

Вронський Сергій Володимирович

Аспірант кафедри системного аналізу та теорії оптимізації,

<https://orcid.org/0000-0002-9960-6437>

ДВНЗ «Ужгородський національний університет, Ужгород

Андрашко Юрій Васильович

Доцент кафедри системного аналізу та теорії оптимізації,

<https://orcid.org/0000-0003-2306-8377>

ДВНЗ «Ужгородський національний університет, Ужгород

МОДЕЛІ ТА МЕТОДИ АДАПТИВНОГО ТЕСТУВАННЯ ЯК ІНСТРУМЕНТ ІНФОРМАТИЗАЦІЇ СИСТЕМИ КОНТРОЛЮ ЗНАНЬ У ЗАКЛАДАХ ВИЩОЇ ОСВІТИ

***Анотація.** Дослідження присвячене огляду сучасних моделей та методів, що використовуються для аналізу ефективності тестової форми контролю знань. Наведено опис теоретичних підходів до оцінювання результатів здійснюваного тестування. Розглянуто математичні моделі сучасної теорії тестування, що описують взаємозв'язок між навичками здобувачів та їх практичним відображенням при виконанні тестових завдань. Наведено опис загальних принципів адаптивного тестування з використанням дихотомічних та політомічних завдань для контролю знань та вмінь здобувачів, набутих у процесі навчання. Проаналізовано тестовий підхід до оцінки рівня знань, оскільки його застосування передбачає об'єктивне оцінювання знань і врахування різноманітних аспектів навчальної діяльності: уміння впізнавати і розуміти матеріал, класифікувати явища й процеси, визначати послідовність логічних операцій, систематизувати і синтезувати ситуації і процеси, в подальшому – проводити самостійні дослідження та застосовувати отримані знання на практиці. У статті досліджено специфіку застосування різноманітних підходів, що описують теоретичну основу функціонування сучасної теорії тестування: однопараметричних моделей Раши та Фергюсона, двопараметричної та трипараметричної моделей Бірнбаума. У науковому дослідженні наведено огляд алгоритму роботи та специфіку використання засобів адаптивного тестування, що використовується для визначення загального рівня знань з певних навчальних дисциплін; розглянуто математичний апарат, що застосовується безпосередньо для формування блоків завдань відповідного рівня складності та для інтерпретації результатів перевірки знань здобувачів різного рівня підготовки. Особливу увагу приділено автоматичному підбору завдань необхідного рівня складності, що безпосередньо базується на результатах тестування в режимі реального часу. Такий алгоритм дозволяє виявити рівень знань тестованого з конкретної дисципліни, водночас оцінюючи його психометричні характеристики: рівень зосередженості, вміння протистояти стресовим ситуаціям, швидкість перемикання між завданнями різного типу та рівня складності. Також важливим є попередній розподіл завдань за блоками відповідно до типу сформульованої умови виконання завдання та кількості логічних кроків, що необхідно здійснити для правильного розв'язання завдання. Отриманий в результаті повного проходження тестування профіль рівня знань тестованого дозволяє виявити не тільки прогалини в знаннях з певного предмету, а й оцінити рівень психомоторної адаптації до процесу тестування в процесі виконання завдань.*

***Ключові слова:** сучасна теорія тестування; адаптивне тестування; якість тесту; тестова форма контролю рівня знань; математичні моделі оцінювання якості тесту*

Постановка проблеми

Проведення моніторингу знань, здобутих у результаті навчання, є невід'ємною складовою освітнього процесу, оскільки саме він забезпечує

можливість об'єктивної оцінки рівня підготовки здобувачів освіти. Такий моніторинг охоплює не лише аналіз засвоєння теоретичних положень і формування практичних навичок, але й здатність застосовувати їх у нових умовах та при розв'язанні

завдань, що виникатимуть у майбутній професійній діяльності. Тобто, йдеться про необхідність комплексного підходу, який включає як перевірку фактичних знань, так і оцінку гнучкості мислення, уміння робити висновки та аргументовано обґрунтовувати власний вибір.

У сучасних умовах, коли освітній процес характеризується масовістю, динамічністю та високим рівнем варіативності, традиційні форми контролю (усні співбесіди, письмові роботи, індивідуальні завдання) часто виявляються трудомісткими, суб'єктивними й недостатньо ефективними. Тому все більшої актуальності набуває застосування тестових систем контролю знань, які дозволяють стандартизувати процедуру оцінювання, забезпечити її прозорість та оперативність, а також значною мірою знизити вплив людського чинника.

Суть тестового контролю полягає у виконанні завдань різних рівнів складності з можливістю вибору правильної відповіді з кількох запропонованих варіантів або у формулюванні власного розв'язку з подальшим обґрунтуванням. При цьому важливим є не лише факт вибору правильного варіанта, а й математичні методи, що лежать в основі побудови тестів, алгоритмів оцінювання та статистичної інтерпретації результатів. Саме якість і адекватність цих методів визначають достовірність та справедливність висновків про рівень підготовки здобувачів.

