УДК 37.091.26:004.9

Круглова Наталія Володимирівна
кандидат фізико-математичних наук, доцентка кафедри математичного аналізу та теорії ймовірностей
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна
ORCID ID 0000-0002-9143-9398
natahak@ukr.net

Диховичний Олександр Олександрович
кандидат фізико-математичних наук, доцент, доцент кафедри математичного аналізу та теорії ймовірностей
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна
ORCID ID 0000-0001-8505-2878
a.dyx@ukr.net

ДОБІР МАТЕМАТИЧНОЇ МОДЕЛІ ДЛЯ АНАЛІЗУ ТЕСТОВИХ ЗАВДАНЬ ТИПУ «ВБУДОВАНІ ВІДПОВІДІ» З МАТЕМАТИЧНИХ ДИСЦИПЛІН

Анотація. У статті досліджується питання вибору математичної моделі для аналізу комп'ютерних тестів з математичних дисциплін. Особливістю даного дослідження є розгляд тестових завдань типу «вбудовані відповіді». Такі тестові завдання мають декілька пов'язаних за певною логікою «кроків» і були використані саме для дистанційного контролю знань з математичних дисциплін під час карантину в НТУУ «Київський політехнічний інститут імені Ігоря Сікорського». Введення таких завдань у тест ставить перед викладачем питання: як аналізувати якість таких тестів? Автори побудували дослідження на основі загальновідомих методів Класичної Теорії Тестів (КТТ) та Сучасної Теорії Тестів (IRT), які довели свою ефективність як у педагогічному тестуванні, так і в інших галузях, у яких знаходить застосовування тестовий підхід.
Дослідження проводилось на прикладі модульної тестової контрольної роботи з теорії ймовірностей для студентів другого курсу факультету інформатики та обчислювальної техніки НТУУ «Київський політехнічний інститут імені Ігоря Сікорського», сформованої на платформі MOODLE під час дистанційного навчання в листопаді 2020 р. Під час дослідження було: проведено аналіз якості тесту як у цілому, так і окремих тестових завдань; оцінено латентні параметри завдань, перевірено адекватність відповідних моделей даним тестування; побудовано характеристичні та інформаційні криві; за допомогою інформаційних критеріїв вибрано моделі IRT, які краще підходять для аналізу комп'ютерних тестів із завданнями типу «вбудовані відповіді»; доведено, що найбільш прийнятними для аналізу подібних завдань є як дихотомічна модель Раша, так і політомічна модель Муракі.
Обчислення проводилося за допомогою спеціальних бібліотек (пакетів) мови програмування R, у яких реалізовано алгоритми КТТ та IRT, а саме, **eRm, ltm, mirt**.
Показано, як за допомогою завдань з «вбудованими відповідями» можна проаналізувати певні вміння й знання студентів.
Результати дослідження сприяють підвищенню компетентності викладачів щодо оцінювання якості тестів з вищої математики, що є дуже актуальним при складанні тестів з різноманітних дисциплін, особливо під час дистанційного навчання студентів.

Ключові слова: комп'ютерне тестування; «вбудовані відповіді»; модель Раша; політомічна модель; середовище програмування R; пакети eRm і ltm; MOODLE.

1. ВСТУП

Постановка проблеми. Сучасний стан пандемії та тимчасового карантину висунули нові вимоги до якісного контролю знань студентів в умовах дистанційної освіти. Одним з найбільш поширених засобів дистанційного контролю знань

© Н.В. Круглова, О.О. Диховичний, 2022.

залишаються комп'ютерні тести, які отримали стрімке та широке розповсюдження особливо під час карантину. Розширення застосування комп'ютерних тестів висуває перед викладачами різних рівнів освіти ряд нових проблем. Серед них однією з найважливіших є проблема якості складених у короткі строки тестів. Окрім того, виникла очевидна проблема банального «списування» студентами, зокрема з використанням дуже поширених онлайн ресурсів з математики.

Під час карантину у КПІ ім. Ігоря Сікорського було створено широкий спектр дистанційних курсів, які було розміщено на платформі дистанційного навчання «Сікорський» [https://www.sikorsky-distance.org]. Основним інструментом для розміщення і розробки дистанційних курсів у КПІ було обрано платформу MOODLE (Modular Object Oriented Dictance Learning Environment) [1], версії 3.10 і вище, яка добре зарекомендувала себе протягом останнього десятиліття.

У платформі MOODLE існує велика кількість форм сучасних тестових завдань. В останніх версіях користувачам надана можливість створювати завдання типу «вбудованої відповіді» (ВВ) [2, 3]. Ці завдання в комплексі з більш простими завданнями типу «числовий», «відповідність», «багатоваріантне питання», «розрахунковий» складають основу сучасних тестів з математичних дисциплін у КПІ ім. Ігоря Сікорського.

Підвищена увага до тестових завдань типу ВВ зумовлена наступними факторами:

1. В умовах дистанційного тестового контролю знань викладачі стикаються зі звичайним списуванням або використанням відомих онлайн ресурсів для розв'язування математичних задач. Наявність у тесті завдань типу ВВ понижує можливості студентів по залученню таких ресурсів, оскільки є можливість перевірити проміжні етапи розв'язання задачі, які не завжди прописані на онлайн ресурсах для розв'язання математичних задач;
2. За допомогою одного тестового завдання можна водночас перевірити декілька знань або вмінь студента, наприклад, у завданні на рис.2 – знання розподілу системи випадкових величин, уміння обчислювати математичне сподівання, уміння перевіряти незалежність дискретних випадкових величин;
3. Можна відслідкувати, як визначається певне вміння студента в різних завданнях, наприклад, у представленій контрольній роботі – уміння інтегрувати;
4. Оскільки завдання типу ВВ містять декілька простих математичних кроків, пов'язаних між собою певними логічними зв'язками (див. приклад на рис.2), то вони дозволяють не тільки перевіряти окремі навички, а й логіку їх цілісного використання;
5. Достатньо ефективно реалізуються за допомогою платформи MOODLE;
6. Краще розрізняють навички і вміння студентів.

Хоча очевидно, що система MOODLE не є ідеальною, зокрема в ній дещо обмежені засоби побудови і аналізу складних логічних конструкцій, наприклад, як у системі[4], але наявність шаблону завдань типу ВВ суттєво компенсує це обмеження. До того ж, функція перегляду спроб студента дає можливість відслідкувати відповіді на окремі кроки завдання.

Важливим є те, що складання таких завдань суттєво підвищує компетентність викладача щодо оцінювання якості тестів з вищої математики [5] саме за допомогою таких факторів:

- уміння добирати окремі кроки;
- об'єднувати їх єдиною логічною схемою;
- правильно використовувати можливості гнучкої «розбаловки» для збалансованої підсумкової відповіді;

– користуватись сучасними комп'ютерними засобами для формування завдань, зокрема створювати тестові завдання мовою HTML.

Методику аналізу якості дихотомічних, політомічних завдань, завдань «множинного вибору» у комп'ютерних тестах з вищої математики досліджено зокрема у роботах [6], [7]. Основу цих досліджень складають Класична Теорія Тестів (КТТ) та Сучасна Теорія Тестів (IRT). Саме аналіз тестових завдань типу ВВ у цій методиці, а також у інших джерелах не розглядався. Тому виникає проблема вибору підходу для аналізу завдань з ВВ.

З огляду на це автори визначили такі завдання дослідження на прикладі тестової контрольної роботи з теорії ймовірностей:

– провести аналіз як тесту в цілому, так і окремих тестових завдань;
– оцінити латентні параметри завдань;
– перевірити адекватність відповідних моделей даним тестування;
– побудувати графіки та проаналізувати характеристичні та інформаційні криві;
– на основі проведеного аналізу і використання інформаційних критеріїв дослідити та обґрунтувати, за допомогою яких моделей IRT краще проводити аналіз комп'ютерних тестів із завданнями типу «вбудовані відповіді».

Аналіз останніх досліджень і публікацій. Питанню застосування комп'ютерного тестування для контролю знань присвячено багато досліджень. Серед них слід відзначити монографію В. Ю. Бикова, Ю. М. Богачкова [8], у якій автори надають основну інформацію, необхідну для організації і проведення якісного тестування на основі ІКТ. Принципи проєктування і побудови ефективної автоматизованої системи педагогічної діагностики наведено в роботі О. Г. Колгатіна [9], також у роботі того ж автора [10] представлено автоматизовану систему діагностики знань з математичної статистики з адаптивними тестами. У роботі Ю.М. Богачкова та П.С. Уханя [11] досліджуються проблеми використання довідників у проєктуванні дистанційних курсів, а саме банків тестових завдань.

Розширення та розвинення комп'ютерного тестування потребує всебічного аналізу якості тестів, які використовуються. Аналіз якості тестів вибудовано у двох напрямах: перший напрям - опрацювання результатів тестування методами КТТ, тут відзначимо роботи Дж. Алгіна (J. Algina), Л. Крокера (L. Crocker) [12], Н. Б. Челишкової [13] та ін. На думку цих авторів аналіз якості тесту на засадах КТТ зокрема передбачає оцінювання надійності, валідності (змістовної, критеріальної та прогностичної) та аналіз розподілу вибірки індивідуальних балів. Найбільш поширеною числовою характеристикою надійності тесту є коефіцієнт Кронбаха, а аналіз розподілу вибірки результатів полягає в графічному представленні даних у вигляді гістограми та обчисленні вибіркових статистичних характеристик.

