

**ОЦІНЮВАННЯ ЯКОСТІ ТЕСТІВ НА ОСНОВІ СПЛАЙН-МОДЕЛЕЙ
ПРОФІЛІВ**

© Роман Дубан, 2011

Основна увага приділяється тесту як інструменту вимірювання знань і необхідності його попереднього неодноразового вдосконалення та оцінки якості тесту і властивостей кожного з питань до початку його використання. Особливість підходу у відкритій моделі формування тесту та сплайн-моделі профілів.

Ключові слова: інформаційна система, тестовий контроль знань, сплайн-модель.

Emphasis is placed on the test as a tool for measuring the knowledge and the need to pre-refillable improvement, quality assessment and the properties of each of his subject before its use. Feature of the approach lies in the open model of the formation test and the spline-model profiles.

Keywords: information system, test control knowledge, spline-model.

Вступ. Важливе місце серед інформаційних освітніх технологій займає тестовий контроль знань. Відомими перевагами тестового контролю знань є простота процесу перевірки, об'єктивність, низька вартість процесу тестування, можливість дистанційного проведення. Однак створення якісних тестів є складною задачею, що потребує значних ресурсів. Основні вимоги до тестів та процедура їх створення і перевірки викладені в комплексі нормативних документів МОН [1], монографіях [2] та навчальних посібниках [3]. На жаль, в переважній більшості випадків, при тестуванні застосовуються тести, що не пройшли належної перевірки і мають невизначені властивості, як інструмент вимірювання знань. Ймовірною причиною такого положення є відсутність практично зорієнтованих методів, алгоритмів і програмного забезпечення для оцінювання тестів.

Наявний інструментарій розробки тестів. В умовах достатнього розвитку комп'ютерних технологій процес складання тестів і тестових завдань доцільно виконувати з використанням комп'ютерної техніки та баз даних. Найбільш відоме та вживане вільне програмне забезпечення – це Moodle, iTest, та OpenTEST2, що дозволяють створювати бази тестів та проводити тестування з використанням комп'ютерної техніки, а в OpenTEST2 і проведення статистичного аналізу результатів. Комерційні програмні продукти

забезпечують технічну підтримку у вирішенні проблем і мають покращений вигляд інтерфейсу та вдосконалений функціонал. Але сам підхід до створення тестів в цих та інших програмних засобах має закритий характер, автор чи група авторів розробляють тест і використовують його для вимірювання знань без попереднього рецензування та випробування. Інформаційні системи які охоплюють весь життєвий цикл тесту від створення до використання невідомі.

Є значна кількість наукових досліджень присвячених теорії тестування з погляду теорії ймовірності та статистики. Однією з центральних проблем досліджень є побудова адекватної моделі умовної ймовірності правильної відповіді на запитання. Найбільш відомими моделями є однопараметрична функція Г.Раша [4], двох параметрична та трьох параметрична функції А.Бірнабаума [5]. Ці моделі є основою, за допомогою якої досліджуються оцінки характеристик тесту. Практично важливою є побудова вказаних моделей за результатами тестування. Однак серйозною складністю є нелінійна залежність вказаних моделей від параметрів.

Розробка нової інформаційної системи. З огляду на перераховані проблеми постає задача проектування інформаційної системи підтримки тестового контролю, що охоплює весь життєвий цикл тесту. Тест обов'язково, до його використання в якості вимірювального інструмента, має пройти стадії створення, рецензування, багаторазового випробування та вдосконалення. Лише після отримання тесту з нормованими характеристиками якості, його можна рекомендувати до використання. Така система має бути максимально відкритою і безкоштовною, створеною на засадах вільного програмного забезпечення. Основною метою такої системи є спрощення процедур тестового контролю і одночасно підвищення їх якості. З огляду на те, що інформаційна система підтримки тестового контролю знань має бути максимально відкритою та безкоштовною, технології що використовуються для її створення доцільно обирати серед вільного програмного забезпечення. Для забезпечення спрощеного доступу до системи планується використовувати web-технології.