Водночас тестові системи не позбавлені недоліків. До них можна віднести ризик формального підходу, коли здобувач орієнтується на механічне запам'ятовування готових відповідей; можливість випадкового отримання високого результату за рахунок вгадування; складність у вимірюванні творчих і дослідницьких компетентностей. Тому виникає потреба у критичному аналізі математичного апарату, який використовується в тестових системах, та його порівнянні з традиційними педагогічними методами оцінювання успішності. Це дозволить виявити сильні та слабкі сторони обох підходів і запропонувати напрями їхнього вдосконалення.

Мета статті

Метою даної статті є аналітичний огляд математичних методів, моделей та засобів, що використовуються при оцінюванні рівня підготовки здобувачів у системах тестової перевірки знань, розробка пропозицій щодо вдосконалення їх використання.

Аналіз основних досліджень і публікацій

Проблемам тестування як методу оцінки рівня знань присвячені дослідження таких вітчизняних

науковців, як М. В. Карташов [1], І. І. Оксенюк [2], Н. В. Сачанюк-Кавецька та О. П. Прозор [3], С. С. Шевчук [4]. В їх роботах розглянуто питання теорії створення сучасних тестів, проаналізовано тестування з точки зору методу оцінювання навчальних досягнень та вимірювання рівня якості засвоєних у процесі навчання знань. Вивченню проблеми розроблення принципів тестової перевірки рівня знань здійснювали іноземні вчені, серед яких: С. L. Bae [5], W. Linden [7], G. Rasch [9], W. Wang [13], M. Wiberg [14] та велика кількість інших науковців.

Незважаючи на значну увагу до проблеми, є питання, що потребують подальшого теоретичного і практичного вивчення, особливе місце серед яких займають дослідження математичної складової аналізу ефективності алгоритму системи тестового контролю рівня знань. Оскільки вдосконалення та впровадження нових форм та методів навчання вимагають вдосконалення та введення нових форм контролю, можна сказати, що сучасна, правильно побудована та структурована система контролю сприяє покращенню якості підготовки молодих фахівців.

Виклад основного матеріалу

У дослідженні запропонована методика зменшення рівня впливу сторонніх факторів на результати проведеного тестування, що дозволить не тільки більш ефективно підійти до формування нових видів і типів тестових завдань, проте й дієвіше виявляти рівень знань тестованих, що насправді відповідає їхньому рівню здібностей та навичок.

Разом із повсюдним застосуванням тестового контролю знань в освіті, більше уваги стали приділяти і характеристикам тесту, які забезпечують точність та достовірність одержаних результатів виміру.

Контроль знань та рівня підготовки здобувачів дедалі частіше реалізовується з використанням технологій педагогічного тестування. Об'єктивність, надійність і стійкість оцінок, одержуваних під час тестування, залежить, передусім, від якості розроблених тестів, оскільки саме тест і тестові завдання виступають інструментом оцінювання латентних (прихованих) характеристик здобувача.

Використання тестів дозволяє здійснювати навчальний контроль за успішністю здобувачів, простежувати динаміку засвоєння навчального матеріалу та здатність використовувати здобуті знання при вирішенні практичних завдань (таблиця).

Таблиця – Види та суть навчального контролю

Вид навчального контролю	Суть контролю
Вхідний	Дає можливість викладачу планувати та формулювати пріоритетні напрями подальшої навчальної діяльності
Поточний	Відстеження рівня засвоєння знань і вмінь здобувачів
Рубіжний	Контроль знань після завершення вивчення визначених тем (розділів) або наприкінці семестру
Контроль залишкових знань	Перевірка залишкових знань та збереженості вмінь через деякий час після вивчення теми, розділу або курсу
Підсумковий	Екзаменаційний тест, що охоплює теми всього навчального курсу

Складено автором на основі джерела: [4]

Кожен із вказаних типів тестового навчального контролю дозволяє оцінити не тільки ступінь засвоєння нового навчального матеріалу, проте також виявити, наскільки нові отримані знання доповнили поточну систему володіння матеріалом із певної наукової дисципліни.

В даний час оцінювання результатів тестування проводиться за допомогою класичного математичного апарату (classical test theory) або сучасної теорії тестування (item response theory – IRT).

Кожна з даних теорій має власні переваги та недоліки [11]. Основною відмінністю класичної теорії від сучасної є підхід до оцінювання здобувача, згідно з яким отриманий результат – це підсумковий бал з конкретного тесту з поправкою на можливо допущену помилку.

Проте існує проблема апроксимації функціями, що використовуються представленими моделями, емпіричних даних тестування. Функції нелінійно залежать від параметрів. У зв'язку з цим необхідно застосовувати методи нелінійної оптимізації та оцінювання, які не гарантують отримання глобального мінімуму критерію наближення. Тому процес оцінювання параметрів наявними методами проводиться індивідуально для кожного тестового завдання та тестованого під контролем досвідченого оператора. Поряд із зазначеними функціями IRT розроблені та досліджуються інші різноманітні функції [13]. Ця різноманітність ілюструє відсутність функції, яка здатна досить повно задовольнити наявні теоретичні та практичні потреби тестового контролю. Особливі труднощі виникають, коли

емпіричні дані виявляються далекі від очікуваних теоретичних. У класичній теорії кількість тестових балів, набраних здобувачами, залежить від рівня складності завдань у тесті, а складність рівня завдання – від вибірки здобувачів, що були обрані для складання тесту. З іншого боку, варто відзначити прозорість та ясність висновків при використанні класичної теорії. З метою отримання результатів високого рівня істинності, вибірка тестованих не має бути великою – обсяг може варіюватися до сотні або менше здобувачів [8].