Другий напрям – аналіз тестів методами IRT, який розпочато ще в роботі Г.Раша [14] і який продовжує дуже інтенсивно розвиватись. Основи аналізу тестів методами IRT висвітлено в монографіях М. Уілсона (M. Wilson) [15] та Ф. Бейкера (F. Baker) [16]. Найбільш повний огляд сучасних підходів та застосування моделей IRT до аналізу тестів як педагогічних, так і психологічних, соціологічних та ін. наведено у відомій монографії В. Ліндена (W. Linden) та Р.К. Хемблетона (R. Hambleton) [17]. Відзначимо також роботи В. С. Аванесова [18], Т. В. Лісової [19], у яких в основному донесено базові ідеї IRT до вітчизняного науковця. Сумісне застосування методів КТТ та IRT при опрацюванні результатів тестування з вищої математики, об'єднаних в єдину методику оцінювання якості тестів з вищої математики, досліджувалось у роботі [6]. Оскільки нас цікавив аналіз завдань типу ВВ, які реалізовані на платформі MOODLE,

то в дослідженні ми обирали найбільш перевірені базові моделі IRT, такі як моделі: Раша; 1,2,3-PL моделі, описані в [17], а також модель Муракі (Muraki) [20]. Зауважимо, що для дихотомічних тестових завдань типу «числове» недоречно вводити параметр угадування, оскільки це призводить до надмірної складності моделі. Своєю чергою, на думку Крокера (L. Crocker) [12], це може призводити до неточності в обчисленні оцінок для параметрів тесту і до хибного трактування результатів тестування. Оскільки більшість завдань типу ВВ, на прикладі яких проводився аналіз, складалися з завдань-кроків числового типу, то ми керувалися [12] при виборі математичної моделі для аналізу тестових завдань такого типу.

Очевидно, що в умовах карантину і дефіциту часу для якісного статистичного аналізу тестів неможливо обійтись без алгоритмів та методів IRT та їх відповідної комп'ютерної реалізації. У середовищі програмування R [21] існує цілий розділ, який містить бібліотеки (пакети) з широким розмаїттям функцій з аналізу тестів. Приміром, це пакети: **eRm, ltm, mirt**. Перевагою обраного середовища програмування для програмної реалізації дослідження є безкоштовність, висока швидкість обробки великих об'ємів даних та гарантованість правильної роботи алгоритмів. Дослідження в галузі IRT із застосуванням системи програмування R проведено в монографіях [22] та [23] Але, незважаючи на дуже велику кількість робіт з аналізу тестів, питання статистичного аналізу тестових завдань типу ВВ за допомогою методів IRT окремо не розглядалось.

У роботі Дудко А.Ф. [5] визначено компетентність викладача щодо оцінювання якості тестів з вищої математики.

У відповідності до завдань дослідження сформулюємо його мету.

Мета дослідження: обґрунтувати вибір математичних моделей для аналізу тестів з математичних дисциплін із завданнями типу ВВ та проаналізувати результати проведеного тестування на підставі обраних моделей з метою забезпечення підвищення рівня компетентності викладачів математичних дисциплін щодо оцінювання якості тестів.

2. МЕТОДИКА ДОСЛІДЖЕННЯ

Дослідження проводилось на підставі аналізу результатів комп'ютерного тестування з різних математичних дисциплін, але в основу даної статі, як типовий приклад, покладено результати контрольної роботи з теорії ймовірностей.

У межах КТТ у дослідженні для аналізу тесту обрані такі характеристики: валідність, надійність, розподіл вибірки результатів.

Під валідністю тесту будемо розуміти характеристику тесту, яка відображає відповідність результатів тестування тій меті, заради якої тест було створено. Розрізняють змістовну, критеріальну та прогностичну валідності.

Змістова валідність – характеристика відповідності змісту тесту навчальному плану та програмі, змісту підручників, а також формулюванню мети навчання [13].

Критеріальна валідність передбачає наявність зовнішнього критерію, кореляція з яким визначає валідність тесту [24]. У якості критерію беруться оцінки експертів, виставлені при традиційній перевірці знань студентів без використання тестів. Отже, для оцінювання критеріальної валідності обчислюється коефіцієнт кореляції між тестовими індивідуальними балами студентів і їх балами за традиційну письмову контрольну роботу. Такі контрольні роботи також проводились і під час дистанційного навчання. А викладачі кафедри математичного аналізу та теорії ймовірності з багаторічним величезним досвідом викладання математики, безумовно, є експертами в цій царині.

Прогностична валідність – це характеристика тесту, що відображає ступінь упевненості, що отримані за тест оцінки адекватно прогнозують майбутні досягнення іспитника [25].

Надійність тесту визначалась за коефіцієнтом надійності Кронбаха α [12]:

$$\alpha = \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} S_j^2(X)}{S^2(X)}\right),$$

де $S_j^2(X)$ – вибіркова дисперсія розподілу балів за j-е завдання; $S^2(X)$ – вибіркова дисперсія розподілу індивідуальних балів.

Основна увага в дослідженні була приділена вибору такої математичної моделі IRT, яка б найточніше відповідала даним. Аналіз тестових завдань за допомогою моделей IRT побудовано шляхом введення латентних параметрів іспитників (студентів) і тестових завдань. Нижче наведено обрані моделі, у яких ми будемо використовувати такі позначення:

- n – кількість іспитників,
- m – кількість завдань у тесті.

Було відібрано такі моделі:

Модель Раша(G. Rasch) [14]. Застосовується для завдань дихотомічного типу (завдань, у яких тільки два варіанта відповіді: правильний або неправильний). У [13] зазначено, що цю модель можна використовувати в завданнях відкритого типу.

Модель описується наступними параметрами:

- Складністю завдання тесту - $\alpha_j,\ j = \overline{1,m}$;
- Підготовленістю іспитника - $z_i, i = \overline{1,n}$.

Імовірність правильної відповіді i-го іспитника на j-те завдання тесту визначається за формулою:

$$P_{ij} = \frac{1}{1+e^{\alpha_j - z_i}}, i = \overline{1,n},\ j = \overline{1,m}.$$

Модель Раша виправлена (1-PL) [17]. Відрізняється від попередньої моделі тільки множником перед параметрами моделі. Імовірність правильної відповіді i-го іспитника на j-те завдання тесту визначається за формулою:

$$P_{ij} = \frac{1}{1+e^{1.7(\alpha_j - z_i)}}, i = \overline{1,n},\ j = \overline{1,m}.$$

Модель Бірнбаума (2-PL) (A. Birnbaum) [17]. Ця модель є узагальненням моделі Раша, до якої додатково вводиться параметр диференціюючої здатності. Параметрами моделі є:

- $D_j,\ j = \overline{1,m}$ - диференціююча здатність завдання;
- $z_i, i = \overline{1,n}$ - підготовленість і-того іспитника;
- $\alpha_j,\ j = \overline{1,m}$ - складність завдання тесту.

Імовірність правильної відповіді i-го іспитника на j-те завдання тесту визначається так:

$$P_{ij} = \frac{1}{1+e^{-D_j(z_i - \alpha_j)}}, i = \overline{1,n},\ j = \overline{1,m}.$$

3-PL модель [17]. Ця модель теж використовується для завдань дихотомічного типу. Але в ній вводиться додатковий параметр – імовірність угадування правильної відповіді. Модель характеризується наступними латентними параметрами:

- D_j, $j=\overline{1,m}$ - диференціююча здатність завдання;
- $z_i, i=\overline{1,n}$ - підготовленість і-того іспитника;
- α_j, $j=\overline{1,m}$ - складність завдання тесту;
- γ_j, $j=\overline{1,m}$ - імовірність угадування правильної відповіді j-го завдання.

$$P_{ij}=\gamma_j+(1-\gamma_j)\left[\frac{1}{1+e^{-D_j(z_i-\alpha_j)}}\right], i=\overline{1,n},\ j=\overline{1,m}.$$

Модель Муракі (The Generalized Partial Credit Model) (Muraki) [20]. Ця модель використовується для політомічних питань. ЇЇ параметрами є:

- D – константа, рівна 1.7;
- a_j - диференціююча здатність завдання тесту;
- α_{jk}, $j=\overline{1,m}, k=\overline{0,m_j}$ - параметри складності k-го рівня j-го завдання;
- $z_i, i=\overline{1,n}$ - підготовленість і-того іспитника;
- m_j - кількість рівнів завдання .

У політомічній моделі ймовірність досягнення і-м іспитником k-го рівня j-ого завдання визначається за формулою:

$$P_{ijk}=\frac{e^{\sum_{s=0}^{k}Da_j(z_i-\alpha_{js})}}{\sum_{k=0}^{m_i}e^{\sum_{s=0}^{k}Da_j(z_i-\alpha_{js})}},\ i=\overline{1,n},\ j=\overline{1,m},k=\overline{1,m_j}.$$

Перевірка адекватності моделей даним тестування здійснювалась за критерієм χ^2 [26].

