

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ

Кафедра фундаментальної та прикладної математики

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему: «ЗАСТОСУВАННЯ МАТЕМАТИЧНИХ МЕТОДІВ ТА
ТЕСТОВИХ ТЕХНОЛОГІЙ ДЛЯ ДІАГНОСТИКИ НАВЧАЛЬНИХ
ДОСЯГНЕНЬ ЗДОБУВАЧІВ МАТЕМАТИЧНОЇ ОСВІТИ »

Виконала: студентка 2 курсу, групи 8.1112-з
спеціальності 111 Математика
(шифр і назва спеціальності)
освітньої програми Математика
(назва освітньої програми)
Н.О. Бібла
(ініціали та прізвище)

Керівник доцент кафедри фундаментальної та прикладної
математики, доцент, к.ф.-м.н. Панасенко Є.В.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Рецензент доцент кафедри програмної інженерії,
доцент, к.ф.-м.н. Кудін О.В.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Запоріжжя

2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет математичний

Кафедра фундаментальної та прикладної математики

Рівень вищої освіти магістр

Спеціальність 111 Математика

(шифр і назва)

Освітня програма Математика

ЗАТВЕРДЖУЮ

Завідувач кафедри
фундаментальної та прикладної
математики, д.т.н., професор

_____ Гребенюк С.М.
(підпис)

« _____ » _____ 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТЦІ

Біблі Ніні Олександрівні

(прізвище, ім'я та по батькові)

1. Тема роботи Застосування математичних методів та тестових технологій
для діагностики навчальних досягнень здобувачів математичної освіти

керівник роботи Панасенко Євген Валерійович, к.ф.-м.н., доцент

(прізвище, ім'я та по батькові, науковий ступінь, вчене звання)

затверджені наказом ЗНУ від « 01 » травня 2023 року № 643-с

2. Строк подання студентом роботи _____

3. Вихідні дані до роботи 1. Постановка задачі.

2. Перелік літератури.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Постановка задачі

2. Основні теоретичні відомості

3. Задача

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

Презентація

6. Консультанти розділів роботи |

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Постановка задачі та плану роботи	08.09.2023	
2.	Збір вихідних даних	12.09.2023	
3.	Обробка методичних та теоретичних джерел	20.09.2023	
4.	Розробка першого та другого розділу	06.10.2023	
5.	Розробка третього розділу	03.11.2023	
6.	Оформлення та нормо контроль кваліфікаційної роботи	24.11.2023	
7.	Захист кваліфікаційної роботи	17.12.2023	

Студент _____
(підпис)Н.О. Бібла _____
(ініціали та прізвище)Керівник роботи _____
(підпис)Є.В. Панасенко _____
(ініціали та прізвище)**Нормоконтроль пройдено**Нормоконтролер _____
(підпис)О.Г. Спиця _____
(ініціали та прізвище)

РЕФЕРАТ

Кваліфікаційна робота магістра «Застосування математичних методів та тестових технологій для діагностики навчальних досягнень здобувачів математичної освіти» містить: 57 сторінок, 10 малюнків, 7 таблиць, 15 джерел.

ВАЛІДНІСТЬ, ІНДЕКС ДИСКРИМІНАТИВНОСТІ, КОЕФІЦІЄНТ КОРЕЛЯЦІЇ, КОМП'ЮТЕРНА СИСТЕМА ТЕСТУВАННЯ ЗНАНЬ, КОРЕЛЯЦІЙНА МАТРИЦЯ.

Об'єкт дослідження – тестові технології та математичні методи дослідження.

Метою дипломної роботи є оволодіння методикою математичних методів дослідження результатів тестування та застосування їх для діагностики навчальних досягнень здобувачів математичної освіти

Метод дослідження – аналітичний.

У першому розділі подано дослідження методу тестування в системі педагогічних вимірювань: розглянута змістовна сутність педагогічної діагностики, класифікація тестів,

У другому розділі досліджено застосування комп'ютерних технологій для тестування знань, розглянуто його переваги та недоліки, класифікацію та основні вимоги до комп'ютерних систем тестування знань.

У третьому розділі зроблено дослідження статистичного методу аналізу результатів тестування. Також зроблено статистичний аналіз результатів тематичного тесту «Матриці».

Практичне значення роботи полягає у оволодінні методикою застосування математичних методів та тестових технологій для діагностики навчальних досягнень здобувачів математичної освіти.

SUMMARY

Master's Qualifying Theses «Test Technologies and Their Application for Diagnosing Educational Achievements of Students in the Mathematics»: contains 57 pages, 10 figures, 7 tables and 15 sources.

VALIDITY, DISCRIMINATION INDEX, CORRELATION COEFFICIENT, COMPUTER KNOWLEDGE TESTING SYSTEM, CORRELATION MATRIX.

The object of the research is testing technologies and mathematical methods of research.

The aim of the thesis is to master the methodology of mathematical methods of research of test results and their application to diagnose the educational achievements of students of mathematics education.

The research method is analytical.

The first chapter presents a study of the testing method in the system of pedagogical measurements: the substantive essence of pedagogical diagnostics and the classification of tests.

The second chapter examines the application of computer technologies for knowledge testing, considers its advantages and disadvantages, classification, and basic requirements for computer knowledge testing systems.

In the third chapter, a study of the statistical method of analysis of test results is carried out. A statistical analysis of the results of the thematic test "Matrix" was also made.

The practical significance of the work consists in mastering the methodology of applying mathematical methods and test technologies for diagnosing educational achievements of students of mathematics education.

ЗМІСТ

Завдання на кваліфікаційну роботу	2
Реферат	4
Summary	5
Вступ.....	7
1 Тестові технології для діагностики навчальних досягнень	9
1.1 Історія розвитку створення тестів	9
1.2 Метод тестування у системі педагогічних вимірювань	11
1.3 Класифікація тестів. Форми тестових завдань.....	20
2 Комп'ютерне тестування, як метод контролю якості знань з математичних наук	24
2.1 Комп'ютерне тестування знань, його переваги та недоліки.....	24
2.2 Класифікація та основні вимоги до комп'ютерних систем тестування знань.....	27
2.3 Характеристика комп'ютерних систем тестування знань	30
3 Статистичний аналіз результатів тестування	35
3.1 Класичні статистичні методи аналізу результатів тестування	35
3.2 Структура тестової контрольної роботи, що аналізується.....	40
3.3 Аналіз якості тесту в цілому	44
Висновки	54
Перелік посилань.....	55

ВСТУП

Процес європейської інтеграції вже значно впливає на всі сфери життя держави, і вища освіта в Україні теж відчуває його вплив. Наша країна чітко визначила свої цілі щодо вступу в освітній і науковий простір Європи та активно працює над покращенням якості освіти, враховуючи європейські стандарти.

З цієї причини питання підвищення якості освіти стає дедалі актуальнішим, і вдосконалення процесів управління підвищенням ступінь вартості вищої освіти передбачає покращення ефективності контролю і оцінки рівня підготовки студентів на різних етапах навчання. Відтак, педагогічний контроль стає необхідністю на кожному етапі навчання.

Вибір методів діагностики в освіті базується на відповідних дидактичних принципах вимірювання результатів навчання. Лише добре підготовлений комплекс завдань дозволяє використовувати певні діагностичні методи для об'єктивної оцінки рівня знань та вмінь студентів. Таким чином, в сучасній педагогіці активно використовуються кількісні методи педагогічного контролю. Зокрема, метод тестування є одним з науково обґрунтованих засобів об'єктивного контролю.

Своєчасний контроль знань допомагає студентам зосередитися на головних аспектах інформації та робить навчальний процес більш цікавим. Результати контролю надають можливість студентам оцінити свій прогрес у засвоєнні курсу, а також допомагають викладачам адаптувати методи навчання до різних рівнів знань студентів. Використання електронних тестів для контролю знань є ефективним і об'єктивним підходом.

Комп'ютерне тестування стає все важливішим і актуальнішим засобом контролю за рівнем засвоєння знань студентами вищих навчальних закладів. Проведення електронних контрольних і іспитів в університетах стають

вельми важливими, оскільки ця форма контролю відзначається високою ефективністю та об'єктивністю.

Існують декілька причин, які підкреслюють необхідність впровадження електронних засобів контролю в освітньому процесі. По-перше, зменшення кількості годин аудиторних занять та акцент на самостійній роботі студентів вимагають більш ефективного контролю та управління навчальним процесом. Своєчасний контроль знань допомагає студентам зосередити увагу на ключових аспектах матеріалу і робить навчання більш продуктивним.

Крім того, електронне тестування забезпечує можливість об'єктивно оцінювати знання студентів, оскільки воно виключає суб'єктивний фактор, який може виникнути при інших методах оцінювання. Результати контролю дозволяють кожному студентові оцінити свій прогрес і активізують навчальний процес.