У разі застосування методів сучасної теорії тестування підсумковий бал вираховується як результат сукупної взаємодії прихованих параметрів, саме істинного рівня підготовки здобувачів і складності розглянутих завдань [14]. Сучасна теорія тестів полягає в теорії латентно-структурного аналізу, застосовуючи який можна визначити латентну характеристику, тобто оцінити рівень здібностей здобувача на основі розроблених тестових завдань, різних за рівнем складності, що визначається емпіричним шляхом під час тестування.

Прихований параметр, що оцінюється, повинен бути одномірним (принцип дії: один тест – один фактор – тест на інтелект). Існує два класи латентних параметрів [10], від яких залежить вірогідність правильної відповіді на тестове завдання, – рівень підготовки тестованого та рівень складності пропонованого завдання.

Відповіді індивіда не залежать одна від одної. Імовірність того, що при проходженні тесту екзаменований дасть конкретну послідовність відповідей, визначається як добуток ймовірностей відповідей на окремі завдання. Найбільш вагома відмінність сучасної теорії тестування від класичної полягає в наступному [7]: в IRT нехтують вирішенням проблеми емпіричної валідності та надійності тесту.

Тест є однофакторним, співвідноситься лише з однією властивістю та вважається априорі валідним. У класичній теорії тестів рівень результативності (або індивідуальний бал) є деяким постійним значенням. У сучасній теорії тестів прихований параметр розглядається як безперервна змінна.

Переважає більшість моделей IRT розрізняються за кількістю змінних, що використовуються в них. Найбільш відомі – це однопараметричні моделі Раша та Фергюсона, двопараметрична та трипараметрична моделі Бірнбаума.

Дихотомічна модель Раша є простим зв'язком між здобувачем, що проходить тестування, і завданнями [9]. Основним принципом, що застосовується при використанні даної моделі, є бінарна схема оцінювання правильності відповідей: 0 – відповідь є неправильною, 1 – відповідь правильна.

Зв'язок між рівнем здатності тестованого та складністю завдання виражається дихотомічною моделлю Раша:

$$\ln\left(\frac{P_1}{P_0}\right) = B - D, \quad (1)$$

де P_1 – ймовірність того, що тестований з рівнем знань B правильно відповість на завдання складністю D ; P_0 – ймовірність того, що відповідь була неправильною.

Одиницею виміру інтервальної шкали, побудованої за цією моделлю, є логіт – бінарний показник, що отримується в результаті взяття лінійної комбінації значень змінних величин, що можуть набувати будь-якого значення в діапазоні $\pm\infty$, і перетворення цих значень на шкалу ймовірності, тобто в діапазон від 0 до 1.

Відстань у логітах вздовж одномірної шкали вимірів між тестованим, який має 50% шансів відповісти правильно на тестове завдання ($\ln(50\% / 50\%) = 0$), тобто рівень здатності знаходиться в тій же точці на шкалі, що і рівень складності завдання, і тестованим, який має 75% шансів відповісти правильно ($\ln(75\% / 25\%) = 1,1$), дорівнюватиме 1,1 логіт.

Виразивши з рівняння ймовірність правильної відповіді, отримаємо:

$$P_1 = \frac{e^{B-D}}{1+e^{B-D}}. \quad (2.1)$$

З рівняння (2.1) можна помітити його схожість із логістичною функцією, графіком представлення якої є крива сімейства. Логістична функція широко застосовується також для визначення рівня ефективності процесу навчання та оцінки ефективності зворотного зв'язку із рівнем навчального навантаження та отриманим результатом знань та вмінь [6]:

$$f(x) = \frac{e^x}{1+e^x} \quad (2.2)$$

Підставивши в рівність (2.1) коефіцієнт 1,7, що використовується для сумісності моделі Раша та моделі Фергюсона [8], де ймовірність правильної відповіді виражена інтегралом нормального розподілу, отримаємо:

$$P_{ni1} = \frac{e^{1,7 \times (B_n - D_i)}}{1 + e^{1,7 \times (B_n - D_i)}}, \quad (3)$$

де індекс n відповідає ідентифікаційному номеру тестованого; i – ідентифікаційний номер завдання.

Дана залежність (3) визначає можливість правильної відповіді як функцію одного параметра, ще її називають однопараметричною моделлю IRT.

Коректне використання моделі Раша дозволяє відокремити оцінки рівнів підготовки випробуваних від оцінки рівнів складності завдань.

