпіричними даними. Тобто необхідно вирішити завдання згладжування емпіричних даних деякою функцією. Застосування функції Г.Раша пов'язане з труднощами оцінювання параметрів, які входять у рівняння нелінійно. Вибір профілюючої функції значною мірою довільний і ґрунтується на інтуїції й досвіді оператора, а це ускладнює автоматизацію процедури. Однак є функції, які характеризуються оптимальною адаптивною здатністю у поєднанні з простотою розрахунків і лінійною залежністю від параметрів, – сплайни [4]. У практиці опрацювання даних найбільш відомі інтерполяційні кубічні сплайни з двома неперервними похідними [3]. Для їхнього розрахунку необхідно розв'язати систему інтерполяційних рівнянь. Зручніше користуватися кубічними ермітовими сплайнами, які мають неперервність лише першої похідної і є локальними. При інтерполяції з їхньою допомогою немає потреби розв'язувати систему рівнянь. Зміст параметрів моделі збігається зі змістом вхідних даних – точок спостережень [6].

Для оцінювання вектора A – параметрів сплайнової моделі – застосуємо метод найменших квадратів (МНК). Такий вибір зумовлено нормальністю випадкової складової і простотою методу. Параметрами оцінювання є значення сплайна в точках стикування фрагментів. Особливість сплайнів – в істотній залежності їхньої форми від кількості і схеми розміщення точок стикування – вузлів: (tu, A) , тому алгоритм має це враховувати. Є два варіанти вирішення. Розглянемо послідовну побудову сплайна з контролем відсутності невідповідної складової в залишках наближення. Звичайно, перевірка випадковості має характер перевірки статистичної гіпотези, для чого використаємо непараметричний критерій інверсій. Фрагмент сплайна продовжується доти, доки не буде відкинуто гіпотезу про випадковість залишків. Позитивом такого алгоритму є автоматичне визначення числа вузлів. Імовірність помилки першого роду варто вибирати так, щоб імовірність випадковості залишків була максимальною.

Послідовна побудова сплайна за МНК реалізує модуль `sersplm`, приклад застосування якого показано нижче:

```
alfa=0.5; % рівень помилки першого роду
[tu, A, Am, Ap, PInv]=sersplm(Q(:,nt)/(M+1),qp(:,nt),0,alfa); % послідовна побудова сплайна
sp=splherm(r,tu,A); % одержання графіка оціненого профілю
```

```
p=rashp(r,ptp(nt)); % одержання істинного профілю
```

```
plot(r,p,r,sp,tu,Am,'o',tu,Ap,'o',Q(:,nt)/(M+1),qp(:,nt),'+');title([' quest ',num2str(nt)]);
```

Для одержання МНК оцінок при заданій схемі розміщення вузлів варто скористатися модулем `lsqsplh`, що разом із оцінками розраховує й довірчі інтервали для вузлів і точок спостережень [там само]:

```
[A,di,s,ds]=lsqsplh(tu,Q(:,10)/(M+1),qp(:,10),0.95);
```

Інший спосіб оптимізації полягає в застосуванні методу покоординатного спуску з обмеженнями. При заданій початковій сітці вузлів на кожному етапі оптимізації здійснюється пошук оптимального положення вузла щодо сусідніх. Оптимізація здійснюється за різними критеріями. Рекомендується вибирати критерій мінімуму середньоквадратичного відхилення. Початкове положення вузлів можна знайти шляхом послідовної побудови, а уточнити їх – покоординатною оптимізацією. Приклад виклику модуля оптимізації показано нижче:

```
[wu,wi,op]=optspl(tu,Q(:,4)/(M+1),qp(:,4),4);
```

На рис. 1 зображено теоретичні профілі трьох тестових запитань зі складностями 0.238, 0.57, 0.90 і, відповідно, емпіричні результати моделювання процесу тестування. Узагальнення опрацювання модельних результатів показано на рис. 2–4. Сплайнова модель досить точно передає характерні риси профілів тестів з різними рівнями складності. При цьому забезпечується майже достатній рівень вірогідності. Іншим важливим аспектом моделі є компактне подання моделі профілю тестових запитань. Уся інформація про моделі міститься в значеннях вузлових точок сплайна. Вірогідність