Аналіз результатів контролю також надає можливість викладачам вдосконалювати методи навчання, а також адаптувати підходи до студентів із різними рівнями знань. Електронні тести є зручним і швидким способом контролю, оскільки вони дозволяють оцінювати різні аспекти знань, включаючи теоретичний матеріал, практичні навички та розрахункові завдання.

Загалом, комп'ютерне тестування є перспективним і об'єктивним методом контролю знань, який забезпечує ефективну та об'єктивну оцінку навчальних результатів студентів. Використання онлайн тестів дозволяє докладно перевіряти рівень засвоєння матеріалу та здійснювати оперативну діагностику кожного студента, що сприяє покращенню навчального процесу та ефективному використанню часу.

У практичній частині даної кваліфікаційної роботи, якій присвячений третій розділ, розглянуто статистичні методи дослідження та проведено статистичний аналіз результатів тестової тематичної перевірки знань студентів.

1 ТЕСТОВІ ТЕХНОЛОГІЇ ДЛЯ ДІАГНОСТИКИ НАВЧАЛЬНИХ ДОСЯГНЕНЬ

1.1 Історія розвитку створення тестів

Першими прикладами шкільних тестів для оцінки успішності учнів були розроблені у 1864 році Джорджем Фішером під назвою «Шкальовані книги», а також у 1894 році Дж. М. Раїсом, який створив тести у вигляді таблиць для перевірки орфографії учнів [7]. Проте, наукові дослідження в області теорії тестів почали з'являтися лише на початку ХХ століття. Особливий внесок вніс Ф. Гальтон, який сформулював три ключові принципи в теорії тестування:

- статистичний аналіз результатів;
- визначення критеріїв оцінки;
- застосування серії однакових досліджень до великої кількості людей, які проходять тест.

Карл Пірсон у 1896 році вніс важливий внесок у розвиток теорії кореляції. Того ж року психолог Дж. М. Кеттел в своїй роботі «Розумові тести» висловив ідею використання тестування як наукового інструмента та встановив ряд вимог до проведення тестів, включаючи однакові умови для учнів, обмежений час (не більше 60 хвилин), відсутність наглядачів тощо.

У 1916 році Л.М. Термен розробив шкалу для оцінки знань учнів, яка пізніше була відома як Станфордська шкала Біне, і ввів поняття коефіцієнта IQ, що відображав співвідношення інтелекту до віку досліджуваних осіб. Після цього в Америці стали розробляти «інтелектуальні» тести для школярів.

У ХХ столітті американський дослідник В.А. Макколл класифікував тести на два типи: психологічні і педагогічні.

Едуард Лі Торндайк, американський психолог, автор книги «Вступ до теорії психології та соціальних вимірювань», вважається батьком педагогічних тестів. М. Стоун створив перший стандартизований тест (з арифметики) під керівництвом Едуарда Лі Торндайка. У 1915 році Єркс склав неординарний комплект тестів, в якому змінювалися методи підрахунку [8].

З 60-х років ХХ століття тестування отримало новий розквіт. Видатний американський педагог Р. Тайлер розробив програму "National Assessment of Educational Progress" (NAEP), яка зробила вплив на впровадження методу тестування в галузь освіти США. Головною ідеєю цієї програми було надання педагогам можливості самостійно вдосконалювати систему освіти та навчальний процес.

В Україні розвиток тестування розпочався лише після отримання незалежності, оскільки наслідки колишнього СРСР впливають на сучасну систему освіти. Проте, з 90-х років ХХ століття Україна почала активно впроваджувати тестування як частину системи прийому до вищих навчальних закладів. Спочатку це стосувалося лікарів, коли за підтримки міжнародних фондів, у 1993 році Міністерство охорони здоров'я України впровадило тестування для ліцензування медичних працівників за допомогою системи «Кроки». Система «Кроки» стала першою в Україні системою тестування на державному рівні, і до неї був створений банк завдань, який включає всі дані тестування та проводить їх аналіз.

З 2002 по 2005 роки в Україні був запроваджений експеримент з впровадження зовнішнього незалежного оцінювання (ЗНО) в систему загальної середньої освіти за підтримки Центру тестових технологій, Міністерства освіти і науки України, та Міжнародного фонду «Відродження». У першому етапі дослідження взяли участь 3121 учень та 4 університети. З роками кількість учасників і університетів, які долучилися до ЗНО, збільшувалася. У 2005 році було вже 8700 випускників ЗНЗ та всі вищі навчальні заклади України. З 2006 року проведення ЗНО стало національним і обов'язковим для вступу до вищих навчальних закладів. Зарахування до

першого курсу університету почало здійснюватися тільки на підставі результатів ЗНО [9].

1.2 Метод тестування у системі педагогічних вимірювань

Сьогодні на практиці було підтверджено твердження про те, що правильна оцінка є ключовим елементом у визначенні змісту навчання. Це означає, що без детального вимірювання того, що вже маємо, неможливо здійснити зміни та вдосконалення.

В розробці питання про вимірювання важливі три питання: чому це потрібно, що і яким чином вимірювати? Відповідь на перше питання досить проста і пов'язана з встановленням цілей контролю. Якщо ціль – оцінка досягнень учнів, то основна увага приділяється перевірці та визначенню рівня та якості їх підготовки.

Для оцінки досягнень учнів найчастіше використовують рівень та якість підготовки як об'єкт вимірювання. У контексті вимірювання знань учнів тести використовуються як інструмент вимірювання.

Тестовий контролю має такі функції: управляючу, мотиваційно-орієнтуючу, оціночну, навчальну, освітньо-виховну, розвиваючу, коригуючу і прогнозуючу. Основною й найочевидніша функція тестування є діагностична функція, яка полягає у виявленні рівня знань, умінь, навичок студентів. Ця функція діагностування за об'єктивністю і швидкістю тестування перевершує всі інші форми і методи педагогічного контролю.

Процес вимірювання рівня знань включає такі основні категорії: метод вимірювання, інструмент вимірювання (за допомогою якого відбувається вимірювання), процедура вимірювання (як саме відбувається вимірювання), та процедура оцінювання (як виміряна величина отримує кількісне визначення).

Метод тестування передбачає, що тест є інструментом вимірювання, процедурою вимірювання є тестування, а методом оцінювання є шкалювання. З 50-х років ХХ століття педагогічна наука розвинула окрему галузь, відому як тестологія.

Тестологія – це наукове дослідження в галузі теорії та практики вимірювання та оцінювання якісних характеристик, що притаманні людині, де методом вимірювання є тестування.

Переваги використання тестів порівняно з іншими методами оцінювання полягають у наступному: спрощення процесу оцінювання та зменшення суб'єктивного впливу вчителя; можливість використання технології, включаючи комп'ютери, для створення, проведення та оцінки тестів; економія часу; здатність виміряти не тільки кінцеву відповідь, але й логічний процес, що призвів до неї.

Основні переваги тестів включають наукову обґрунтованість, точність, надійність, об'єктивність та можливість використання для контролю навчальних досягнень та перевірки відповідності стандартам.

Тести - це система тестових завдань різної складності, яка дозволяє ефективно вимірювати рівень та структуру підготовки учнів. Вилучення навіть одного завдання з тесту може призвести до необ'єктивності вимірювання навчальних досягнень.

Тести представляють собою комплекс завдань різної складності, який дозволяє об'єктивно та ефективно визначати рівень та структуру підготовки учнів, студентів та, взагалі, людей ,які виконують заданий комплекс завдань та прикладів [8].

Важливо звернути увагу на вираз «система завдань». Вилучення будь-якого або кількох завдань з тесту може призвести до необ'єктивності в оцінці навчальних досягнень учнів [5].

Тест повинен відповідати певним критеріям: мати валідність (вимірювати те, що він призначений вимірювати), точність (мінімізувати випадкові та систематичні похибки вимірювань), надійність (забезпечувати

стабільні та консистентні результати вимірювань незалежно від зовнішніх факторів, а також різних «нестандартних» випадків) і об'єктивність (гарантувати, що отримані результати відповідають реальності та є достовірними, точними).

Таблиця 1 – Загальна характеристика критеріїв тесту

Критерій	Зміст поняття	Показник
Об'єктивність	Мінімізація впливу суб'єктивних факторів	Не введений
Надійність	Ступінь стійкості результатів	Коефіцієнт надійності
Валідність	Відповідність результатів тому, що треба вимірювати	Коефіцієнт валідності
Точність	Мінімальна помилка	Коефіцієнт точності

Розглядаючи характеристики тесту, важливо підкреслити, що кількість завдань, які утворюють тест (його довжина), не може бути випадковою і визначається декількома факторами.