У зв'язку з тим, що до процесу розробки тесту має залучатись група користувачів системи з різними правами доступу, система обов'язково повинна мати можливість авторизації користувачів та розподілу прав. Адміністратор інформаційної системи – технічний спеціаліст, що володіє знаннями про функціонування системи та виконує її обслуговування, допомагає у вирішенні питань всім користувачам, виправляє помилки виявлені в програмі та розвиває її функціонал. Менеджери – найбільш досвідчені та поважні користувачі системи, що розуміються на розробці тестів та, здебільшого, мають науковий ступінь, вся відповідальність за просування тесту на наступну стадію лежить на них. Звичайні зареєстровані користувачі інформаційної системи можуть

мати певні ролі або їх комбінації, а саме: тестувальник, рецензент, автор. Менеджер має всі можливі з перерахованих ролей користувачів та може надавати їх будь якому користувачу. Гості є незареєстрованими користувачами системи, що можуть лише переглядати списки тестів та, можливо, проходити тестування. Роль тестувальника надає можливість користувачу проводити тестування на групі респондентів на етапі випробування тесту, після чого він має внести опис групи та результати. Роль рецензента надає можливість користувачу переглядати цікаві йому тести їх розділи та питання, та залишати рецензії на них у вигляді зауважень або коментарів. Автор має змогу створювати нові тести, визначати їх розділи та наповнювати їх питаннями, публікувати чи знімати з публікації питання, відповідати на рецензії. Автор має змогу запропонувати іншому автору, в цікавий йому тест, своє питання, але опублікувати це питання має змогу лише автор тесту.

Новий тест створений автором має містити тему та опис. Будь-який менеджер, краще якщо спеціаліст по темі тесту, вирішує чи потрібен такий тест і публікує його, призначивши відповідальним за контроль над тестом менеджера – «адміністратора тесту». Задача адміністратора тесту полягає у відповідальності за переведення тесту на наступну стадію «життєвого циклу». Першою стадією – є «розробка», на яку тест переходить відразу після його публікації і призначення менеджером адміністратора. На першій стадії автор тесту визначає розділи та наповнює їх питаннями, інші автори можуть запропонувати свої питання до певних розділів тесту, автор тесту може їх або прийняти із збереженням авторства або відхилить, рецензенти мають переглядати питання та робити зауваження і коментарі до тесту в цілому і до окремих питань. Після того, як тест має достатню кількість питань, не менш як тридцять на кожний розділ, та прийняті до уваги усі зауваження рецензентів, або подані пояснення до них, автор може зробити запит до адміністратора тесту на перехід до наступної стадії. Адміністратор тесту перевіривши виконання необхідних умов переводить тест на наступну стадію – «випробування», або висловлює свої зауваження. На етапі «випробування» проводиться тестування користувачами тестувальниками на групах респондентів у звичайному паперовому вигляді, або з використанням комп'ютерної техніки. Важливим на цьому етапі є використання усіх питань тесту, для цього зручнішим буде проведення тестування по розділах. Після отримання перших результатів тестування починається процес статистично-математичного аналізу по певним групам респондентів. З надходженням результатів наступних тестувань розрахунки будуть уточнюватись а їх достовірність збільшуватиметься. Отримані показники мають

використовуватись рецензентами та самим автором для роботи над вдосконаленням тесту.