Однак однопараметрична модель Раша не може бути застосована в тому випадку, якщо тест містить завдання з різною диференційною здатністю або

дискримінативністю (здатність тестів виявляти сильних і слабких здобувачів, диференціювати піддослідних з їхньої підготовки). Вирішення цієї проблеми запропонував А. Бірнбаум, який запровадив ще один параметр – a :

$$P_{ni} = \frac{e^{1,7 \times a_i \times (B_n - D_i)}}{1 + e^{1,7 \times a_i \times (B_n - D_i)}} \quad (4)$$

Параметр a_i визначає диференціюючу здатність i -го завдання. Для ще більш точної відповідності емпіричним даним Бірнбаум ввів третій параметр – коефіцієнт вгадування c_i :

$$P_{ni1} = c_i + (1 - c_i) \frac{e^{1,7 \times a_i \times (B_n - D_i)}}{1 + e^{1,7 \times a_i \times (B_n - D_i)}} \quad (5)$$

Також існує ще чотирьохпараметрична модель, яка включає додатково параметр «неуважності», – d_i :

$$P_{ni1} = c_i + (1 - c_i - d_i) \frac{e^{1,7 \times a_i \times (B_n - D_i)}}{1 + e^{1,7 \times a_i \times (B_n - D_i)}} \quad (6)$$

Проте розрахунок всіх параметрів чотирьохпараметричної моделі є складним завданням. Якщо моделі Раша отримано послідовний алгоритм знаходження параметрів моделі, то інших моделей знаходження параметрів використовують чисельні методи з урахуванням методу максимальної правдоподібності, що з великому обсязі даних вимагає значних обчислювальних ресурсів [1].

На рис. 1 показана характеристична крива моделі тестової системи оцінювання якості рівня знань, що базується на чотирьох параметрах. Горизонтальна вісь показує приблизний рівень вмінь та навичок здобувача, що проходить тестовий контроль, на вертикальний відображається ймовірність того, що надана відповідь буде правильною.

Представлена функція є монотонно зростаючою з однією точкою перегину (точка a), в якій рівень складності завдання, що пропонується для вирішення здобувачу, відповідає рівню його знань. Кут нахилу прямої b відображає різкість зміни рівня складності завдань.

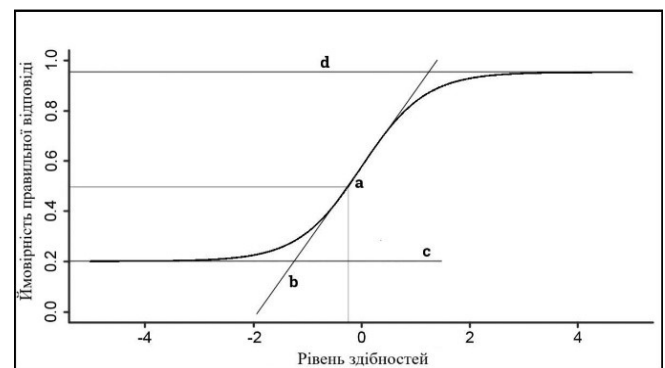


Рисунок 1 – Характеристична крива чотирьохпараметричної тестової моделі
Джерело: [2]

Вплив кожного з розглядуваних параметрів на графіку показаний за допомогою критичних точок та прямих:

- а – точка, у якій ймовірність правильної відповіді завдання, характеристична крива якого зображено на рисунку, дорівнює 0,5 (50%), тобто складність завдання дорівнює рівню підготовки здобувача, що складає тест;
- b – асимптота, кут нахилу якої відносно осі Ох відображає диференціовальну здатність представленого питання;
- с – асимптота, що характеризує коефіцієнт вгадування правильної відповіді на тестове завдання;
- d – асимптота, що характеризує коефіцієнт неуважності тестованого.

Інший клас моделей застосовується до політомічних завдань, де кожна відповідь відбиває часткову правильність виконання завдання. Їх застосовують у випадку, коли відповідь на запитання в тесті не може бути визначена як «правильно» чи «неправильно». У цьому випадку кожен варіант відповіді на тестове завдання має свою певну вагу, яка характеризує ступінь близькості відповіді до правильної відповіді.

Як правило, тестові завдання з множинними відповідями, завдання на відповідність з відкритим числовим рішенням є політомічними завданнями. Загальним прикладом оцінювання для завдань такого типу цього є шкала Лікерта [12], за якої, в залежності від повноти правильності відповіді, дається певна кількість балів – наприклад, у діапазоні «від 0 до 4».

При використанні комп'ютерного адаптивного тестування необхідним фактором, що варто враховувати при його застосуванні є специфіка автоматизованого процесу оцінювання рівня підготовки здобувачів [5].

На першому етапі тестування необхідно підготувати достатній банк тестових завдань, проранжованих за рівнем складності. Найбільш поширеним варіантом є банк дихотомічних завдань із кількома варіантами відповіді. Кожне завдання має складність, яка заздалегідь розрахована згідно з використовуваними моделями тестування.

Для використання певного типу завдань у процесі тестування, необхідним є виконання наступних умов:

- тест повинен мати властивість одномірності, тобто за одне тестування можуть бути визначені латентні параметри лише в одній галузі знань;
- банк тестових завдань має бути відкалібрований за рівнем складності на основі моделей IRT;
- тестований ознайомлений зі всіма умовами щодо вибору тестових завдань, умов зупинки проходження тестування та рівня точності оцінювання.