Перший фактор – це передбачена точність вимірювання. Якщо ви маєте намір вимірювати рівень знань більш точно, то тест має бути більшим за обсягом. Це є оптимальними показниками. Зрозуміло, що довжина тесту обмежується певними психофізіологічними характеристиками тестованої аудиторії.

Другий визначальний фактор, який визначає довжину тесту – це структурований зміст, який має бути включений і відображений у кількості тестових завдань. Матеріал, який піддається оцінці через тест та представлений у формі завдань, називається матрицею тесту.

Розглянемо детальніше деякі критерії оцінювання тесту: валідність та надійність.

Надійність тесту (ρ) зростає з підвищенням узгодженості результатів одного студента під час повторної перевірки його знань за допомогою того ж

тесту або еквівалентної його версії (паралельного тесту). Рівень узгодженості результатів можна виміряти за допомогою коефіцієнта кореляції Пірсона. Значення коефіцієнта у діапазоні 0,80-0,89 вказує на високу надійність тесту, а якщо воно 0,90 або вище, то надійність можна вважати дуже високою [7].

Оцінку надійності загального тесту можна проводити за допомогою коефіцієнта кореляції $r_{1/2}$, який розраховується за формулою Спірмана-Брауна [5]:

$$\rho = \frac{2r_{1/2}}{1 + r_{1/2}}$$

Ще один спосіб оцінки надійності розщепленого тесту базується на формулі Рюлона:

$$\rho = 1 - \frac{s_d^2}{s_y^2};$$

де s_y^2 – дисперсія сумарних балів результату;

s_d^2 – дисперсія різниць між результатами кожного студента по обох половинах тесту.

Розглядаючи внесок окремого завдання в надійність тесту та узгодженість його результатів з результатами усього тесту, можна визначити індекс надійності цього завдання (таблиця 2).

Таблиця 2 – Характеристика тесту за коефіцієнтом надійності

Інтервали коефіцієнта надійності	Характеристика тесту
0,90–0,99	Відмінний
0,85–0,89	Дуже хороший
0,80–0,84	Хороший
0,75–0,79	Задовільний
0,70–0,74	Мало задовільний

Валідність тесту відображає ступінь його ефективності у досягненні поставлених цілей. Це комплексна оцінка, яка відображає обґрунтованість, значущість та адекватність результатів тестування в контексті його призначення. Визначення коефіцієнта валідності тесту означає визначення ступеня відповідності результатів тесту іншим незалежним методам оцінки знань студентів. Для цього потрібен зовнішній незалежний критерій, який може бути представлений, наприклад, експертною оцінкою викладача. Коефіцієнт валідності визначається як коефіцієнт кореляції між результатами тестування і цим експертним критерієм. Якщо маємо числову послідовність оцінок експерта для студентів, отриманих незалежно від процедури тестування, як Y_1, Y_2, \dots, Y_n , то коефіцієнт валідності тесту може бути розрахований за наступною формулою:

$$V = \frac{\frac{\sum_{i=1}^n (Y_i \times y_i)}{n} - \bar{Y} \times \bar{y}}{s_Y \times s_y} \cdot \frac{n}{n-1};$$

де \bar{Y} – середнє арифметичне експертних оцінок;

s_Y – стандартне відхилення цих оцінок.

Валідність тесту визначається шляхом порівняння його результатів з показниками, отриманими незалежно від тестування, такими як традиційні оцінки, експертні висновки або результати інших тестів, які були раніше валідовані. Зазвичай вимірюється за допомогою коефіцієнтів кореляції. Отже, з двох тестів, спрямованих на досягнення одного й того ж результату, більш ефективним вважається той, який дозволяє швидше і точніше виміряти рівень знань. Під «швидкістю» в даному контексті розуміється кількість питань, необхідних для отримання достовірного уявлення про рівень знань студента.

Розуміємо, що чим вища валідність, тим більш обґрунтованим є використання результатів тестування для аналізу, висновків і прогнозів. Валідність тестів може бути поділена на різні види, включаючи критеріальну

валідність (яка відображає відповідність результатів тесту реальним досягненням рівня знань чи навичок), змістову валідність (яка показує, наскільки завдання відображають зміст певної предметної області і її важливість для даної діяльності), та конструктивну валідність (яка вказує на відповідність вимірюваного конструкту в теоретичній області знань).

Система тестування містить два види вимірювання, а відповідно і оцінювання (тестів): нормативно-орієнтоване та критеріально-орієнтоване (таблиця 3).

Таблиця 3 – Типи вимірювання

Критерії порівняння	Нормативно–орієнтоване	Критеріально–орієнтоване
Основна задача	розрізнити (диференціювати) екзаменованих між собою	встановити абсолютний рівень досягнень індивідуума (незалежний від інших)
Домен	великий; вимірюється в «ширину», інформація про загальні риси оволодіння великими доменами	вузький; вимірюється в «глибину», інформація про оволодіння конкретними елементами домену
Тестові завдання	середньої складності	складність відповідає складності оцінюваного домену
Результат	Відносний (порівняння із norm-group)	Абсолютний (порівняння з критерієм, стандартом)
Сфера використання	відбір, ранжування, порівняння оцінювання програм, порівняння освітніх інституцій	сертифікація, ліцензування, оцінка мінімальної компетентності

Таким чином, нормативно-орієнтоване оцінювання дозволяє ранжувати тих, кого випробують, за рівнем знань. Результати тестування за нормативно-орієнтованим тестом дозволяють порівняти навчальні досягнення экзаменованих один з одним. Якість такого тесту (як інструменту вимірювання) визначається за допомогою статистичних методів аналізу результатів його апробації [4].

Аналіз відповідності розробленого тесту вищезазначеним характеристикам проводиться через апробацію та статистичний аналіз отриманих результатів. Цей етап технологічного циклу стандартизованого тестування надзвичайно важливий з двох основних причин. По-перше, доки ми не переконаємося у якості інструменту вимірювання, неможливо робити висновки щодо отриманих результатів. По-друге, необхідно аналізувати компоненти тесту – тестові завдання – та відсікати ті, що не відповідають вимогам якості.

Під час оцінки якості інструменту вимірювання застосовуються аналітичні (експертні) та емпіричні (статистичні, математичні) методи. Загальну сукупність емпіричних методів для оцінки якості інструменту вимірювання називають психометричним аналізом. Серед таких методів включають: середнє (M), мінімальне та максимальне значення; середнє квадратичне відхилення σ ; коефіцієнт надійності R (α -Кронбаха або Кьюдер–Ричардсона); помилка вимірювання – E ; коефіцієнт валідності (якщо наявний); коефіцієнт асиметрії – A_x (асиметричність кривої); коефіцієнт ексцесу – E_x (зміна гостроти кривої у вершини).

Цей аналіз визначає відповідність тесту визначеним критеріям та допомагає вдосконалити тест для досягнення більш точних та надійних результатів.

Аналіз тестових завдань та тесту в цілому може включати такі етапи:

- а) створення та оцінка матриці результатів: Аналізуючи відповіді учасників, формується матриця результатів, яка використовується для подальшого дослідження;

- б) оцінка складності та дискримінативності завдань: визначення того, наскільки важко чи відмінно кожне завдання виявляється для учасників, що дозволяє зрозуміти його ефективність у вимірюванні знань;
- в) визначення варіації балів: розрахунок різниці між найвищим та найнижчим балом, що отриманим учасниками, для оцінки різноманітності результатів;
- г) розрахунок мір центральної тенденції: визначення середнього балу або медіани для отримання загального уявлення про типовий рівень відповідей;
- д) побудова графічних представлень розподілу балів: використання гістограм та полігонів частот для наглядного відображення розподілу результатів;
- е) оцінка надійності тесту: вивчення консистентності тесту для визначення його надійності та вірогідності вимірювань.

При цьому, послідовність проведення аналізу може бути різною. Основними психометричними властивостями тестового завдання є його рівень складності та здатність розрізняти між учасниками (індекс дискримінативності ID, point-biserial) [8].

Складність тестового завдання визначається тим, наскільки воно відповідає рівню знань та навичок конкретної групи учнів, їхнього рівня, класу, школи і т.д. Іншими словами, тестове завдання є складним або легким відносно тих, хто його виконує. Складність завдання визначається шляхом розрахунку відсотку учнів, які правильно відповіли на нього, і виражається в індексі складності, що може приймати значення від 0 до +1,0. Формула для розрахунку складності (P-value) виглядає наступним чином:

$$P_i = \frac{R_i}{N};$$

де R_i – кількість учнів, які виконали завдання правильно;

N – загальна кількість учнів.

Отже, чим вищий показник, тим легше тестове завдання. Наприклад, якщо 100% учнів відповіли правильно, то показник дорівнює 1.