Вибір моделі для аналізу завдань типу ВВ побудовано на застосуванні спектру спеціальних інформаційних критеріїв AIC, BIC тощо, і найкраща модель забезпечує мінімальне значення критеріїв. Детальний опис можна знайти в [26].

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

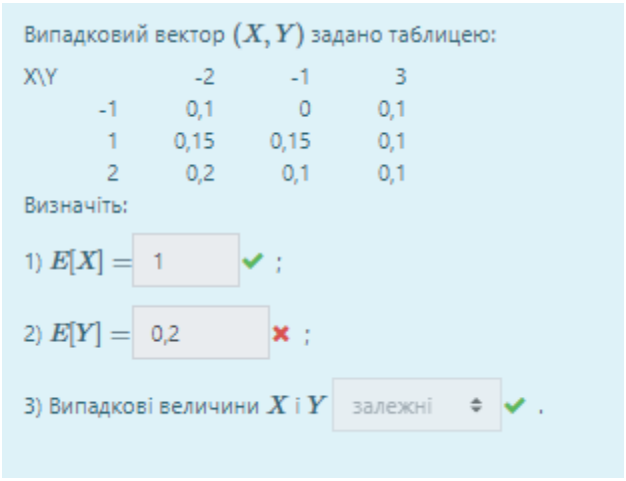
Дослідження проводилось за результатами модульної контрольної роботи з теорії ймовірностей на тему «Випадкові величини та їх характеристики», яку було запропоновано студентам факультету інформатики та обчислювальної техніки другого курсу КПІ ім. Ігоря Сікорського, спеціальність 121. Контрольна складалась із 8 завдань типу ВВ і двох завдань типу «множинний вибір» зі штрафними балами у разі вибору неправильного варіанту.

Ми проведемо аналіз саме завдань типу ВВ. Як вже зазначалось, завдання типу ВВ об’єднують у єдине завдання декілька більш коротких кроків. Відповідями на кроки можуть бути:

- коротка відповідь;
- числова відповідь з похибкою;

– множинний вибір у різних формах.

Водночас допускається гнучкість при «розбаловці» результатів окремих кроків. Усе це дуже розширює можливості викладача при складанні тестів. Для пояснення наведемо типові приклади таких завдань з даної контрольної. У завданні на рис. 1 результат знаходження відповіді на кожен крок задачі не залежить від попереднього кроку.


Випадковий вектор (X, Y) задано таблицею:

X\Y	-2	-1	3
-1	0,1	0	0,1
1	0,15	0,15	0,1
2	0,2	0,1	0,1

Визначіть:

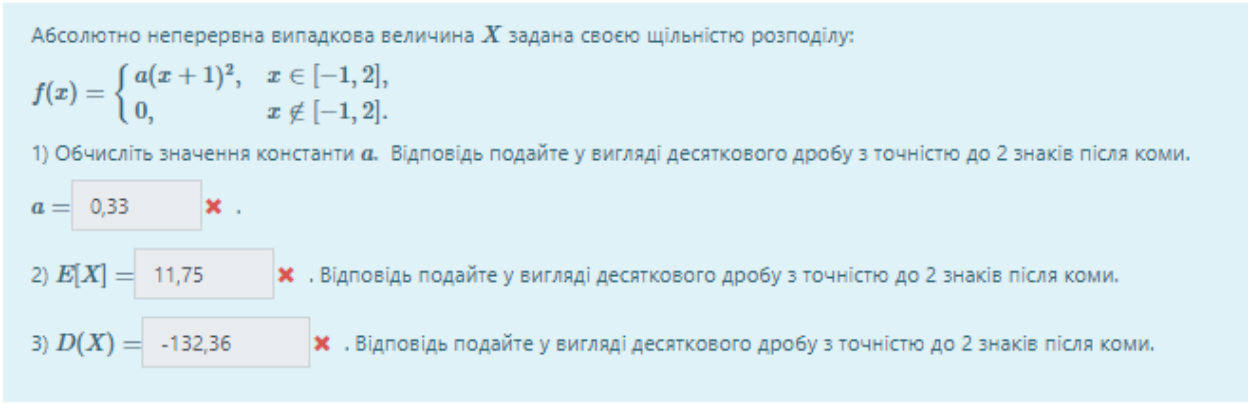
1) $E[X] =$ 1 ✔ ;

2) $E[Y] =$ 0,2 ✖ ;

3) Випадкові величини X і Y залежні ✔ .

Рис.1. Завдання з не зв'язаними кроками

У наступному тестовому завданні (рис.2) кроки логічно пов'язані між собою. У результаті, якщо буде неправильно розв'язаний перший крок, то це потягне за собою помилки в наступних кроках. У цьому завданні стала a визначає щільність випадкової величини, за допомогою якої обчислюються математичне сподівання і дисперсія випадкової величини. Коли константа знайдена неправильно, математичне сподівання і дисперсія теж будуть пораховані з помилкою.


Абсолютно неперервна випадкова величина X задана своєю щільністю розподілу:

$$f(x) = \begin{cases} a(x+1)^2, & x \in [-1, 2], \\ 0, & x \notin [-1, 2]. \end{cases}$$

1) Обчисліть значення константи a. Відповідь подайте у вигляді десяткового дробу з точністю до 2 знаків після коми.

$a =$ 0,33 ✖ .

2) $E[X] =$ 11,75 ✖ . Відповідь подайте у вигляді десяткового дробу з точністю до 2 знаків після коми.

3) $D(X) =$ -132,36 ✖ . Відповідь подайте у вигляді десяткового дробу з точністю до 2 знаків після коми.

Рис.2. Завдання із зв'язаними кроками

Можливим є й змішаний варіант (рис.3): деякі кроки між собою можуть бути пов'язані, а деякі - ні. Кроки 1 і 2 пов'язані між собою, оскільки коваріацію можна порахувати, використовуючи результат першого кроку. Кроки 3 і 4 не залежать від результатів попередніх кроків.

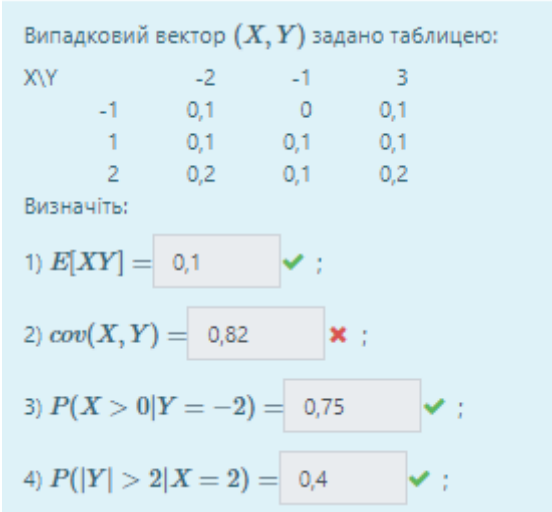


Рис.3. Завдання з частково зв'язаними кроками

Відсутність чи наявність подібних зв'язків між кроками у тестовому завданні потрібно враховувати при виборі відповідної моделі. Як видно з наведених прикладів, у різних завданнях перевіряються певні вміння студента, наприклад, уміння обчислювати числові характеристики випадкових величин.

Згідно з навчальною програмою (сілабусом) спеціальності 121 [https://do.ipo.kpi.ua/course/view.php?id=2358] можна виділити такі основні знання з теорії ймовірностей, які виявляє дана контрольна:

- означення дискретних і неперервних випадкових величин;
- означення та властивості функції розподілу випадкових величин;
- обчислення числових характеристик випадкових величин;
- системи випадкових величин та їх розподіл;
- обчислення числових характеристик систем випадкових величин;
- перевірка незалежності випадкових величин.

На думку авторів, найпростішим способом аналізу завдання типу ВВ було б застосування політомічної моделі Муракі GPCM, оскільки цей метод можна було застосовувати до матриці первинних балів, завантаженої безпосередньо з MOODLE. Але спочатку було застосовано інший метод. Було сформовано матрицю дихотомічних моделей відповідей. Для цього 8 питань типу ВВ було розбито на 28 дихотомічних завдань і сформовано єдину матрицю результатів. Дослідження тесту проводилось у наступні етапи:

A. Аналіз загальних характеристик тесту методами КТТ;
B. Аналіз тесту на основі дихотомічних моделей завдань;
C. Аналіз тесту на основі політомічних моделей завдань.

Розглянемо детальніше ці етапи.

A. Аналіз загальних характеристик тесту методами КТТ. На цьому етапі було побудовано гістограму первинних балів (рис.4) та оцінено такі вибіркові характеристики:

Вибіркове середнє = 21.77;
Середньо квадратичне відхилення = 6.27;
Коефіцієнт асиметрії = -1.88.

Вибіркові характеристики та деякий зсув управо пов'язані з високим рівнем підготовки студентів ФІОТ з математики, для яких середній бал ЗНО з математики вищий за 180 балів.