Безпосередньо адаптивне тестування включає наступні послідовні етапи, що є складовими цілісного ітераційного процесу:

- оцінюється рівень здібностей тестованого на основі доступної на даний момент інформації (попередніх відповідей на завдання). Визначається попередній рівень знань учасника тестування;
- обирається наступне для виконання завдання з бази тестових завдань, що ще не було відібрано для випробуваного відповідно до методу відбору та параметрів завдання (важкості, дискримінативності тощо);
- проводиться адміністрування вибору тестового завдання випробуваного, тобто оновлюється шаблон відповіді (фіксується, правильно чи ні учасник тестування відповів);
- кроки 1 – 3 повторюються.

Найбільш важливим кроком даного алгоритму є визначення методів оцінювання рівня здібностей учасника тестування. Методи оцінювання базуються на різних моделях та критеріях оцінювання [3]. Серед методів можна виділити метод максимальної правдоподібності (ML estimator), коли шаблон відповідей передбачає наявність лише двох подій – правильної та неправильної відповіді на тестове завдання (бінарне оцінювання правильності відповіді), і метод, що базується на припущенні нормального розподілу відповідей (BM estimator) для інших типів завдань. У процесі вибору наступного завдання також можуть використовуватись інші евристичні алгоритми, що базуються на теорії байєсівських мереж, нечіткої логіки та інших підходах.

Існують різні критерії вибору наступного тестового завдання, кожен з яких володіє рядом переваг та має ситуативне застосування, залежно від віку здобувача, що проходить тестування, рівня складності завдань та наукової дисципліни [15]:

- критерій максимальної інформативності (maximum information criterion-MFI);
- мінімальної очікуваної подальшої дисперсії (minimum expected posterior variance MEPRV);
- критерій максимальної правдоподібності зваженої інформації (maximum likelihood weighted information – ML, maximum posterior weighted information – MPWI);
- максимально очікуваної інформації (maximum expected information – MEI);
- комбінованого випадкового вибору.

Етап завершення адаптивного тестування передбачає використання одного з трьох правил:

- досягнуто максимальної довжини тесту (the length criterion);
- досягнуто необхідної величини точності оцінювання (precision criterion) – випробуваний стійко відповідає на завдання даного рівня складності;

– учасник тестування пройшов класифікацію рівня знань (classification criterion).

Максимальну довжину тесту задає адміністратор. Базуючись на типі тесту та цілях проходження тестування, вимог до точності оцінювання, визначається оптимальна кількість тестових завдань для виконання. Занадто мала кількість не дасть достовірної оцінки рівня здібностей випробуваного, тоді як результат занадто великого тесту може бути спотворений помилками по неухважності, оскільки учасники тестування можуть втомитися під час виконання надто великої кількості тестових завдань.

Необхідна величина точності визначається з урахуванням стандартної помилки оцінки рівня здібностей. Порогове значення також задається адміністратором тесту залежно від виду тесту та цілей тестування. Критерій класифікації використовується тоді, коли метою тестування є класифікація учасників, а чи не оцінювання їх здібностей, тобто випробуваний досяг деякого порогового значення здібностей і не потрібно продовжувати тестування.

На завершальному етапі тестування проводиться підсумкове оцінювання рівня здібностей випробуваного з урахуванням отриманого шаблону відповідей адаптованого тесту. Розраховується стандартна помилка та результати оцінювання можуть відображатися у вибраній шкалі оцінювання.

Двопараметрична модель Раша має недолік, який утруднює аналіз тесту: характеристичні криві, отримані в результаті проведення тестування в різних за рівнем здібностей групах здобувачів можуть сильно відрізнятися за кінцевими результатами. Одне і те саме завдання для однієї групи опитаних може виявитися легшим за інше, а для іншої групи – навпаки, що не завжди має обґрунтоване пояснення.

Приблизно такі за характером складнощі спостерігаються при тестуванні та розробці тестових завдань на основі багатопараметричних моделей Фергюсона та Бірнбаума. Визначення актуального значення коефіцієнтів, що використовуються при аналізі та розробці тестів на основі цих моделей, є складним процесом через надмірну кількість даних, що впливають на поведінку моделі оцінювання загалом.

Суттєвим недоліком розглянутих математичних моделей є також неврахування міжфакторної взаємодії та можливої взаємної залежності між ключовими показниками окремо взятих моделей. Кожна зі змінних, що входить до характеристичного рівняння, описує дискретний вплив окремо взятого фактора, без врахування дії їх сумарного ефекту.

З метою усунення вищенаведених недоліків, характерних даним моделям, під час проведення попереднього тестування здійснюється калібрування завдань за рівнем складності та створення математичної моделі, в якій комплексно будуть враховані характеристики, що можуть вплинути на оцінювальну якість тесту. Реалізація такого підходу відбувається при фіксації значень всіх, за виключенням одного, з параметричних показників, вплив якого підлягає аналізу; такий алгоритм застосовується послідовно до всіх параметрів, що відображають механізм дії моделі.

Під час попереднього тестування для визначення рівня складності завдання, що пропонується для виконання наступним за рахунком (n), необхідним врахувати результати, що були продемонстровані тестованим при виконанні попередніх k завдань. Рівень складності початкових k завдань (базових для оцінки) ґрунтується на статистичних даних рівня складності аналогічних завдань тестувань попередніх років.