Залежно від значення складності, тестове завдання класифікується як:

- надто складне, якщо $0,20 < P < 0,36$;
- середньої складності, якщо $0,36 < P < 0,84$;
- надто легке, якщо $P > 0,84$ [1].

Ця методика класифікації складності відповідає міжнародній практиці. Український центр оцінки якості освіти використовує наступні діапазони коефіцієнтів складності тестового завдання (таблиця 4).

Таблиця 4 – Діапазони рівнів складності тестових завдань

Інтервал індексу складності	Характеристика завдань
Понад 0,8	Дуже легкі
0,60–0,79	Легкі
0,40–0,59	Оптимальні
0,20–0,39	Складні
0,19 і нижче	Дуже складні

Ефективність тесту вважається оптимальною, коли він включає 20% завдань легкого рівня, 60% завдань середньої складності і 20% завдань складного та дуже складного рівня. Збільшення кількості завдань складного рівня за рахунок зменшення кількості легких та середнього рівня може призвести до зниження результатів вимірювання, і навпаки, зменшення кількості завдань складного рівня за рахунок збільшення кількості легких та середнього рівня може призвести до завищення результатів.

1.3 Класифікація тестів. Форми тестових завдань

Процедура тестування дозволяє студентові самостійно виявляти пропуски в структурі своїх знань, приймати заходи для їх ліквідації, черпати необхідну інформацію. У таких випадках виявляється значний навчальний потенціал тестових завдань, практична реалізація принципу єдності і взаємозв'язку навчання і контролю. Це дисциплінує, організовує і спрямовує діяльність студентів, формує прагнення розвинути свої здібності [5].

Тести, як правило, класифікують за цільовими, функціональними, смисловими та формальними ознаками. Наочно відобразимо це за допомогою таблиці 5.

Таблиця 5 – Класифікація дидактичних тестів

Критерії класифікації	Види тестів
а) цільові-функціональні та смислові ознаки ціль (мета застосування)	констатуючі, діагностуючі, прогностичні (прогнозуючі)
вид контролю	стартові (вхідні), поточні (проміжні), рубіжні (семестрові), підсумкові, заключні (вихідні)
об'єкт контролю	тести знань та навичок, тести інтелекту, тести інтересів, тести соціальних здібностей, характерологічні (особистісні) тести, тести визначення окремих психічних функцій
статус впровадження	загальнонаціональні тести, тести відомчі (міністерські), тести рівня навчального закладу, тести кафедральні, тести особисті (неформальні)
рівень уніфікованості	стандартизовані, не стандартизовані

Продовження таблиці 5

б) формальні ознаки співвідношення із нормами (критеріями)	тести, орієнтовані на норму; на групу; на критерій
однорідність тестових завдань	однорідні, неоднорідні
гомогенність завдань	на швидкість, на складність
форма тестування	тести для групового тестування, тести для індивідуального тестування
застосування технічних засобів	безмашинні тести (бланкові з ручною обробкою результатів); бланкові тести з комп'ютерною обробкою результатів; комп'ютерні тести
форма оцінки відповідей	одно канална (правильна відповідь плюс один бал, неправильна – нуль), двоканальна-ліберальна (всі відповіді правильні - один бал, деякі відповіді правильні – пів бала, всі неправильні – нуль балів), три канална (плюс один бал, 0 балів, мінус один бал) тощо.

Класифікація діагностичних тестів у різних країнах і наукових спільнотах розрізняє їх за найменуваннями, такими як тести навчальних досягнень, тести успішності, дидактичні тести, педагогічні тести, а навіть тести викладача. Форма тесту визначає його зовнішній вигляд. Основний поділ полягає в розрізненні тестів на відкритий і закритий типи.

Тести закритого типу є найпоширенішими. У них до кожного запитання наводиться кілька варіантів відповідей, з яких студент повинен вибрати правильний. Цей тип тестів є простішим у підготовці та використанні, оскільки основні зусилля студентів зосереджені на розв'язанні завдань, а не

на введенні відповідей у комп'ютер. У тестах закритого типу виділяють такі підтипи:

- дихотомічні тести;
- тести з множинним вибором;
- тести на відповідність;
- тести на відновлення послідовності.

Кожен з цих типів тестів має свою специфіку. Наприклад, тести з множинним вибором перевіряють розуміння математичних понять та тверджень, а також уміння аналізувати та вибирати альтернативи.

Відкриті тестові завдання визначаються відсутністю заздалегідь готових відповідей, які студент може вибрати. Замість цього, вони передбачають, що студент самостійно надасть відповідь у стислій та конкретній формі, наприклад, числове значення виразу. Завдання відкритого типу вважаються більш перевагою в порівнянні з завданнями закритого типу, основним чином через те, що вони сприяють мінімізації вірогідності вгадування під час їх виконання [10].

При створенні математичних тестів, особливо з таких дисциплін, як математичний аналіз та теорія ймовірностей, важливо переконатися, що перевірка набутих знань та навичок відбувається на належному рівні. Тестові завдання відкритого та закритого типів, хоч і корисні, можуть бути недостатніми для оцінки процесу розв'язання задачі.

У зв'язку з цим, кафедрою математичного аналізу та теорії ймовірностей було розпочато розробку так званих «покрокових тестів». Ці тести представляють собою послідовність завдань різних типів, які об'єднані в одну задачу з метою перевірки кожного кроку в процесі розв'язання та остаточної відповіді. Це дає можливість викладачам більш точно оцінювати рівень розуміння та засвоєння матеріалу студентами [12].

Такий підхід дозволяє не тільки перевіряти правильність відповіді, а й аналізувати логічні кроки та методи розв'язання, що застосовуються студентами. Він сприяє більш глибокому розумінню предмету та може бути

корисним інструментом для покращення якості освіти в математичних дисциплінах.

Завданням розробки тестів є створення інструментів для об'єктивної оцінки знань студентів. Тести можуть відрізнятися за різними параметрами, такими як їх орієнтованість та метод генерації питань.

Одним з параметрів, за якими розрізняють тести, є їх орієнтованість. Критеріально-орієнтований тест призначений для виміру рівня засвоєних студентом знань або відповідності його до певного критерію. Нормативно-орієнтовані тести, натомість, оцінюють досягнення студента у порівнянні з іншими студентами, а не просто його знаннями. Важливо зазначити, що обидва типи тестів можуть містити однакові завдання, і вони відрізняються лише в інтерпретації результатів [4].

Щодо методу генерації питань, існують два основних типи тестів. Сталі тести, або «авторські», складаються з фіксованого набору питань, які залишаються незмінними для кожного студента. Такі тести корисні для екзаменів і самоконтролю, і вони дозволяють оцінити рівень успішності групи в цілому.

Тематичні тести, або тести, що генеруються випадковим чином, формуються на основі заданої тематики і кількості питань для кожної теми. Питання генеруються автоматично з банку даних питань, і цей тип тесту забезпечує можливість створювати практично безліч різних варіантів тестів. Цей метод зручний для контрольних тестів та дозволяє автоматично створювати індивідуальні тести для кожного студента для перевірки його рівня знань.

Також варто відзначити адаптивні тести, які адаптуються до відповідей студента. Ефективне виконання завдання веде до пред'явлення більш важких завдань, а неефективне виконання – до менш важких. Такий підхід дозволяє більш точно визначити рівень знань студента і підбирати завдання, що відповідають його поточному рівню.

2 КОМП'ЮТЕРНЕ ТЕСТУВАННЯ, ЯК МЕТОД КОНТРОЛЮ ЯКОСТІ ЗНАНЬ З МАТЕМАТИЧНИХ НАУК

2.1 Комп'ютерне тестування знань, його переваги та недоліки

У сучасному інформаційному суспільстві важливо забезпечити якісну освіту при обмеженому часі для її засвоєння. Для вирішення цього завдання можна обрати один із найкращих методів: використання комп'ютерного тестування як складової педагогічних новацій. Сучасний стан розвитку інформаційних технологій та їх широке застосування в освіті призводить до переходу від бланкового до комп'ютерного тестування. Він є одним з найоптимальніших засобів контролю, який задовольняє вимоги щодо об'єктивності отриманої оцінки, якості процесу контролю. Це викликало підвищений інтерес до цього методу оцінки якості знань та в результаті спричинило його досить широке використання [11].