Профілі складності питань та рівня знань респондентів. Застосування функції Раша пов'язане із труднощами оцінки параметрів, які входять у рівняння нелінійно. Вибір профільюючої функції значною мірою довільний і ґрунтується на інтуїції й досвіді оператора, а це утрудняє автоматизацію цієї процедури. Однак існують функції, які мають гарну адаптивну здатність у сполученні із простотою розрахунків і лінійною залежністю від параметрів – це сплайни. У практиці обробки даних найбільш відомі інтерполяційні кубічні сплайни із двома неперервними похідними. Для їхнього розрахунку необхідно вирішити систему інтерполяційних рівнянь. Зручніше користуватися кубічними ермітовими сплайнами, які мають неперервність лише першої похідної і локальні. При інтерполяції з їхньою допомогою немає необхідності вирішувати систему рівнянь. Зміст параметрів моделі співпадає із змістом вхідних даних: точок спостережень [6]. Іншим важливим аспектом моделі є компактне подання моделі профілю тестових питань. Вся інформація про моделі міститься в значеннях вузлових точок сплайна. Вірогідність моделі описується шириною довірчих інтервалів у точках стикування. Представлені алгоритми оцінювання характеристик питань досить загальні й не враховують деяких особливостей оброблюваних даних (рис 1.).

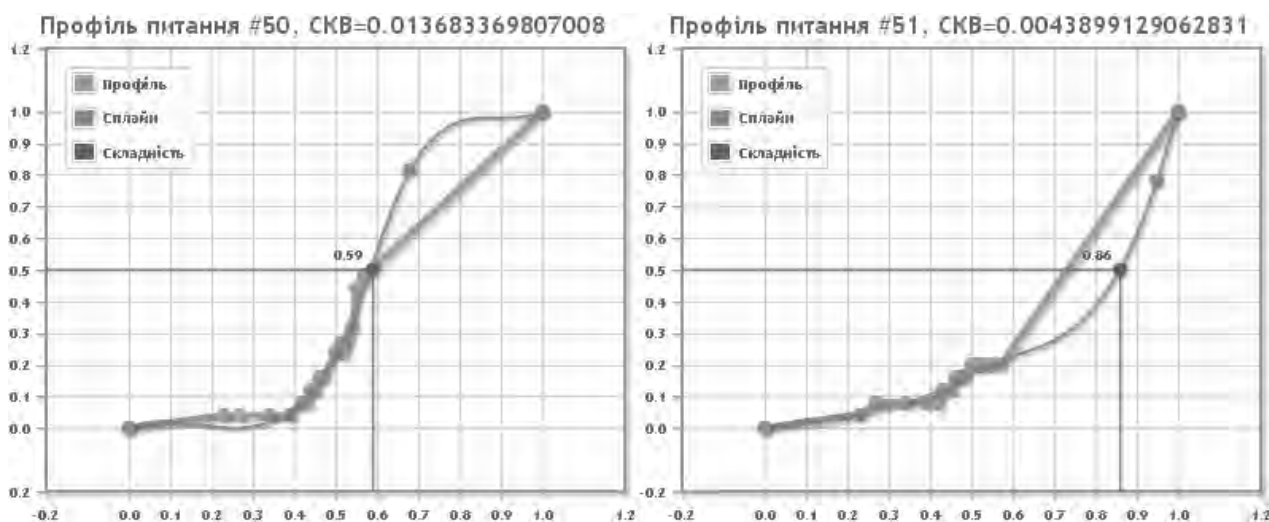


Рис. 1. Профілі питань та сплайн-моделі, що їх описують

Подальше вдосконалення алгоритму полягає в застосуванні складеного кубічного ермітового сплайн з накладенням умов, що більш адаптований до особливостей даних. Застосування зваженого МНК нажаль не дало очікуваних результатів.

Оцінка профілів тестових питань вимагає значного числа респондентів з різним рівнем знань. Тому практично доцільною є ітераційна процедура оцінки характеристик тесту. Спочатку виявляють найбільш очевидні недоліки й установлюються загальні характеристики на деякій контрольній групі респондентів.

При досягненні необхідного рівня вірогідності оцінок профілів тестових питань виконується оптимізація тесту (наприклад, вилучення малоінформативних питань і питань зі статистично близькими профілями). Формується паспорт тесту, що включає профілі питань. Наявність цих характеристик дає можливість більш об'єктивно оцінювати рівень знань респондентів. Із цього моменту тест можна вважати завершеним інформаційним продуктом з нормованими характеристиками. Тест переводиться на стадію “застосування”, а зміни в нього не вносяться, за винятком умов, що потребують перегляду тесту. Тобто одержуємо своєрідний вимірювальний прилад, що вимірює знання в певній області.