Кожному з них присвоюється певний ваговий коефіцієнт a_i , що відображає «внесок» рівня складності цього завдання в поточне:

$$d_n = \sum_{i=1}^k a_i d_i, \quad (7)$$

де a_i – ваговий коефіцієнт, що відповідає «внеску» в рівень складності поточного завдання кожного із попередніх; d_n – складність завдання з номером n (у логітах); k – кількість попередніх завдань, результати вирішення яких враховуються при визначенні рівня складності поточного для виконання завдання.

При визначенні вагових коефіцієнтів впливу складності попередніх k завдань на рівень складності поточного завдання n , необхідною умовою є:

$$\sum_{i=1}^k a_i = 1. \quad (8)$$

Відхилення суми вагових коефіцієнтів у бік значень, сумарна величина яких є більшою або меншою за одиницю, призведе до розбіжності кінцевого значення ряду або його збіжності до нескінченно малої величини (рис. 2):

– при $\sum_{i=1}^k a_i < 1$ кінцевий рівень складності становитиме: $\lim \sum_{i=1}^k a_i d_i \rightarrow \frac{1}{\infty}$;

– при $\sum_{i=1}^k a_i > 1$ відповідно: $\lim \sum_{i=1}^k a_i d_i \rightarrow \infty$.

Важливим є той факт, що збіжність чи розбіжність числового ряду не залежить від початкового розподілу кількості балів, отриманих за базисні завдання. Встановлення рівноважного значення складності завдання в логітах, що відповідає рівню знань тестованого, досягається при строгому дотриманні умови (8).

На основі проведеного попереднього тестування відбувається розрахунок параметрів,

значення яких будуть включені як факторні характеристики до кінцевої математичної моделі, що буде використовуватися при проведенні тестування.

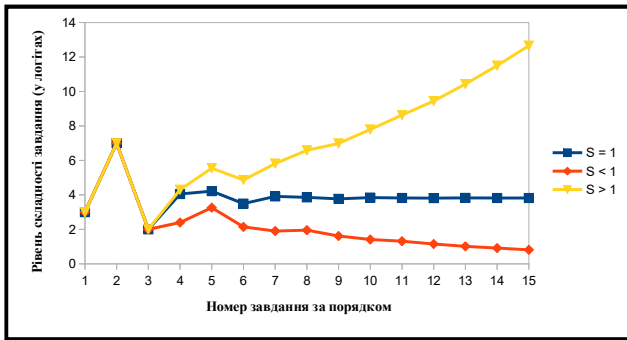


Рисунок 2 – Динаміка зміни рівня складності завдань тесту, залежно від значень суми вагових коефіцієнтів моделі
Розроблено автором на основі джерела: [7]

Оскільки рівень складності завдання вимірюється в логітах, то при переведенні отриманих після проходження тестового контролю результатів в бали, отриманим логітам необхідно присвоювати відповідний еквівалент у балах. У кількісному вираженні в більшості випадків один логіт не є еквівалентом одного балу. В залежності від рівня складності завдання, кількість балів, що відповідає одному логіту змінюється: за правильно вирішене завдання вищого рівня складності нараховується більша кількість балів за один отриманий логіт, ніж при вирішенні завдання середнього чи легкого рівня складності.

Додатковою перевагою використання такого підходу при калібруванні завдань за складністю та присвоєним ним логітним значенням, є те, що при переході від завдань одного блоку (рівень складності якого нижчий) до наступного, в обрану вибірку з k елементів потраплятимуть завдання з різним рівнем складності, що дозволить здійснювати перехід без різкої зміни навантаження на тестованого.

Завдяки використанню покрокового ітераційного алгоритму зміни рівня складності завдання, нівелюються латентні параметри та частково усувається вплив фактора неувважності на кінцевий результат складання тесту. Отриманий кінцевий показник відображатиме наближений до реального рівень знань тестованого, вплив на який випадкових факторів значно зменшений.

Застосування параметричних моделей в тестових завданнях дозволяє побудувати ефективні методи обробки та аналізу даних, отриманих в результаті проведення тестового контролю рівня знань та навичок здобувачів. Головними наслідками використання моделей параметричного типу є:

- підвищення рівня диференційованості завдань за рівнем складності, що дозволяє

розташовувати їх в тестових блоках в логічному порядку зростання кількості та складності операцій, необхідних для правильного вирішення;

- дані, отримані при складанні тесту, структура якого відповідає моделям параметричного типу, дозволяють здійснити більш ефективний розподіл здобувачів, що складають тести, за рівнем знань та вмінням їх застосовувати при вирішенні представлених та аналогічних ним завдань;

- багатопараметричність математичних моделей, що використовуються при аналізі результатів тестового контролю, дозволяє оцінити не тільки кількісні параметри рівня знань, тобто накопичену їх кількість, проте і вміння їх застосовувати за умови комбінування між собою при вирішенні завдань підвищеного рівня складності, представлених у завершальних блоках тесту;

- використання параметричних моделей дозволяє оцінити рівень знань здобувачів незалежно від рівня складності завдання, шляхом виявлення зв'язку між латентними параметрами тестованого та результатами виконання тесту.