Потенційні переваги комп'ютерного тестування очевидні, оскільки вони дозволяють:

- генерувати індивідуальні завдання для кожного учасника, використовуючи випадковий вибір питань з банку завдань;
- зменшувати витрати часу та фінансові витрати на проведення тестування, особливо при великих обсягах, за рахунок автоматизованих процесів;
- дозволяти учням самостійно оцінювати свої знання, а викладачам - ефективно систематизувати та спростити процес оцінювання.;
- звільняти викладача від рутинної роботи під час проведення іспитів та проміжного контролю;
- автоматизувати обробку результатів для забезпечення об'єктивності та швидкості оцінювання;

- підвищувати гнучкість контролю знань за рахунок використання гнучкого графіку проведення тестування, включаючи індивідуальний графік;
- покращувати якість аналізу навчального процесу та надавати всебічний аналіз результатів тестування, включаючи хронологічні дані;
- централізовано проводити аналіз якості великої кількості питань для визначення найскладніших розділів та коригування процесу навчання;
- використовувати мультимедійні технології у тестах, що неможливо використовувати у письмовій формі, наприклад, технології перетягування «drag-and-drop» для операцій з графічними зображеннями;
- підвищувати рівень інформаційної безпеки тестових завдань [7].

В першу чергу ці переваги стосуються традиційного, очного навчального процесу, оскільки при використанні дистанційних технологій комп'ютерне тестування стає основним засобом контролю знань.

Проте, комп'ютерне тестування має свої недоліки:

- 1) Банк тестових завдань, з якого створюється тест, має містити від 100 до 300 питань.
- 2) В тесті рекомендується використовувати різні форми тестових завдань, такі як множинний вибір, встановлення відповідності або відновлення правильної послідовності, і відкрита форма відповіді. Бажано дотримуватися такого розподілу в тестуванні: 70% завдань мають бути у форматі множинного вибору, 20% – у форматі встановлення відповідності або відновлення правильної послідовності, і 10% – у формі відкритої відповіді.
- 3) Уникайте перевірки тривіальних або надмірно спеціалізованих знань.
- 4) Тестове завдання не повинно містити дискусійних питань або відповідей.

5) Кількість варіантів відповідей в тестовому завданні слід обмежувати 4-6.

6) У загальних налаштуваннях тесту рекомендовано використовувати випадковий порядок для розташування як питань, так і варіантів відповідей тесту.

7) Інформація в одному тестовому завданні не повинна розкривати відповідь на інше тестове завдання.

8) У загальних настройках тесту можна обмежити час на проходження тесту та кількість спроб.

9) Формулювання умови тестового завдання повинно бути лаконічним, зрозумілим та містити чітко визначене завдання.

10) Умови та варіанти відповідей бажано формулювати позитивно.

11) Структура тестового завдання повинна бути граматично коректною.

12) У тестовому завданні не повинно бути повторів одних і тих самих слів.

13) Відповіді повинні бути однорідними за змістом, а також короткими і простими за своєю структурою.

14) Формулювання правильної відповіді повинно бути аналогічним за структурними характеристиками до інших варіантів відповідей даного завдання.

15) Усі неправильні варіанти відповідей мають бути правдоподібними.

16) При оцінці відповідей на питання різної складності рекомендується використовувати диференційований підхід [13].

Ці принципи і правила допомагають забезпечити якість комп'ютерного тестування в умовах дистанційного навчання і зменшити ризики недостатньої підготовки викладачів у цьому плані.

2.2 Класифікація та основні вимоги до комп'ютерних систем тестування знань

На сучасний момент існує значна кількість систем комп'ютерного тестування, що вимагає проведення їх класифікації з урахуванням різних характеристик, принципів побудови, схем взаємодії з користувачем та областей застосування. Виокремлюються різні типи систем контролю знань, зокрема за можливістю оновлення бази питань і внесення змін до програмного забезпечення, за схемою взаємодії з користувачем, за середовищем застосування (локальні та глобальні) та за предметно-галузевою направленістю (профільні та універсальні) [2].

Розглянемо класифікацію комп'ютерних систем тестування знань.

1) За ступенем універсальності:

- Універсальні: використовують і для створення тестових завдань у різних сферах.
- Спеціалізовані: орієнтовані на конкретну галузь, наприклад, математику чи хімію.
- За рівнем автономності:
 - В складі LMS (Learning Management System) тобто систем управління навчанням, таких як Moodle або SharePoint Learning Kit.
 - Окремий модуль.
 - Комплексний набір модулів.

2) За складом модулів:

- Наявність окремих модулів для складання тестів і проведення тестування.
- Наявність спеціального модуля лише для складання тестів.

3) За характером проведення тестування:

- З використанням інструментальних оболонок, спрямованих саме на проведення тестування.

- Через браузер, що іноді відоме як WWW-тестування або Інтернет-тестування [14].

Такий підхід дозволяє враховувати різноманітні аспекти систем тестування та визначає їхню придатність для різних освітніх контекстів та завдань [14].

Оптимальна структура комп'ютерних систем тестування знань передбачає модульну побудову, що включає три основні компоненти::

- 1) Модуль для створення, підготовки та редагування тестових завдань.
- 2) Модуль для проведення тестових завдань, іноді відомий як «Player тестів».
- 3) Модуль для адміністрування [14].

Комп'ютерна система тестування знань має виконувати свою ключову функцію – забезпечення ефективного тестування знань, що включає різноманітні аспекти. Деякі з них стосуються системи в цілому, в той час як інші аспекти пов'язані з конкретними модулями. В умовах конкретного використання можуть бути встановлені специфічні вимоги до системи тестування. Тим не менше, можна виділити кілька основних вимог, які зазвичай має відповідати комп'ютерна система тестування знань.

- 1) Універсальність. Комп'ютерна система тестування знань повинна бути придатною як для внутрішнього, так і для зовнішнього тестування. Цю універсальність можна досягти шляхом гнучкої конфігурації системи, яка враховує відмінності між внутрішнім і зовнішнім тестуванням. Наприклад, система може розділяти користувачів на різні ролі з відповідними правами, підтримувати різні системи оцінювання та шкали. Для внутрішнього користувача, такого як «свій студент», можна використовувати систему оцінок, що відповідає навчальному плану, тоді як для зовнішнього користувача можна застосовувати іншу систему, наприклад, накопичувальну для оцінки успішності тестування.

- 2) Механізм керування користувачами з розподілом прав доступу: це аспект тісно пов'язаний із функцією універсальності. Комп'ютерна система тестування знань повинна підтримувати безпечний та універсальний механізм управління користувачами, який передбачає розподіл прав доступу до різних функцій та компонентів системи. Наприклад, можливі різні типи користувачів, такі як "свій" студент, студент, не зарахований у конкретний навчальний заклад, викладач, розробник тестів та адміністратор, кожен з яких має свої відповідні права доступу.
- 3) Захищеність інформаційної бази тестів та результатів тестування: лише користувачі з відповідними правами повинні мати доступ до інформаційної бази тестів і результатів тестування, забезпечуючи високий рівень захищеності цієї інформації.
- 4) Шифрування інформаційного вмісту тестів, особливо щодо відповідей: це особливо важливо для забезпечення конфіденційності та інформаційної безпеки. Шифрування дозволяє зберігати відповіді у безпечному форматі, недоступному без відповідного дозволу.
- 5) Підтримка основних форм тестових завдань: система повинна підтримувати різні форми тестових завдань, такі як завдання з простим і складним множинним вибором, відкритою формою, завдання на встановлення відповідності та на відновлення правильної послідовності.
- 6) Підтримка двох типів проведення тестування: система повинна мати можливість проводити два типи тестування згідно з їх функціональним призначенням. Перший тип тестування призначений для проведення соціологічних або психологічних тестів, де особа, яка тестується, відповідає на всі завдання тесту, а результати обробляються згідно з певними правилами для отримання певного висновку. Другий тип тестування призначений для проведення тестів знань або контрольних тестів. Тут особа

відповідає лише на заздалегідь визначену кількість випадково відібраних контрольних завдань в рамках певних розділів або тем, визначених на етапі налаштування системи..

- 7) Підтримка різних режимів тестування:
 - Екзаменаційний режим. У цьому режимі за невірні відповіді віднімаються бали, а окремі контрольні завдання дозволяється пропускати, за що бали не додаються і не віднімаються.
 - Режим самотестування або тренажера. Під час тестування виводяться повідомлення про помилки або рецензія на відповідь.
 - Режим навчання. Може містити вступ і пояснення до завдання, різноманітні навчальні матеріали у текстовому або мультимедійному вигляді [9].
- 8) Можливість імпорту тестів: система повинна підтримувати можливість імпорту тестів у текстовому форматі і форматі HTML. Також важливо, щоб була можливість експорту тестів з електронних текстових версій паперових тестів або з інших комп'ютерних систем тестування знань.
- 9) Наявність української локалізації. Наявність докладної документації українською мовою. Технічна підтримка комп'ютерної системи тестування знань.

2.3 Характеристика комп'ютерних систем тестування знань

Комп'ютерні системи тестування знань відкривають широкі можливості для персоналізації навчання, автоматизації процесів тестування та миттєвого збору результатів. Забезпечуючи користувачам можливість створення, редагування та адаптації тестів з легкістю, ці сервіси вирішують проблеми традиційних методів оцінювання, дозволяючи пристосовувати тести до індивідуальних потреб студентів.