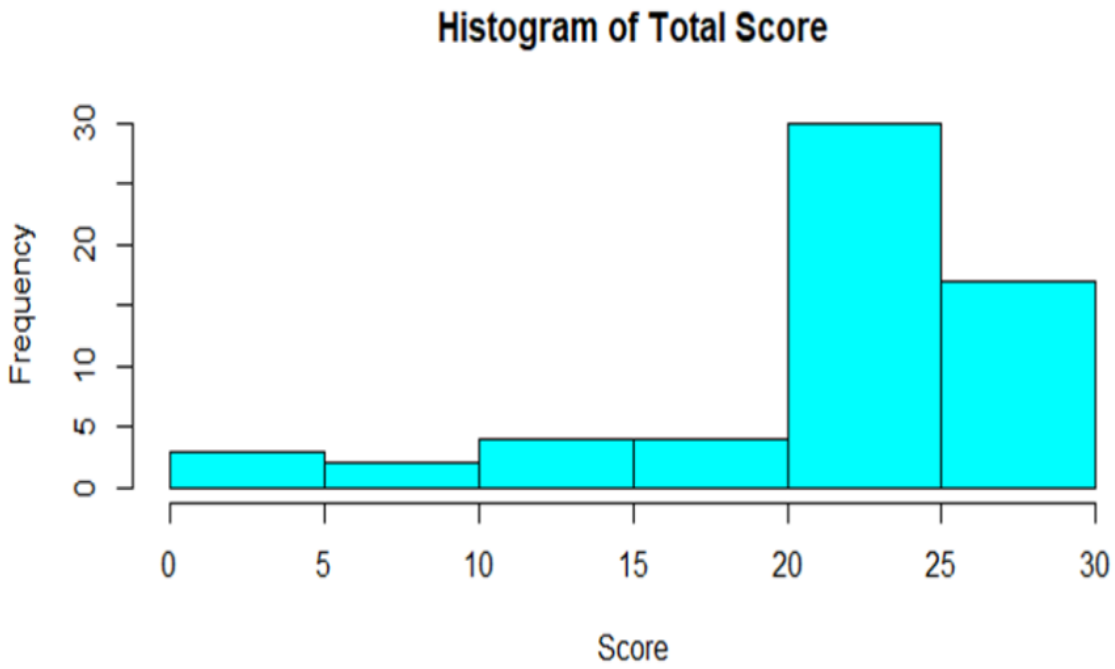


Рис.4. Гістограма розподілу первинних балів

Змістовну валідність забезпечує складання тесту експертами у відповідності до навчальної програми.

Для перевірки критеріальної валідності порівняно результати тесту і традиційної контрольної. Письмову контрольну роботу здавало 60 студентів спеціальності 121. Відповідний коефіцієнт кореляції між балами за тестову і письмову контрольні роботи складає 0.714. А для прогностичної валідності коефіцієнт кореляції між балами за тест і балом, отриманим на заліку, складає 0.806. Вочевидь кореляція є досить високою. Можна також за t-критерієм перевірити гіпотезу про рівність відповідних вибіркових середніх. Середній бал, отриманий за семестр -77.57, середній бал за тест за стобальною шкалою - 77.66. Для гіпотези H_0 про рівність середніх p-value = 0.9748>0.05. Отже, приймаємо нульову гіпотезу (середні рівні).

Для обох матриць (дихотомічних та політомічних моделей) було обчислено коефіцієнти надійності Кронбаха (пакет **psychometric** функція **alpha**):

Коефіцієнт Кронбаха для вихідної матриці балів:

$\alpha = 0.8486021$.

Коефіцієнт Кронбаха для дихотомічної матриці балів:

$\alpha = 0.9266837$.

Значення коефіцієнтів Кронбаха свідчать при високий рівень надійності тесту.

В. Аналіз тесту на основі дихотомічних моделей завдань. За сформованою єдиною матрицею дихотомічних моделей було побудовано такі моделі IRT: 1-PL; Раша; 2-PL; 3-PL. Оцінено латентні параметри моделей; побудовано характеристичні криві(ICC); інформаційні криві завдань (IIC) та інформаційну криву тесту (TIF).

Для оцінки параметрів було застосовано пакет **ltm**, функція **rasch.** Оцінені параметри дихотомічних моделей зібрані в таблицю1.

Таблиця 1

Латентні параметри моделей 1-PL;Раша;2-PL;3-PL

№ завдання	1-PL	Раша	2-PL		3-PL		
	α_j	α_j	α_j	D_j	α_j	D_j	γ_j
1	-3,15	-1,85	-1,07	3,33	-1,30	2,63	4,18E-07
2	-2,64	-1,51	-0,90	2,99	-0,95	2,98	7,91E-10
3	-2,25	-1,25	-0,88	2,24	-0,93	2,20	7,8E-10
4	-3,85	-2,34	-1,47	3,23	-1,03	29,45	0,268289
5	-2,88	-1,66	-0,87	4,00	-0,96	3,77	1,88E-10
6	-1,91	-1,03	-0,95	1,60	-1,06	1,45	6,39E-07
7	-1,63	-0,85	-0,96	1,28	-0,97	1,30	1,07E-09
8	-1,77	-0,94	-1,20	1,11	-1,33	1,01	3,91E-08
9	-1,50	-0,77	-0,59	1,95	-0,61	1,91	6,35E-11
10	-1,37	-0,69	-0,05	27,28	-0,01	91,00	3,83E-18
11	-1,77	-0,94	-0,07	35,34	-0,03	88,44	1,27E-17
12	-1,63	-0,85	-0,08	24,33	-0,02	84,87	1,46E-17
13	-1,77	-0,94	-0,06	38,32	-0,03	84,71	5,01E-17
14	-2,07	-1,14	-0,92	1,86	-1,08	1,59	4,6E-07
15	-2,43	-1,37	-1,34	1,47	-0,78	2,11	0,275293
16	-2,64	-1,51	-1,73	1,20	0,00	73,21	0,700007
17	-1,25	-0,61	-0,53	1,71	-0,55	1,66	1,18E-11
18	-0,82	-0,34	-0,24	1,96	-0,25	1,89	7,1E-10
19	1,68	1,25	8,73	0,15	7,30	0,18	0,001474
20	1,46	1,11	3,75	0,32	3,56	0,34	4,94E-06
21	-1,77	-0,94	-0,36	5,65	-0,37	5,39	6,94E-13
22	-1,63	-0,85	-0,33	5,25	-0,34	5,01	2,16E-11
23	-0,82	-0,34	-0,50	1,06	-0,60	0,93	1,26E-09
24	-1,91	-1,03	-0,77	2,07	-0,80	2,04	2,47E-11
25	-2,64	-1,51	-0,67	31,47	-0,41	96,50	0,300005
26	-2,07	-1,14	-0,65	3,02	-0,84	2,20	0,000165
27	-1,91	-1,03	-0,54	3,44	-0,03	40,94	0,300005
28	-2,07	-1,14	-0,62	3,22	-0,80	2,32	2,74E-05

Як видно з таблиці 1, аналізуючи коефіцієнти складності завдань, можна зробити висновок, що завдання мають невисоку складність. Але диференціююча здатність є додатною для всіх завдань. Проведено аналіз адекватності моделі 1-PL (пакет **ltm**, функція **GoF.rasch**) у цілому за критерієм χ^2, який показав її адекватність з рівнем значущості p-value= 0.05.

Також за критерієм χ^2 (пакет **ltm**, функція **item.fit**) проведено перевірку адекватності окремих завдань. Можна вважати, що не задовольняє моделі тільки задання 19, для нього p-value=0.0062. Для решти завдань показник p-value є вищим. Аналогічно проведено аналіз адекватності моделі Раша (пакет **ltm**, функція **GoF.rasch**) у цілому за критерієм χ^2, для якого p-value= 0.055, а також за критерієм χ^2 (пакет **ltm**, функція **item.fit**) проведено перевірку адекватності окремих завдань. Як і у випадку моделі 1- PL завдання 19 не підходить, для нього p-value=0.006. Для решти завдань p-value також є вищим.

Було побудовано ансамблі характеристичних кривих завдань та інформаційних кривих завдань (пакет **ltm**, функція **plot**) (рис. 5,6). В ансамблях кривих дещо осторонь розташовані криві завдань 19, 20, оскільки вони мають підвищену складність. Решта

кривих розташована достатньо рівномірно. Але завдання 19 і 20 виявились занадто складними через недоліки формулювання самих питань.