Висновки

Імплементація нових методів контролю та оцінки навчальних досягнень є одним з основних показників покращення якості сучасної системи вищої освіти. На сьогоднішній момент використання тестів дає можливість реалізувати всі види контролю знань та відповідає основним його принципам.

Широке впровадження у систему освіти комп'ютерного оцінювання знань, також використання дистанційного навчання, призвело до необхідності розробки якісного та швидкого моніторингу результатів підготовки здобувачів. З одного боку, отримані дані повинні бути легко інтерпретованими, з іншого – підходи, що застосовуються, повинні відповідати всім необхідним критеріям якості: завдання повинні бути належним чином відкалібровані, тести – валідними та надійними, складність завдання – максимально відповідати рівню підготовки респондента тестування.

Дане наукове дослідження присвячене розгляду математичних моделей сучасної теорії тестування, оскільки наукові підходи, що використовують такі методи вважаються більш потужними, ніж підходи класичної теорії тестування. В роботі описані основні характеристики даних моделей, важливою перевагою яких є те, що оцінки можуть бути використані для адаптації процесу тестування з точки здібностей та вмінь здобувачів. Такий підхід використовується при використанні комп'ютерного адаптивного тестування.

У роботі наведено огляд загального алгоритму роботи та специфіку використання адаптивного тестування, розглянуто математичний апарат, що

застосовується безпосередньо для формування рівня складності пропонованих для вирішення завдань та для інтерпретації результатів перевірки знань здобувачів різного рівня підготовки.

Розглянуто недоліки математичних моделей, що використовуються при побудові та аналізі ефективності завдань тестової форми, запропоновано альтернативний спосіб спрощення їх застосування з метою підвищення ефективності роботи при визначенні рівня знань тестованих.

При розробці тестових завдань необхідно враховувати також поєднання між собою логічних блоків системи тестового контролю рівня знань. Тести необхідно гармонійно поєднувати з іншими формами та методами контролю, що сприяє як об'єктивному оцінюванню успішності, уможлиблює отримання правильних висновків про рівень їх навченості, так і забезпечує формування важливих професійних якостей у майбутньому.

Список літератури

1. Карташов М. В. Імовірність, процеси, статистика. Київ : ВПЦ «Київський університет», 2007. 504 с.
2. Оксенюк І. І. Організація діагностики та контролю знань засобами комп'ютерного тестування. *Педагогічний пошук*. 2022. № 1 (113). С. 31–35.
3. Сачанюк-Кавецька Н. В., Прозор О. П. Особливості тестового контролю знань студентів вищих навчальних закладів. *Dynamics of the development of world science : the 8th International scientific and practical conference (April 15-17, 2020)*. Vancouver, Canada : Perfect Publishing, 2020. P. 796–803.
4. Шевчук С. С. Вектори підвищення якості професійної підготовки фахівців у сучасних умовах. *Імідж сучасного педагога : електрон. наук. фах. журнал*. 2020. № 3 (192). С. 5–9.
5. Bae C. L., Theriault D. J., Redifer J. L. Investigating the testing effect: Retrieval as a characteristic of effective study strategies. *Learning and Instruction*. 2019. Vol. 60 (1). P. 206–214.
6. Han J., Morag C. The influence of the sigmoid function parameters on the speed of backpropagation learning. *From Natural to Artificial Neural Computation : Lecture Notes in Computer Science*. 1995. Vol. 930. P. 195–201.
7. Linden W., Hambleton R. Handbook of modern item response theory. New York : Springer, 2018. 512 p.
8. Loken E., Rulison K. L. Estimation of Four-Parameter Item Response Theory Model. *British Journal of Mathematical and Statistical Psychology*. 2010. № 63. P. 509–525.
9. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago : The University of Chicago Press, 1980. 199 p. (бажано вказати кількість сторінок).
10. Riggs C. D., Kang S., Rennie O. Positive Impact of Multiple-Choice Question Authoring and Regular Quiz Participation on Student Learning. *CBE Life Sci Educ*. 2020. № 19 (2).
11. Ronald K. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: issues and practice*. 1993. Vol. 12, Iss. 3. P. 21–28. (додано том/випуск для повноти).
12. Rusch T., Lowry P., Mair P., Treiblmaier H. Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management*. 2017. Vol. 54 (2). P. 189–203.
13. Wang W. Are All Item Response Functions Monotonically Increasing : Dissertation Ph.D. Kansas, 2012. 166 p.
14. Wiberg M. Classical test theory vs. item response theory. *Emergency Medicine*. 2004. № 50. URL: http://www.edusci.umu.se/digitalAssets/59/59529_em-no-50.pdf (дата звернення: 14.01.2026).
15. Ying C. Computerized adaptive testing: New developments and applications. University of Illinois at Urbana-Champaign, 2008. 92 p.