Розглянемо характеристику он-лайн сервісів для створення тестів зробимо детальний огляд ключових можливостей, переваг та перспектив використання цих інструментів у сучасній освітній практиці.

1) Google Форми представляють собою частину офісного інструментарію Google Drive, призначеного для створення тестових форм. Для автоматичної перевірки та виставлення оцінок за визначеними критеріями є потреба у встановленні плагіну Flubaroo. Сервіс є безкоштовним, але для використання йому потрібен акаунт Google. Це один з найшвидших і простих способів створення власного опитування або тесту. Процес включає написання завдань та вибір типу відповідей, таких як вибір з кількох варіантів або написання власних. Отриманий тест можна зручно розповсюджувати серед учнів шляхом відправлення електронною поштою або вбудовувати його на власний веб-сайт за допомогою спеціального коду..

Переваги Google форм:

- Учні можуть проходити тестування он-лайн, просто перейшовши за посиланням.
 - Створені тести можна вбудовувати в блог або на сайт, відправляти по електронній пошті.
 - Є набір тем для оформлення тесту.
 - Можна зібрати певну статистику з відповідей слухачів.
 - Дають можливість спільного доступу для редагування тесту.
 - Існує можливість автоматичної оцінки відповідей, нарахування балів, коментарів до відповіді, відкладеного показу результатів
 - Доступне індивідуальне налаштування – показ питань на основі відповідей користувачів
 - Інтуїтивно зрозумілий інтерфейс, можливість копіювати питання з текстового редактора [15].
- 2) «MyTest» – це зручний та потужний набір програм для створення та проведення комп'ютерного тестування, а також для збору та аналізу

результатів. Складається з редактора тестів (MyTestEditor), модулю тестування (MyTestStudent) та журналу результатів (MyTestServer). Завдяки своїй простоті, цей комплекс ідеально підходить для використання в різних навчальних закладах. Хоча він в основному призначений для шкіл, важливо відзначити, що «MyTest» не має розширених можливостей масштабування. З іншого боку, він є дуже простим та ефективним програмним забезпеченням, яке працює ефективно навіть на обмежених обчислювальних ресурсах і дозволяє проводити тестування одночасно на кількох комп'ютерах [14].

- 3) Moodle – інноваційне об'єктно-орієнтоване навчальне середовище, що відоме як система управління навчанням, система управління курсами, або віртуальне навчальне середовище. Ця платформа для навчання надає широкий спектр інструментів для комп'ютеризованого навчання, зокрема дистанційного. Для створення тесту необхідно спочатку створити «Банк питань». Це включає введення всіх тестових запитань з навчальної дисципліни до системи, які потім можна використовувати для створення тестів. Для цього у меню слід вибрати блок «Керування» та пункт «Банк питань». Потім створити «Категорію», до якої будуть віднесені завдання (можна використовувати, наприклад, назву теми, до якої вони відносяться), та натискати кнопку «Створити нове питання» [10].

Цей розділ дозволяє вчителю створювати різнотипові тести, які включають такі види питань. Множинний вибір – питання, де можна обирати одну або декілька відповідей із заданого списку. Есе – дозволяє відповіді у вигляді декількох речень або абзаців, які оцінюються вчителем вручну. Перетягування в тексті – заповнення пропущених слів у тексті за допомогою перетягування. Коротка відповідь – вимагає введення відповіді одним словом з клавіатури. Числова відповідь – вимагає введення числової відповіді з клавіатури. Відповідність – встановлення відповідності між елементами. Правильно-неправильно –

вимагає вибору між двома варіантами відповіді на питання «правильно» чи «неправильно». Інші типи питань, які можуть бути додані. Це дозволяє створювати різноманітні тести для ефективної оцінки знань учнів.

Модуль «Тест» дозволяє викладачеві ефективно оцінювати відповіді великої кількості студентів, оскільки система автоматично аналізує їх відповіді та безпосередньо вносить їх в журнал оцінок.

- 4) Платформа Kahoot! дозволяє використовувати опитування та тести для подання навчального матеріалу. Для забезпечення зворотного зв'язку викладач може створювати ігрові теми з простими питаннями та відповідями, а також проводити більш детальне тестування. Спільна робота в класі забезпечується тим, що викладач відображає матеріал на головному екрані, а учні відповідають на питання та обговорюють інформацію через спеціальний клієнт на комп'ютерах або браузерах на смартфонах (Android, iOS, Windows Phone). Для участі в віртуальному класі студенти використовують унікальний код, наданий викладачем. Платформа дозволяє відстежувати відповіді кожного студента та створювати діаграми успішності для всієї групи. Студенти також можуть слідкувати за своїми результатами у спеціальних таблицях. Kahoot! є безкоштовним та доступним після реєстрації [11].
- 5) CLASSMARKER – це сервіс для швидкого створення тестових завдань та опитувань англійською мовою з широким спектром форматів відповідей. Педагог може особисто створювати та редагувати тести, зберігаючи їх у своєму профілі. Для формування класу вчитель може використовувати розсилку реєстраційних даних учнів на їхні електронні адреси або надсилати реєстраційні коди.
- 6) Proprofs – цей інструмент надає можливість створювати та оприлюднювати освітні курси, додавати опитування, формувати тести з різними варіантами відповідей або з розгорнутою відповіддю. Також можна вставляти в завдання зображення або відео файли.

7) Майстер-Тест це безкоштовний інтернет-сервіс, який дозволяє створювати тести (як онлайн, так і скачати для використання без підключення до інтернету).

Крім розглянутих вище, існує ще дуже багато різних сервісів для проведення комп'ютерного тестування знань.

3 СТАТИСТИЧНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

3.1 Класичні статистичні методи аналізу результатів тестування

Класична методика оцінки тестів ґрунтується на використанні статистичних методів для аналізу отриманих результатів тестування. Для статистичного опрацювання результатів тестування знань у своїй роботі я розглянула найпростіші і найнеобхідніші методи оцінки якості тесту, які найчастіше використовують.

Кількісну, тобто числову оцінку успішності виконання j -ого завдання i -им студентом позначимо через x_{ij} . Результати тестування звичайно для проведення статистичного аналізу представляють у вигляді матриці $\{x_{ij}\}$ з n рядками та m стовпцями ($i = 1, \dots, n; j = 1, \dots, m$). Дана матриця результатів показує результат, який отримали учасники при виконанні всіх завдань тесту. На практиці прийнято, найчастіше, застосовувати дихотомічну шкалу оцінок результатів. Якщо студент правильно виконав завдання, то він отримує один бал $x_{ij} = 1$, в протилежному випадку – нуль балів $x_{ij} = 0$. І тоді бал виражає кількість правильно виконаних завдань, але результат можна оцінювати тільки нулем чи одиницею, але й присвоювати певний коефіцієнт, який буде відповідати складності завдання [3].

Розглянемо детальніше, по кроках процес статистичної обробки матриці результатів тестування, яка використовується в даній роботі.

На першому кроці потрібно сформулювати таблицю з результатами тестування. Для цього необхідно обчислити індивідуальні бали всіх студентів, які проходять тест $x_i, i = \overline{1, N}$, (N – кількість людей, що проходять тест) та загальні бали за кожне завдання $R_j, j = \overline{1, K}$, (K – кількість завдань, які входять до тесту).

Обчислимо індивідуальні бали іспитників (результат тестування кожного):

$$y_i = \sum_{j=1}^m x_{ij}.$$

Для перевірки статистичних гіпотез, які використовуються в класичній теорії тестів, зазвичай припускається, що сумарні бали студентів мають нормальний розподіл. Це передбачення базується на центральній граничній теоремі, яка стверджує, що сума незалежних випадкових величин зазвичай наближається до нормального розподілу з ростом об'єму вибірки [6].

Проте, якщо дані не мають нормального розподілу, рекомендується досліджувати розподіл частот. Це означає аналізувати кількість випадків, коли значення попередньо визначеної ознаки попадають в певний інтервал. Для порівняння такого розподілу балів з нормальним можна використовувати різні статистичні критерії, які зазвичай використовуються для цієї мети.

Другим кроком обчислюються середні результати \bar{y} сумарних балів студентів, та середні результати \bar{x}_j для кожного завдання j .

Для дихотомічних даних, які можуть бути лише двох видів (наприклад, правильна або неправильна відповідь), традиційно в тестології використовуються показники, позначені як p_j , які відображають міру складності завдання j . Ці показники вказують, який відсоток студентів відповів правильно на дане завдання [9]. Вони можуть бути обчислені за формулою:

$$\bar{p}_j = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Показник p_j допомагає визначити, наскільки складним та важким є конкретне завдання і який процент студентів зміг на нього правильно відповісти.