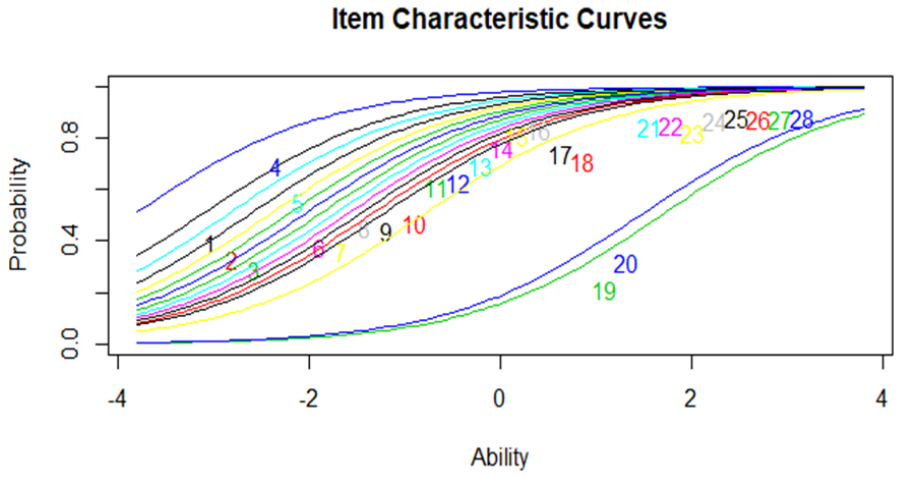


Рис.5. Характеристичні криві дихотомічних завдань для 1-PL моделі

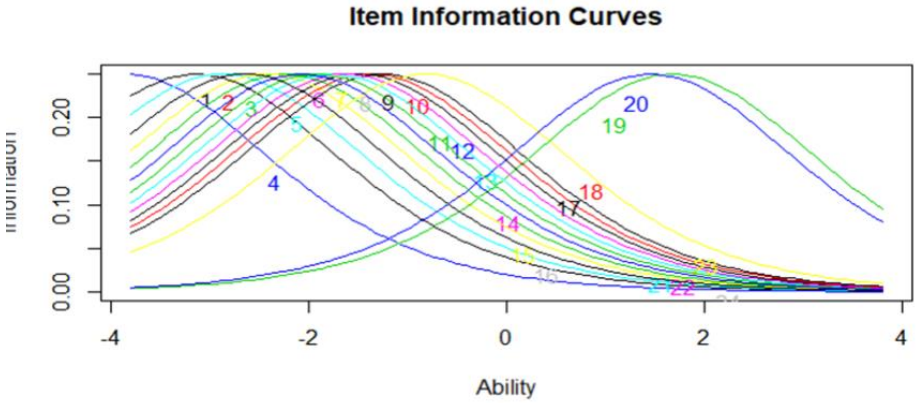


Рис.6. Інформаційні криві дихотомічних завдань моделі 1-PL моделі

Відповідні ансамблі для моделі Раша (функція **ltm**, функція **plot**) наведено на рисунках 7,8. Вони відповідають попередній моделі.

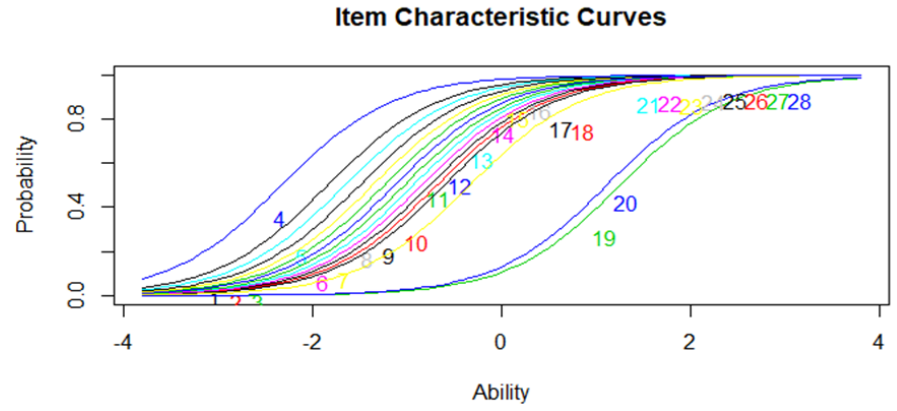


Рис.7. Характеристичні криві завдань для моделі Раша

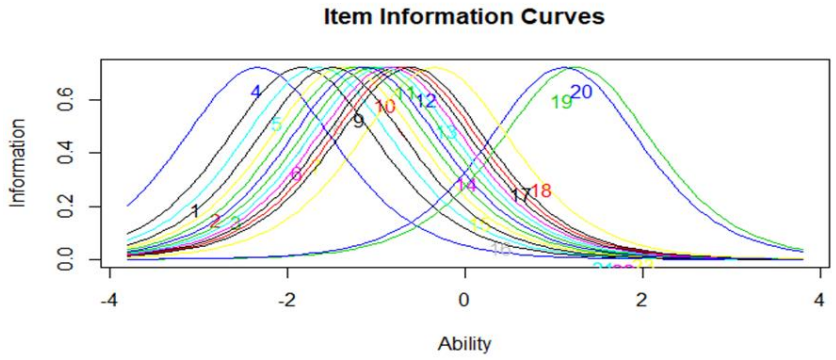


Рис. 8. Інформаційні криві завдань для моделі Раша

Порівняння моделі Раша і 1-PL моделі за інформаційними критеріями AIC, BIC (пакет **stats**, функція **anova**) [27] показали деяку перевагу моделі Раша. (див. таблицю 3). Було побудовано інформаційні функції тесту, які мають «правильну» дзвоноподібну форму (пакет **ltm**, функція **plot**). На рисунку 9 наведено інформаційну криву тесту (TIF) для моделі Раша.

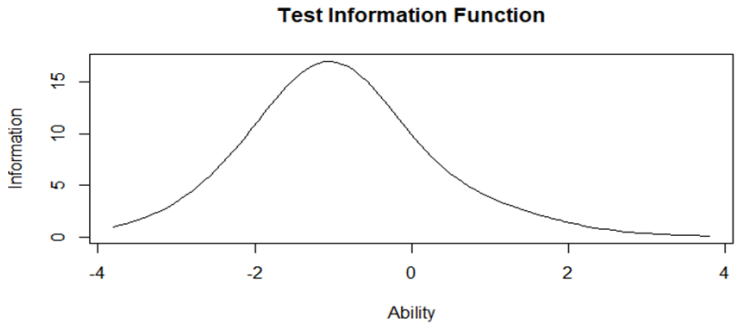


Рис.9. Інформаційна функція тесту

Аналогічно до аналізу за моделями 1-PL та Раша дані були проаналізовані за допомогою 2-PL та 3-PL моделей (пакет **ltm**, функції **tpm, ltm**). Але інформаційні критерії для цих моделей показали гірші значення. Про що й свідчать інформаційні функції завдань (рис.10). Графіки ансамблів характеристичних кривих також мають явну нерівномірність, що зумовлено введенням параметрів угадування у 3-PL модель. Це призвело до занадто низьких імовірностей вгадування більшості завдань.

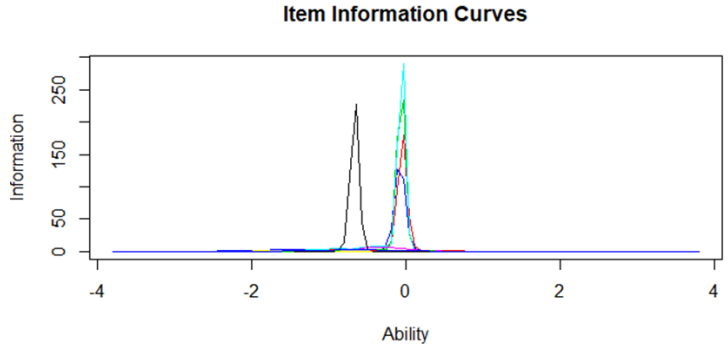


Рис.10. Інформаційні функції завдань

Значення інформаційних критеріїв для досліджених моделей наведено в таблиці 3. Отже, з дихотомічних моделей за інформаційними критеріями краще вибрати або модель Раша, або виправлену модель Раша.

С. Аналіз тесту на основі політомічних моделей завдань. Використовуючи пакет **ltm**, було оброблено дані за допомогою політомічної моделі, а саме моделі Муракі (пакет **ltm**, функція **gpcm**). Оцінено латентні параметри (таблиця 2), побудовано характеристичні криві (рис.11) та інформаційні криві завдань (рис.12), а також тесту загалом (рис.13).