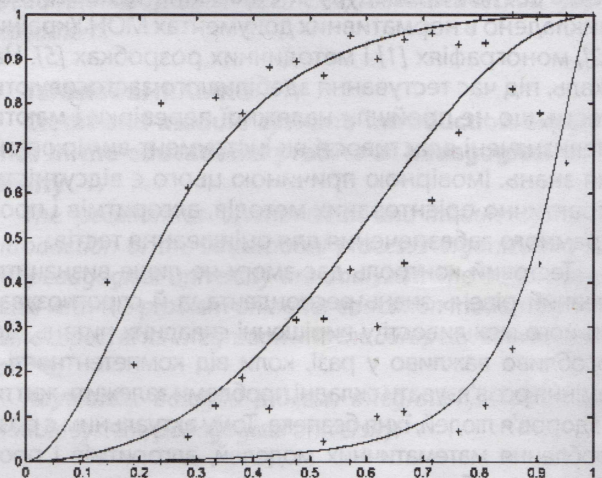


Рис. 1. Теоретичні профілі запитань і змодельовані емпіричні результати тестування

моделі описується шириною довірчих інтервалів у точках стикування. Подані алгоритми оцінювання характеристик запитань досить загальні й не враховують деяких особливостей даних, що опрацьовуються. Їхнє подальше вдосконалення полягає в застосуванні зваженого МНК (дані є оцінками частоти з різним обсягом вибірок) і обліку обмежень рішення (у межах від 0 до 1).