Отже, середні результати у вказують на середній бал студентів у тесті в цілому, а p_j вказує на міру складності окремих завдань. Ці показники допомагають оцінити якість тесту та ефективність навчання.

На третьому кроці знаходимо дисперсію (s_y^2) та стандартне відхилення (s_y) сумарних балів студентів за формулами:

$$s_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \cdot \frac{1}{n-1};$$

$$s_y = \sqrt{s_y^2}.$$

Наступним, четвертим кроком шукаємо дисперсію по кожному j -тому завданню ($j = 1, 2, \dots, m$). Якщо завдання оцінюється лише балами 0 та 1 та застосовуємо формулу:

$$s_j^2 = p_j \cdot (1 - p_j).$$

Коли при оцінюванні завдання використовується множина оцінок, яка має більше, ніж два значення (три ,або більше), застосовуємо таку формулу для знаходження дисперсії:

$$s_j^2 = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \cdot \tau \frac{1}{n-1}.$$

Обчислимо також стандартне відхилення:

$$s_j = \sqrt{s_j^2}.$$

На п'ятому кроці потрібно зробити оцінювання надійності тесту, для цього обчислимо коефіцієнт надійності Кронбаха за формулою:

$$\alpha = \frac{K}{K-1} \cdot \left(1 - \frac{\sum_{j=1}^l s_j^2}{s_y^2} \right);$$

де, K – кількість іспитників,

s_j^2 – вибіркова дисперсія по розподілу балів за j -те завдання;

s_y^2 – вибіркова дисперсія по розподілу індивідуальних балів.

Даний коефіцієнт визначає, наскільки кожне питання або завдання в тесті співвідноситься зі загальним результатом та виявляє ступінь взаємозв'язку між окремими питаннями.

Якщо значення коефіцієнта дорівнює або перевищує 0,7 ($\alpha \geq 0,7$), це свідчить про високу внутрішню узгодженість тесту та його надійність. Узгодженість тесту є важливою ознакою для того, щоб він міг відповідати меті та бути валідним.

Шостим кроком обчислимо індекс I_j ($j = 1, 2, \dots, m$) дискримінативності завдання, тобто міру, яка використовується для оцінки того, наскільки ефективно конкретне завдання в тесті розділяє студентів за їхніми результатами. Цей індекс допомагає оцінити, наскільки завдання відрізняється за відповідями студентів з різними результатами та визначити, чи воно варте включення до тесту. Важливо забезпечити, щоб завдання були достатньо дискримінативними для ефективної оцінки студентів. З загальної сукупності студентів виокремлюють дві групи: тих, що отримали найвищі сумарні бали, і тих, що отримали найнижчі [5].

Наприклад, розподіляючи упорядковану сукупність сумарних балів на три частини, порівнюють результати виконання кожного завдання j серед студентів із першою й останньою третиною [5]. У випадку дихотомічних даних індекс має такий вигляд:

$$I_j = \frac{\sum_{i=1}^{n/3} x_{ij} - \sum_{i=2n/3+1}^n x_{ij}}{n/3}.$$

На наступному сьомому кроці знаходимо коефіцієнт кореляції Пірсона:

$$R_j = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_{ij} - y_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (x_{ij} - y_i)^2}};$$

де x_{ij} – бал, який отримав i -тий студент за j -те завдання;

y_i - середній бал за j -те завдання;

\bar{x} – вибіркове середнє;

x_i - кількість іспитників.

Тест має високу роздільну здатність тоді, коли $R_j \geq 0,7$. Якщо значення $R_j < 0,3$ це означає, що завдання потребує доопрацювання. Якщо ж R_j має від'ємне значення, то це вказує на те, що завдання складено неправильно і його потрібно виключити з тесту.

Після розрахунку всіх необхідних характеристик необхідно обчислити кореляційну матрицю завдань та провести її аналіз. З метою уникнення повторень тестові завдання не повинні мати високий рівень кореляції між собою. Якщо виявляється, що певне завдання має від'ємну кореляцію з більшістю інших завдань, то рекомендується провести заміну цих завдань [13].

3.2 Структура тестової контрольної роботи, що аналізується

Для проведення статистичного аналізу було обрано тематичний тест з предмету лінійна алгебра на тему «Матриця», який проводився у студентів першого курсу у 2023/2024 н. р. за допомогою он-лайн сервісу Google Forms.

У тестування взяли участь студенти групи 6.1212-пі-2 1 курсу математичного факультету освітнього рівня бакалавр, спеціальність 121 інженерія програмного забезпечення Запорізького національного університету. Тест на тему «Матриця» складається з п'ятнадцяти завдань, кожне з яких оцінюється у один бал.

Тематичний тест на тему «Матриця»:

1. Матриця – це:

- А) множина чисел, яка після певних обчислень дорівнює одному числу;
- Б) прямокутний масив чисел, який завжди містить n рядків та n стовпців;
- В) прямокутний масив чисел, який містить n рядків та m стовпців;
- Г) не має вірної відповіді.

2. Одинична матриця – це матриця:

- А) довільного розміру з елементами, що дорівнюють одиниці;
- Б) квадратна з елементами, що дорівнюють одиниці;
- В) квадратна з одиницями на головній діагоналі та нульовими усіма іншими елементами;
- Г) квадратна з нулями на головній діагоналі та одиницями усіма іншими елементами

3. Визначник $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ дорівнює:

- А) $ad + bc$;
- Б) $bc - ad$;
- В) $ac - bd$;

Г) $ad - bc$.

4. Визначник дорівнює нулю:

А) якщо всі елементи головної діагоналі дорівнюють нулю

Б) якщо до будь-якого рядка або стовпчика додати інший рядок або стовпчик, помножений на деяке число

В) якщо він має два однакових рядки або стовпці

Г) якщо елементи будь-якого рядка або стовпчика рівні

5. Рангом матриці називається:

А) сума кількості рядків та стовпців;

Б) число, що дорівнює найбільшому порядку мінору;

В) число, що дорівнює найбільшому порядку ненульового мінору;

Г) добуток кількості рядків та стовпців.

6. При транспонуванні матриці:

А) ранг не змінюється;

Б) ранг буде дорівнювати нулю;

В) ранг збільшиться на одиницю;

Г) зменшиться на одиницю.

7. Обернена матриця існує для:

А) довільної матриці;

Б) довільної квадратної матриці;

В) довільної квадратної не виродженої матриці;

Г) нульової матриці.

8. Яка з наведених нижче операцій над матрицями не виконується:

А) віднімання ;

Б) множення на число;

В) ділення на число;

Г) ділення.

9. При множенні двох матриць дотримуємося умови:

А) кількість рядків першого множника дорівнює кількості стовпців другого множника;

Б) кількість стовпців першої матриці дорівнює кількості рядків другої матриці;

В) кількість стовпців першої матриці дорівнює кількості стовпців другої матриці;

Г) кількість стовпців обох матриць додається.

10. Яка з матриць є виродженою (особливою)?

А) $\begin{bmatrix} 3 & 2 \\ -6 & -4 \end{bmatrix}$;

Б) $\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$;

В) $\begin{bmatrix} 3 & -2 \\ 5 & 4 \end{bmatrix}$;

Г) $\begin{bmatrix} 3 & -2 \\ 1 & 4 \end{bmatrix}$.

11. Застосовуючи властивості визначників і не обчислюючи їх, встановити, який з наступних визначників дорівнює визначнику

$$\begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 8 \\ 0 & 7 & -5 \end{vmatrix}?$$

А) $\begin{vmatrix} -1 & 2 & -3 \\ -1 & 3 & 5 \\ 3 & 1 & -4 \end{vmatrix}$;

Б) $\begin{vmatrix} 1 & -2 & 3 \\ -1 & 3 & 5 \\ 3 & 2 & 4 \end{vmatrix}$;

В) $\begin{vmatrix} 1 & -2 & 3 \\ -1 & 2 & 5 \\ 3 & 1 & 4 \end{vmatrix}$;

Г) $\begin{vmatrix} 1 & -2 & 3 \\ -1 & 3 & 5 \\ 3 & 1 & 4 \end{vmatrix}$.

12. Сумою яких двох матриць є матриця $\begin{bmatrix} 8 & 1 \\ 7 & -4 \end{bmatrix}$?