Таблиця 2

Латентні параметри моделей моделі Муракі

№ Завдання	α_{j1}	α_{j2}	α_{j3}	α_{j4}	a_j
1	-0.069	-0.669	-4.392		0.757
2	-1.326	-1.995	-1.841		1.156
3	-1.514	-0.992	-2.049		0.636
4	-0.450	-1.326	-0.734		2.115
5	-3.177	-1.566	-1.249		1.200
6	1.077	-3.573	3.639	-1.539	0.525
7	-0.078	-1.080	-1.488	-1.539	0.763
8	-0.282	-0.206	-2.360	-2.107	0.959

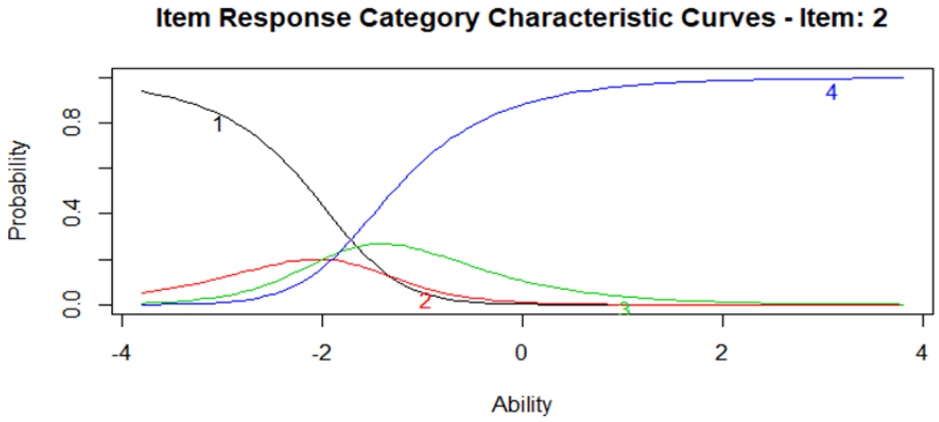


Рис.11. Характеристичні криві кроків завдання 2

Підверджено відповідним критерієм адекватність моделі даним (пакет **ltm**, функція **GoF.gpcm** ,p-value=0.583). Побудовано відповідні графіки інформаційних функцій завдань та тесту (пакет **ltm**, функція **plo**t) (рис.12,13), які мають «правильний» вигляд.

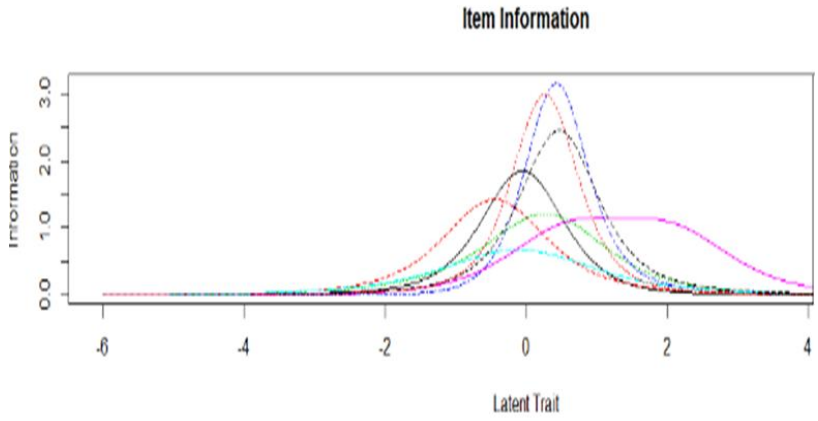


Рис.12. Інформаційні криві завдань

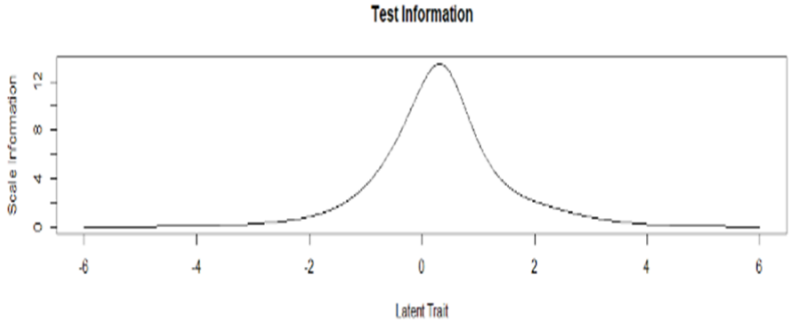


Рис.13. Інформаційна функція тесту

Проаналізуємо наведене на рисунку 2 політомічне задання 4 за допомогою характеристичних кривих. Так, у цьому завданні крок 1) перевіряє знання властивостей щільності випадкової величини та вміння інтегрувати; крок 2), використовуючи результат 1), перевіряє уміння обчислювати математичне сподівання, а також уміння інтегрувати; крок 3) використовує результати 1) і 2), а також знання означення дисперсії і знову уміння інтегрувати.

Аналіз характеристичних кривих (рис.14) дає можливість виявляти, наскільки правильно сформульовані кроки. Так, на рисунку 14 з аналізу взаємного розташування кривих видно, що крок 1) (на малюнку червоним 2) «провисає», тобто ймовірність досягнути цього рівня низька, а, вже перетнувши її, більшість виконує наступні кроки.

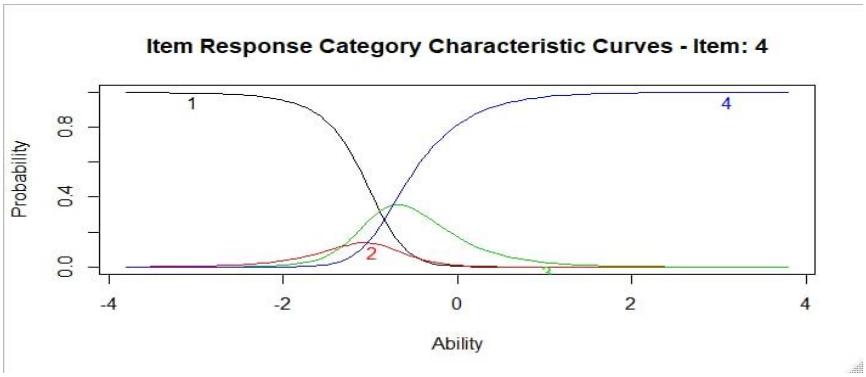


Рис.14. Характеристичні криві кроків завдання 4

Порівнюючи гістограми розподілу рівнів підготовленості, визначених за дихотомічною та політомічною моделями (рис. 15,16), треба зазначити, що в політомічному випадку розподіл ближчий до нормального, що більше відповідає даним.

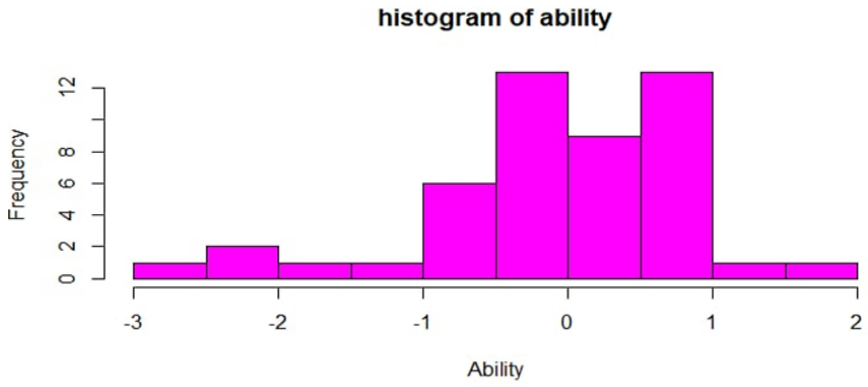


Рис.15.Розподіл рівнів підготовленості, визначеними за дихотомічною моделлю

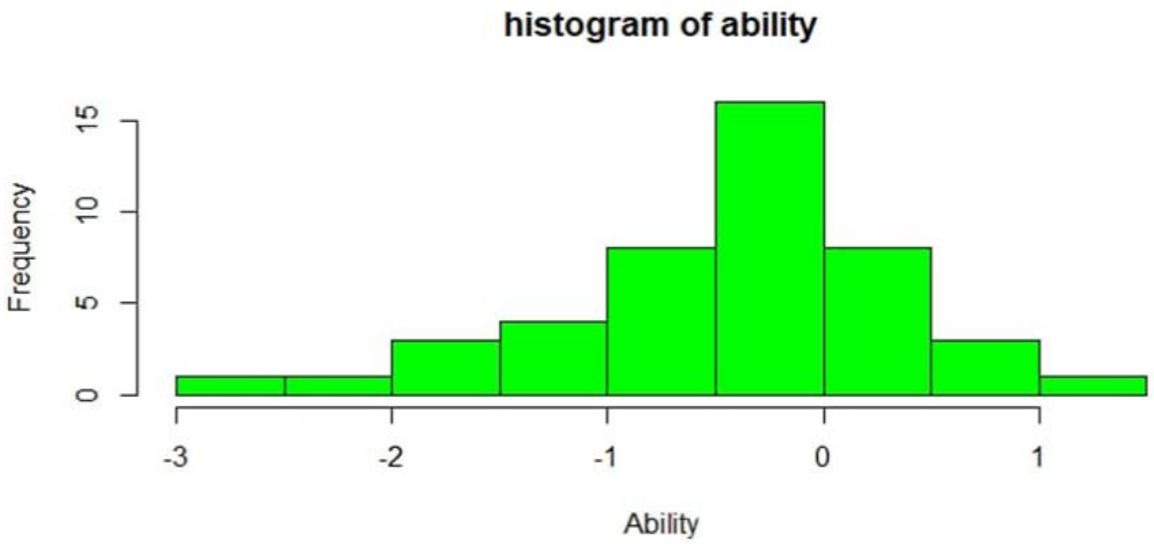


Рис.16. Розподіл рівнів підготовленості, визначеними за політомічною моделлю

Значення інформаційних критеріїв для всіх моделей зібрано у таблиці 3. Для моделі, яка краще відповідає даним, критерії набувають меншого значення.