Стаття надійшла до редколегії 02.11.2025

Vronskiy Serhii

PhD candidate at the Department of Systems Analysis and Optimization Theory,
<https://orcid.org/0000-0002-9960-6437>
Uzhhorod National University, Uzhhorod

Andrashko Yuri

Associate Professor of the Department of Systems Analysis and Optimization Theory,
<https://orcid.org/0000-0003-2306-8377>
Uzhhorod National University, Uzhhorod

MODELS AND METHODS OF ADAPTIVE TESTING AS A TOOL FOR INFORMATIZATION OF KNOWLEDGE CONTROL SYSTEMS IN HIGHER EDUCATION INSTITUTIONS

Abstract. The research is devoted to a review of modern models and methods used to analyze the effectiveness of the test form of knowledge control. A description of theoretical approaches to evaluating the results of testing is provided. Mathematical models of modern testing theory are considered, describing the relationship between candidates' skills and their practical reflection when performing test tasks. A description of the general principles of adaptive testing using dichotomous and polytomous tasks to control the knowledge and skills of applicants acquired in the learning process is provided. The test approach to assessing the

level of knowledge is analyzed, since its application involves an objective assessment of knowledge and taking into account various aspects of educational activity: the ability to recognize and understand the material, classify phenomena and processes, determine the sequence of logical operations, systematize and synthesize situations and processes, and subsequently to conduct independent research and apply the knowledge gained in practice. The article explores the specifics of the application of various approaches that describe the theoretical basis for modern testing theory functioning: one-parameter models of Rasch and Ferguson, two-parameter and three-parameter models of Birnbaum. The scientific study provides an overview of the work algorithm and the specifics of using adaptive testing tools used to determine the general level of knowledge in certain academic disciplines; the mathematical apparatus used directly for forming blocks of tasks of the appropriate complexity level and for interpreting the results of testing the knowledge of applicants of different levels of training is considered. Special attention is paid to the automatic task selection of the required level of complexity, which is directly based on the testing results in real time. Such an algorithm allows for identifying the knowledge level of the test-taker in a particular discipline, while at the same time assessing their psychometric characteristics: the level of concentration, the ability to withstand stressful situations, and the speed of switching between different task types and levels of complexity. Also important is the preliminary distribution of tasks into blocks according to the type of formulated condition for completing the task and the number of logical steps that must be taken to correctly solve the task. The profile of the test-taker's knowledge level obtained as a result of the complete testing allows for identifying not only gaps in knowledge in a particular subject, but also for assessing the level of psychomotor adaptation to the testing process during task performance.

Keywords: modern testing theory; adaptive testing; test quality; test form for controlling the level of knowledge; mathematical models for assessing test quality

References

1. Kartashov, M. V. (2007). *Probability, processes, statistics*. Kyiv: University Publishing and Printing Center. 504 p.
2. Okseniuk, I. I. (2022). Organization of Diagnostics and Control of Knowledge by Means of Computer Testing. *Pedagogical Search*, 1 (113), 31–35.
3. Sachanyuk-Kavetska, N. V., & Prozor, O. P. (2020). Features of test control of knowledge of students of higher educational institutions. In *Dynamics of the Development of World Science: The 8th International Scientific and Practical Conference (April 15-17, 2020)*. Vancouver, Canada: Perfect Publishing, 796–803.
4. Shevchuk, S. S. (2020). Vectors of the specialists' vocational training quality improvement in modern conditions. *Image of the Modern Pedagogue*, 3 (192), 5–9.
5. Bae, C. L., Theriault, D. J., & Redifer, J. L. (2019). Investigating the testing effect: Retrieval as a characteristic of effective study strategies. *Learning and Instruction*, 60 (1), 206–214.
6. Han, J., & Morag, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. *From Natural to Artificial Neural Computation: Lecture Notes in Computer Science*, 930, 195–201.
7. Linden, W., & Hambleton, R. (2018). *Handbook of modern item response theory*. New York: Springer. 512 p.
8. Loken, E., & Rulison, K. L. (2010). Estimation of Four-Parameter Item Response Theory Model. *British Journal of Mathematical and Statistical Psychology*, 63, 509–525.
9. Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
10. Riggs, C. D., Kang, S., & Rennie, O. (2020). Positive Impact of Multiple-Choice Question Authoring and Regular Quiz Participation on Student Learning. *CBE—Life Sciences Education*, 19 (2).
11. Ronald, K. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12 (3), 21–28.
12. Rusch, T., Lowry, P., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management*, 54 (2), 189–203.
13. Wang, W. (2012). *Are All Item Response Functions Monotonically Increasing?* (Doctoral dissertation). University of Kansas, 166 p.
14. Wiberg, M. (2004). Classical test theory vs. item response theory. *Emergency Medicine*, 50. Retrieved from: http://www.edusci.umu.se/digitalAssets/59/59529_em-no-50.pdf
15. Ying, C. (2008). *Computerized adaptive testing: New developments and applications*. University of Illinois at Urbana-Champaign. 92 p.

Посилання на публікацію

- | | |
|------|--|
| APA | Vronskyi, S. & Andrashko, Y. (2025). Models and methods of adaptive testing as a tool for informatization of knowledge control systems in higher education institutions. <i>Management of Development of Complex Systems</i> , 64, 267–275, dx.doi.org/10.32347/2412-9933.2025.64.267-275 . |
| ДСТУ | Вронський С. В., Андрашко Ю. В. Моделі та методи адаптивного тестування як інструмент інформатизації системи контролю знань у закладах вищої освіти. <i>Управління розвитком складних систем</i> . Київ, 2025. № 64. С. 267 – 275, dx.doi.org/10.32347/2412-9933.2025.64.267-275 . |