Висновки. Такий підхід до формування, випробування й застосування тестів здатний гарантувати необхідний (або заданий) рівень вірогідності тестування, перетворивши тест у вимірювальний інструмент із нормованими характеристиками щодо похибок. Забезпечення функціональних можливостей для цього процесу планується реалізувати на web-платформі, що дозволить оперативно обмінюватись інформацією між учасниками. Результати поточних розробок інформаційної системи “Logit” подані на сайті за адресою <http://kdpu.edu.ua/logit/>.

Література

10. *Комплекс нормативних документів для розробки складових систем стандартів вищої освіти. Додаток 1 до наказу Міносвіти N 285 від 31 липня 1998 р.* – К. 1998. – 115 с.
11. *Доровской, В. А. Дихотомическое тестирование и оценивание эффективности обучения диспетчеров полиэргатических систем / В. А. Доровской.* – Кривой Рог, 2001. – 412 с. – ISBN 966-7508-39-0.
12. *Чельшкова, М. Б. Теория и практика конструирования педагогических тестов : учебное пособие / М. Б. Чельшкова.* – М. : Логос, 2002. – 432 с.
13. *Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. With a Foreword and Afteword by BD Wright. The Univ. of Chicago Press. Chicago & London. 1980, 199 p.*
14. *Birbaum A. Some Latent Trait Models and Their Use in Statistical Theories of Mental Test Score. Reading, Mass.: Addison-Wesley, 1968.*

15. Шелевицкий, И. В. Слайды в цифровой обработке данных и сигналов [Текст] / И. В. Шелевицкий, М. О. Шутко, В. М. Шутко, О. О. Колганова. – Кривий Ріг : Видавничий Дім, 2008. – 232 с. – ISBN 978-966-2915-86-0.

УДК 378.14

**Ірина Алексєєва, Віктор Гайдей, Олександр Диховичний,
Наталія Коновалова, Лідія Федорова**
Національний технічний університет України «КПІ»

СТАТИСТИЧНИЙ АНАЛІЗ ТЕСТОВИХ ЗАВДАНЬ ІЗ ЗАСТОСУВАННЯМ СУЧАСНИХ МАТЕМАТИЧНИХ МОДЕЛЕЙ

© Ірина Алексєєва, Віктор Гайдей, Олександр Диховичний, Наталія Коновалова, Лідія Федорова, 2011

У праці проінформовано про застосування сучасних математичних моделей тестів до аналізу якості тестових завдань комплекту дистанційної освіти «Вища математика», розробленого в НТУУ «КПІ». Наведено методику оцінювання латентних параметрів і приклади її застосування.

Ключові слова: тестові завдання, латентні параметри, моделі Раша.

The present paper informs about application IRT to the Item Analysis for outlined web-based courses «Higher mathematics» made in NTUU “KPI”. Methodic of estimating latent parameters and examples of application are given.

Keywords: Item Analysis, IRT, Rasch models.

Вступ. Сучасний період розвитку освіти, у тому числі вищої технічної, характеризується всебічним поширенням тестового підходу до контролю знань і потребує застосування сучасних статистичних математичних методів до аналізу якості тестів як в цілому, так і окремих тестових завдань.

Кафедрою математичного аналізу та теорії ймовірностей НТУУ «КПІ» було створено комплект дистанційної освіти «Вища математика» [1]. Комплект містить розгалужену базу тестових завдань по всіх розділах, на основі якої формуються електронні контрольні роботи та іспити. В зв'язку з цим значну увагу приділяється аналізу якості тестових завдань комплекту із застосуванням як класичних методів статистичного аналізу, так і сучасної математичної теорії параметризації тестових завдань, яка носить назву Item Response Theory (IRT) [2].

Статистичний аналіз тестів. В основу статистичного аналізу було покладено математичні моделі тестових завдань Г. Раша та А. Бірнбаума, які