Таблиця 3

Критерій\Модель	Раша	1-PL	2-PL	3-PL	Модель Муракі
AIC	1245.349	1249.867	1253.258	1302.093	857.3843
BIC	1327.404	1331.603	1370.541	1478.017	930.6864
AICc	1238.845	1328.867	3381.258	730.892	962.3843
SABIC	1143.393	1140.391	1194.407	1213.815	820.6021
G2	607.773	670.779	699.16	692	361.07

Отже, за відповідними інформаційними критеріями найбільше підходять політомічні моделі, а також Раша й 1-PL .

4. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

1. Розбиття завдань типу ВВ на дихотомічні є цілком прийнятним і побудована в такий спосіб модель є адекватною даним.
2. Ця процедура є достатньо «затратною», оскільки формування дихотомічної матриці є досить довгим (поки що виконується вручну).
3. Побудова дихотомічних моделей дає можливість перевірити рівень складності окремих кроків завдань типу ВВ та їх відповідність рівню підготовленості (інформаційна крива) іспитників. У нашому випадку має місце відповідність.
4. Дихотомічні моделі не враховують логічні зв'язки між кроками.
5. Більш прийнятними, на думку авторів, є політомічні моделі. Саме в них досягнення певного рівня відповідає мірі виконання завдань типу ВВ. У цих завданнях досить точно встановлюється зв'язок між заявленими вміннями, які перевіряються, і рівнями виконання політомічного завдання.
6. Важливим є те, що в таких завданнях перевіряється набір умінь. Наприклад, у завданні з теорії ймовірностей перевіряються уміння диференціювання, інтегрування, здобуті в попередньому семестрі.
7. Досвід використання бібліотеки CRAN, а саме пакетів **eRm, ltm, mirt**, показує, що вона є цілком достатньою і ефективною для аналізу різноманітних педагогічних тестів з математичних та інших дисциплін.
8. Застосування тестових завдань типу ВВ, очевидно, є ефективним засобом запобігання списуванню з онлайн ресурсів під час проведення дистанційного контролю (див.[7, 28]).
9. Проведене дослідження з формування, складання та аналізу тесту з теорії ймовірностей може бути застосовано й до контролю знань як з тем у вищій математиці, так і з інших дисциплін, а також сприятиме підвищенню компетентності викладачів щодо оцінювання якості тестів з математичних дисциплін, особливо під час дистанційного навчання студентів.
10. Складання більш якісних тестів підвищить об'єктивність та якість контролю знань студентів з математичних та інших дисциплін.
11. У подальшому автори вбачають перспективним застосування для аналізу таких завдань Багатовимірної Теорії Сучасних Тестів (MIRT) [29]. Частково це вже зроблено у [28].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Moodle.org: open-source community-based tools for learning. [Електронний ресурс]. Доступно: http://www.moodle.org

[2] В. П. Сергієнко, В. М. Франчук, Л.О. Кухар, *Методичні рекомендації зі створення тестових завдань та тестів у системі управління навчальними матеріалами MOODLE.,* К.: НПУ імені М.П. Драгоманова, 100 с., 2014. [Електронний ресурс]. Доступно: https://vfranchuk.fi.npu.edu.ua/images/files/statty/47.pdf

[3] А. М. Польовий, О. В. Шаблій О. В., І. А. Хоменко, Збірник методичних вказівок «Організація навчального курсу в СДН Moodle». ОДЕКУ, 182 с., 2019. [Електронний ресурс]. Доступно: http://eprints.library.odeku.edu.ua/id/eprint/6238/1/PolevoyAM_Shabl%C3%ADyOV_Khomenko%C3%8DA_Zb%C3%ADrnik_metodichnikh_vkaz%C3%ADvok_Organ%C3%ADzats%C3%ADya_navchal'nogo_kursu_v_SND_Moodle_%20(yel_vidannya)%20_2019.pdf

[4] П. І. Андронатій, В.В. Котяк, Комп'ютерні технології в освітніх вимірюваннях: Навчально-методичний посібник. Кіровоград: Лисенко В.Ф., 144 с., 2011. [Електронний ресурс]. Доступно: http://moodle.ndu.edu.ua/pluginfile.php/889/mod_page/content/1/Andronatiy_Kotiak_Komp_tehnol_v_Osvit_Vymir.pdf

[5] А. Ф. Дудко, *Комп'ютерно орієнтована методика оцінювання якості тестів з вищої математики викладачами вищої освіти.* Автореферат дисертації на здобуття наукового ступеня кандидата педагогічних наук, Київ, 24 с., 2019.
[6] О. О. Диховичний, та А. Ф. Дудко, "Комплексна методика аналізу якості тестів з вищої математики", *Науковий часопис НПУ імені М.П. Драгоманова. Серія №2. Комп'ютерно-орієнтовані системи навчання: Зб. наук. праць*, №15 (22), с. 140-144, 2015.
[7] Н. В. Круглова, О.О. Диховичний, та І.В. Алєксєєва, Особливості застосування математичних моделей тестів в умовах дистанційного контролю. *Математичні машини і системи,* №. 2, с. 105-116, 2020.
[8] В. Ю. Биков, Ю. М. Богачков, та Ю. О. Жук, *Моніторинг рівня навчальних досягнень з використанням Інтернет-технологій: монографія*. Київ, Україна: Педагогічна думка, 128 с., 2008.
[9] O. H. Kolhatin, Requirements on the design of the automated system of pedagogical diagnostics. *Information Technologies and Learning Tools*, *19*(5), 2010, doi:https://doi.org/10.33407/itlt.v19i5.352
[10] O. H. Kolhatin, Experience of the automated system application for diagnostics of the learning achievements in the mathematical statistics methods. *Information Technologies and Learning Tools*, *5*(1),2008. doi:https://doi.org/10.33407/itlt.v5i1.153
[11] Y. M. Bogachkov, and P. S. Ukhan, Application of qualifiers in the systems of measurement of the level of educational achievements. *Information Technologies and Learning Tools*, *4*(3),2007. doi:https://doi.org/10.33407/itlt.v4i3.169
[12] L. Crocker, and J. Algina, *Introduction to Classical and Modern Test Theory*. Belmont, CA:Wadsworth, 527 p., 2006.
[13] М.Б. Челышкова, *Теория и практика конструирования педагогических тестов.* М.: Логос, 432 с., 2002.
[14] G. Rasch, *Probabilistic models for some intelligence and attainment tests.* Danish Institute for Educational Research, Copenhagen, 126 p., 1960.
[15] M.Wilson, *Constructing Measures: An Item Response Modeling Approach.* Mahwah, NJ: Erlbaum, 2005.
[16] F. Baker, *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, 2001. [Електронний ресурс]. Доступно: http://echo.edres.org:8080/irt/baker/ Дата звернення: Квітень 24, 2020.
[17] W. Linden and R. Hambleton, *Handbook of modern item response theory*. New York: Springer, 2018.
[18] В. С. Аванесов, Item Response Theory: Основные понятия и положения, *Педагогические измерения,* №2, с.3-28, 2007.
[19] Т. В. Лісова, *Моделі та методи сучасної теорії тестів.* Ніжин, Україна: Видавець ПП Лисенко М. М., 2012.
[20] E. Muraki, and B. Carlson, Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement.* Vol. *19*, N 1. pp. 73–90, 1995.
[21] The Comprehensive R Archive Network. [Електронний ресурс]. Доступно: http://cran.us.r-project.org
[22] W. Holmes Finch, Jocelyn E. Bolin, and Ken Kelley *Multilevel Modeling Using R* , 2019 by Chapman and all/CRC Press, 252 p., 2019.
[23] David A. Armstrong II, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal *Analyzing Spatial Models of Choice and Judgment with R*, CRC Press, Boca Raton, 336 p., 2014.
[24] В. С. Ким, *Тестирование учебных достижений. Монография*. Уссурийск: Издательство УГПИ, 214 с., 2007.
[25] *Энциклопедия психологических тестов. Личность Мотивация. Потребность*. /Под ред. А. Карелина. М., 312 с., 1997.
[26] *Handbook of item response theory.* volume two, statistical tools edited by Wim J. van der Linden, Boca Raton, FL, CRC Press, 427 p., 2016.
[27] M. Chambers, Trevor Hastie, *Statistical Models in S*. Wadsworth & Brooks/Cole Advanced Books & Software, 608 p., 1992.
[28] N. Kruglova, O. Dykhovychnyi, D. Lysenko, Application of IRT and MIRT models to analysis of analytical geometry tests. *Міжвідомчий науково-технічний збірник «Адаптивні системи автоматичного управління»*, № 1 (38), pp. 36-49, 2021.
[29] M. D. Reckase, *Multidimensional item response theory*. New York: Springer, 2009.