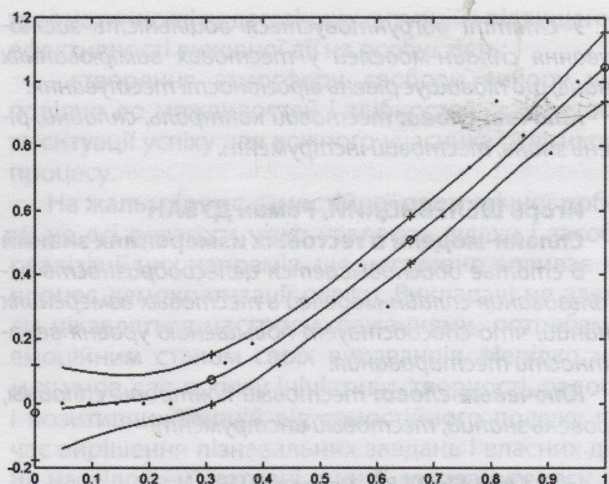


Рис. 2. Оцінювання профілю запитання складності 0.57

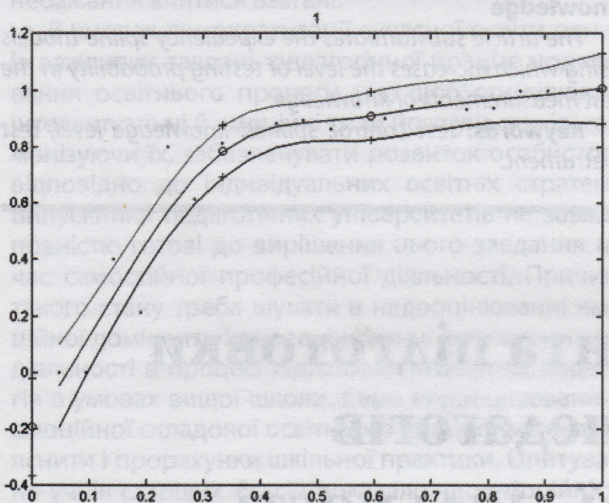


Рис. 3. Оцінювання профілю запитання складності 0.24

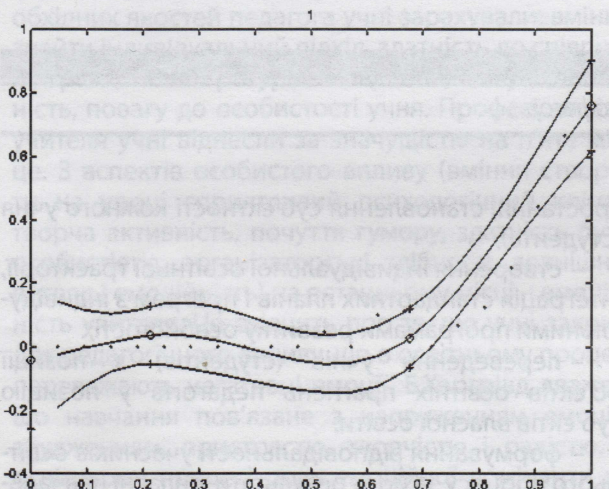


Рис. 4. Оцінювання профілю запитання складності 0.90

Оцінювання профілів тестових запитань потребує значної кількості респондентів із різним рівнем знань, тому доцільною є ітераційна процедура оцінювання характеристик тесту. Перший етап полягає у проектуванні й розробленні тесту й початковому тестуванні. Методика такого аналізу також відома.

На цьому етапі виявляють найбільш очевидні недоліки й визначають загальні характеристики за допомогою контрольної групи респондентів. Для прискорення уточнення характеристик тесту і зниження вартості цієї процедури доцільно залучити до цього процесу якомога більше учасників. Процедура аналогічна бета-тестуванню програмного забезпечення. Тести передаються бета-тестувальникам за умови їхньої зацікавленості і відповідальності. Оперативний зв'язок доцільно здійснювати через електронну пошту або інтерактивні сторінки.

До розробників тесту має надходити інформація про характеристики тестових груп і результати тестування. Після досягнення необхідного рівня вірогідності оцінювання профілів тестових запитань виконується оптимізація тесту (наприклад, вилучення малоінформативних запитань і запитань зі статистично близькими профілями). Формується паспорт тесту, що містить профілі запитань. Наявність цих характеристик дає можливість об'єктивніше оцінювати рівень знань респондентів, а тест можна вважати завершеним інформаційним продуктом із нормованими характеристиками. Тобто одержуємо своєрідний прилад, що вимірює знання в певній галузі.

Звичайно, бажано встановити стандарти на такі вимірювальні інструменти. Це можливо за умови широкої дискусії зацікавлених сторін щодо методологічних, організаційних, соціальних і юридичних аспектів цієї справи. З технічного погляду, тестовий інструмент може ґрунтуватися на ієрархічній структурі об'єктів, що має таку структуру:

- **тест:** сфера знань (наприклад, за УДК), призначення тесту, категорія респондентів, e-mail авторів, список авторів, анотація авторів, дата створення, дата перегляду, умови ліцензування, запитання;
- **запитання:** категорія, група, текст запитання, варіанти відповідей, правильний варіант відповіді, оцінювання профілю запитання;
- **група респондентів:** дата проведення тестування, реквізити відповідального за тестування, анотація групи, апріорне оцінювання рівня знань групи, первинне або повторне тестування;
- **респондент:** ідентифікатор респондента, апріорний рівень знань, професійні реквізити респондента, відповіді на тестові запитання, сумарний бал, оцінювання рівня знань респондента, оцінювання профілю респондента.

Такий підхід до формування, випробування й застосування тестів може гарантувати необхідний (або заданий) рівень вірогідності тестування, перетворивши тест на вимірювальний інструмент із нормованими характеристиками щодо похибок.

Література

1. Доровської В.А. Дихотомічне тестування й оцінювання ефективності навчання диспетчерів поліергатичних систем / В.А. Доровської. – Кривий Ріг, 2001. – 412 с.
2. Комплекс нормативних документів для розробки складових систем стандартів вищої освіти. Додаток 1 до наказу Міністерства №285 від 31 липня 1998 р.