Матеріал надійшов до редакції 19.04 2021 р

SELECTING A MATHEMATICAL MODEL FOR ANALYSIS OF TEST ITEMS OF THE TYPE "EMBEDDED ANSWERS" FOR MATHEMATICAL DISCIPLINES

Nataliia V. Kruglova
PhD of Physical and Mathematical Sciences,
Associate Professor at the Mathematical Analysis and Probability Theory Department
National Technical University of Ukraine “Igor Sicorsky Kyiv Politechnic Institute”, Kyiv, Ukraine
ORCID ID 0000-0002-9143-9398
natahak@ukr.net

Oleksandr O. Dykhovychnyi
PhD of Physical and Mathematical Sciences,
Associate Professor at the Mathematical Analysis and Probability Theory Department
National Technical University of Ukraine “Igor Sicorsky Kyiv Politechnic Institute”, Kyiv, Ukraine
ORCID ID 0000-0001-8505-2878
a.dyx@ukr.net

Abstract. The article examines the issue of choosing a mathematical model for the analysis of computer tests in mathematical disciplines. A feature of this study is the exploration of test items of the "embedded answers (cloze)" type. Such test tasks have several “steps” interconnected by a certain logic and have been used precisely for remote control of knowledge in mathematical disciplines during quarantine at NTUU “Igor Sikorsky Kyiv Polytechnic Institute”.The introduction of such tasks into the test raises the question for the teacher: how to analyze the quality of such tests? The authors built the study on the basis of the well-known methods of Classical Test Theory (CTT) and Item Response Theory (IRT), which have proven their effectiveness in pedagogical testing as well as in other industries in which the test approach is used. The study was carried out on the example of a modular test quizzes on probability theory for second-year students of the Faculty of Informatics and Computer Engineering of NTUU "Igor Sikorsky Kyiv Polytechnic Institute", formed on the MOODLE platform during distance learning in November 2020. In the course of the study, the following was done: the quality of the test was analyzed both as a whole and as individual test items; the latent parameters of the tasks were estimated, the adequacy of the corresponding models to the testing data was checked; characteristic and information curves were built; using information criteria, were selected the IRT models the most suitable for the analysis of computer tests with tasks of the "embedded answers" type; it has been proven that both Rush's dichotomous model and Muraki's political model are the most appropriate for the analysis of such tasks. The calculation was carried out using special libraries (packages) of R programming language, in which the KTT and IRT algorithms are implemented, namely, eRm, ltm, mirt. It is shown how the tasks with “embedded answers (cloze)” can be used to analyze certain skills and knowledge of students. The results of the study contribute to improving the competence of teachers in assessing the quality of tests in higher mathematics, which is very important when compiling tests in various disciplines, especially in the situation of student distance learning.

Keywords: computer tests; "embedded answers (cloze)"; Rasch’s model; polytomous model; R programming environment; packages eRm и ltm; MOODLE.

REFERENCES (TRANSLATED AND TRANSLITERATED)

[1] Moodle.org: open-source community-based tools for learning. [Online]. Available: http://www.moodle.org (in English).

[2] V. P. Serhiienko, V. M. Franchuk, L. O. Kukhar, *Methodical recommendations for creating test tasks and tests in the management system of educational materials MOODLE*, K.: NPU imeni M.P. Drahomanova, 100 ., 2014. [Online]. Available: https://vfranchuk.fi.npu.edu.ua/images/files/statty/47.pdf (in Ukrainian).

[3] A. M. Polovyi, O. V. Shablii, I. A. Khomenko, *Collection of methodical instructions "Organization of a training course in Moodle"*. ODKEU, 182 p., 2019. [Online]. Available: http://eprints.library.odeku.edu.ua/id/eprint/6238/1/PolevoyAM_Shabl%C3%ADyOV_Khomenko%C3%8DA_Zb%C3%ADrnik_metodichnikh_vkaz%C3%ADvok_Organ%C3%ADzats%C3%ADya_navchal'nogo_kursu_v_SND_Moodle_%20(yel_vidannya)%20_2019.pdf (in Ukrainian).

[4] P. I. Andronatii, V. V. Kotiak, *Computer technology in educational measurements*. Navchalno-metodychnyi posibnyk. Kirovohrad: Lysenko V.F, 144 p., 2011. [Online]. Available:

http://moodle.ndu.edu.ua/pluginfile.php/889/mod_page/content/1/Andronatiy_Kotiak_Komp_tehnol_v_Osvit_Vymir.pdf (in Ukrainian).

[5] A. F. Dudko, Computer-oriented methodology of assessing the quality of tests in higher mathematics by teachers of higher education establishments. *Avtoreferat dysertatsii na zdobuttia naukovoho stupenia kandydata pedahohichnykh nauk*, Kyiv, 24 p., 2019. (in Ukrainian) .

[6] O. O. Dykhovychnyi, and A. F. Dudko, "The comprehensive procedure of analysis of tests in higher mathematics", *Naukovyi chasopys NPU imeni M.P. Drahomanova*, *Seriia №2. Kompiuterno-oriientovani systemy navchannia: Zb. nauk. prats*, no. 15 (22), pp. 140-144, 2015. (in Ukrainian).

[7] N. V. Kruglova , O. O. Dykhovychnyi, and I.V. Alekseeva, Features of mathematical application test models in the conditions of remote control. *Mathematical Machines and Systems*, no.. 2, pp. 105-116, 2020. (in Ukrainian).

[8] V. Iu. Bykov, Yu. M. Bohachkov, ta Yu. O. Zhuk, *Monitoring the evaluation of educational achievement using Internet technologies: monograph*. Kyiv, Ukraine: Pedahohichna dumka, 2008. (in Ukrainian).

[9] O. H. Kolhatin. Requirements on the design of the automated system of pedagogical diagnostics. *Information Technologies and Learning Tools*, *19*(5), 2010, doi:https://doi.org/10.33407/itlt.v19i5.352 (in Ukrainian).

[10] O. H. Kolhatin, Experience of the automated system application for diagnostics of the learning achievements in the mathematical statistics methods. *Information Technologies and Learning Tools*, *5*(1),2008. doi:https://doi.org/10.33407/itlt.v5i1.153 (in Ukrainian).

[11] Y. M. Bogachkov, & P. S. Ukhan, Application of qualifiers in the systems of measurement of the level of educational achievements. *Information Technologies and Learning Tools*, *4*(3),2007. doi:https://doi.org/10.33407/itlt.v4i3.169 (in Ukrainian).

[12] L. Crocker, and J. Algina, *Introduction to Classical and Modern Test Theory*. Belmont, CA:Wadsworth, 527 p., 2006. (in English).

[13] M. B. Chelyshkova, *Theory and practice of constructing of pedagogical tests*. Moscow, RF: Logos, 2002. (in Russian).

[14] G. Rasch, Probabilistic models for some intelligence and attainment tests. *Danish Institute for Educational Research*, Copenhagen, 126 p., 1960 (in English).

[15] M.Wilson, *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum, 2005. (in English).

[16] F. Baker, *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, 2001. [Online]. Available: http://echo.edres.org:8080/irt/baker/ Accessed on: Apr. 24, 2020. (in English).

[17] W. Linden and R. Hambleton, *Handbook of modern item response theory*. New York: Springer, 2018 (in English).

[18] V. S. Avanesov, Item Response Theory: The basic concepts and propositions. *Pedagogicheskie izmerenija*, №2, pp. 3-28., 2007, (in Russian).

[19] T. V. Lisova, *Models and methods of the modern test theory*. Nizhyn, Ukraina: Vydavets PP M. M., 2012. (in Ukrainian).

[20] E. Muraki, and B. Carlson,. Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*. Vol. *19*, N 1. pp. 73–90, 1995. (in English).

[21] The Comprehensive R Archive Network. [Online]. Available: http://cran.us.r-project.org

[22] W. Holmes Finch, Jocelyn E. Bolin, and Ken Kelley, *Multilevel Modeling Using R* , 2019 by Chapman and all/CRC Press, 252 p., 2019. (in English).

[23] David A. Armstrong II, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal, *Analyzing Spatial Models of Choice and Judgment with R*, CRC Press, Boca Raton, 336 p., 2014. (in English).

[24] V. S. Kym, *Testing of educational achievements. Monograph.* Ussuryisk: Yzdatelstvo UHPY, 214 p., 2007, (in Russian).

[25] *Encyclopedia of psychological tests. Personality Motivation. Requirement.* Pod red. A. Karelyna. M., 312 p., 1997. (in Russian).

[26] *Handbook of item response theory*. volume two, statistical tools edited by Wim J. van der Linden, Boca Raton, FL, CRC Press, 427 p., 2016. (in English).

[27] M. Chambers, Trevor Hastie, *Statistical Models in S*. Wadsworth & Brooks/Cole Advanced Books & Software, 608 p., 1992. (in English).

[28] N. Kruglova, O. Dykhovychnyi, D. Lysenko, Application of IRT and MIRT models to analysis of analytical geometry tests. *Mizhvidomchyi naukovo-tekhnichnyi zbirnyk «Adaptyvni systemyavtomatychnoho upravlinnia»*, № 1 (38), pp. 36-49, 2021. (in English).

[29] M. D. Reckase, *Multidimensional item response theory*. New York: Springer, 2009. (in English).

(cc) BY-NC-SA

This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.