А) $\begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ 4 & -5 \end{bmatrix}$;

Б) $\begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ -7 & 4 \end{bmatrix}$;

$$\text{В)} \begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ -4 & 3 \end{bmatrix};$$

$$\text{Г)} \begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} + \begin{bmatrix} 3 & -2 \\ 5 & 3 \end{bmatrix}.$$

13. Різницею яких двох матриць є матриця $\begin{bmatrix} 2 & 5 \\ -3 & -11 \end{bmatrix}$?

$$\text{А)} \begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} - \begin{bmatrix} 1 & -21 \\ 3 & 40 \end{bmatrix};$$

$$\text{Б)} \begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} - \begin{bmatrix} 3 & -2 \\ 5 & 4 \end{bmatrix};$$

$$\text{В)} \begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} - \begin{bmatrix} 2 & -1 \\ 5 & 4 \\ 1 & 0 \end{bmatrix};$$

$$\text{Г)} \begin{bmatrix} 5 & 3 \\ 2 & -7 \end{bmatrix} - \begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix}.$$

14. Добутком яких двох матриць є матриця $\begin{bmatrix} 3 & 2 \\ -5 & 6 \end{bmatrix}$?

$$\text{А)} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 4 & 3 \end{bmatrix};$$

$$\text{Б)} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 4 & 3 \end{bmatrix};$$

$$\text{В)} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix};$$

$$\text{Г)} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 3 \\ 2 & 3 & 0 \end{bmatrix}.$$

15. Добутком яких двох матриць є матриця $\begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 0 & 2 \end{bmatrix}$?

$$\text{А)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix};$$

$$\text{Б)} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix};$$

$$\text{В)} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix};$$

$$\text{Г)} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}.$$

Таблиця 6 – Відповіді до тесту «Матриці»

№ завдання	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Правильна відповідь	В	В	В	Г	В	А	В	Г	Б	А	Г	Г	Б	В	Б
Кількість балів	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

3.3 Аналіз якості тесту в цілому

Усього у групі 6.1212-пі-2 1 курсу математичного факультету освітнього рівня бакалавр, спеціальність 121 інженерія програмного забезпечення Запорізького національного університету навчається 26 студенти, а кожне завдання, студент міг отримати максимум 1 бал. Відповідно, максимальний загальний бал за тест, коли вірно відповіли на всі питання, складає 15 балів.

Застосуємо розроблену методику статистичного аналізу результатів якості тестів на проведеному тематичному тесті з предмету «Лінійна алгебра» на тему «Матриці» в цілому. На першому кроці сформовано таблицю результатів яку представлено на рисунку 1.

Зразу ж у сформованій таблиці, знайдемо вибіркоче середнє \bar{y} сумарних балів студентів, та середні результати \bar{x}_j для кожного завдання j (рис. 1).

№	П.І.Б.	Результати тесту															Оцінка
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Белоусова Валерія Вячеславівна	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	13
2	Бондаренко Андрій Максимович	1	0	1	1	0	0	1	0	1	1	1	1	0	0	1	9
3	Васильєв Данило Євгенійович	0	1	1	0	1	1	0	0	1	0	1	1	1	0	0	8
4	Гуржій Данило Борисович	0	0	0	1	1	1	0	0	0	1	0	1	0	1	0	6
5	Дадикіна Олександра Костянтинівна	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	14
6	Демешко Євгеній Андрійович	0	0	0	1	1	0	1	1	1	1	1	1	0	1	1	10
7	Дзюман Максим Андрійович	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	10
8	Зиков Олександр Вікторович	1	1	0	0	0	0	0	0	0	0	1	0	0	1	1	5
9	Капінос Альона Олександрівна	0	1	1	1	0	1	0	1	1	1	0	1	0	0	0	8
10	Костюк Андрій Владиславович	0	0	1	0	0	1	1	0	0	0	0	0	1	1	0	5
11	Лаврусенко Андрій Костянтинівич	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15
12	Логвиненко Поліна Олексіївна	1	0	1	1	1	0	1	1	1	1	1	1	0	1	0	11
13	Мацегора Арсеній Андрійович	0	1	0	1	1	1	1	1	0	1	0	0	1	1	1	10
14	Назаренко Максим Олександрович	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	13
15	Ніколаєнко Нікіта Олексійович	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	12
16	Новіков Антон Володимирович	1	0	0	0	1	1	1	0	0	1	0	1	0	1	1	8
17	Панасенко Віталій Євгенійович	0	1	1	1	1	1	0	1	1	1	1	1	0	0	0	10
18	Пилипенко Артемій Валерійович	0	0	0	1	1	1	0	0	0	1	0	1	0	1	0	6
19	Пушка Дмитро Вікторович	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15
20	Радомський Дмитро Миколайович	0	0	0	1	1	0	1	1	1	1	1	1	0	1	1	10
21	Сирота Максим Сергійович	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	14
22	Сінельніков Матвій Миколайович	0	1	1	1	1	1	0	1	1	1	1	0	0	1	1	11
23	Трушковська Ольга Юріївна	1	1	0	1	0	0	0	0	1	0	1	0	0	1	0	6
24	Фесенко Владислав Юрійович	1	0	0	0	1	0	0	0	1	0	1	1	0	0	0	5
25	Шаповал Віталій Олександрович	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	13
26	Ямацита Рікаєнжела Юкінобівна	1	0	0	0	1	1	0	1	0	1	0	0	1	1	1	8
	R_x	16	16	15	19	20	18	16	16	18	20	18	18	13	19	13	255
	Вибіркове середнє, \bar{x}	0,6	0,6	0,6	0,7	0,8	0,7	0,6	0,6	0,7	0,8	0,7	0,7	0,5	0,7	0,5	9,80769

Рисунок 1 – Таблиця результатів тестування

На основі даних зробимо аналіз частотного розподілу вибірки результатів тестування (рис. 2).

Бали	[0;1]	(1;2]	(2;3]	(3;4]	(4;5]	(5;6]	(6;7]	(7;8]	(8;9]	(9;10]	(10;11]	(11;12]	(12;13]	(13;14]	(14;15]
Частота	0	0	0	0	3	3	0	4	1	5	2	1	3	2	2

Рисунок 2 – Частотний розподіл індивідуальних балів

За допомогою даних таблиці частот побудуємо гістограму частотного розподілу індивідуальних балів (рис. 3)

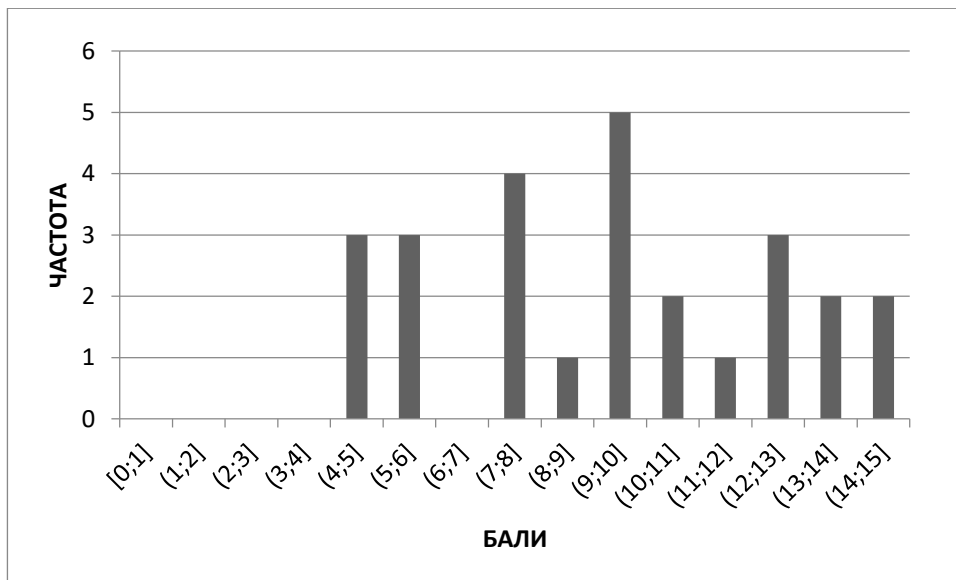


Рисунок 3 – Гістограма частотного розподілу

Після того, як побудували гістограму частот визначимо міри центральних тенденцій та міри мінливості. Зробимо це за допомогою MS Excel, результати обчислень надано на рисунку 4.

Тест вважається правильно складеним, якщо він є унімодальним і значення моди близьке до вибіркового середнього. У нашому випадку мода частотного розподілу становить 10. Це спільно з інтервальним частотним розподілом і гістограмою свідчить про унімодальність розподілу балів.

Вибіркове середнє частотного розподілу балів дорівнює – 9.8, значення медіани – 10 та значення коефіцієнта асиметрії, яке дорівнює – 0,02 при від’ємному ексцесі, який дорівнює – (-1,08) вказує на середню складність тесту, а також на те що даний тест має хорошу розподільну здатність.

Вибіркове середнє частотного розподілу балів, яке є близьким до значення медіани та значення коефіцієнта асиметрії, яке близьке до нуля вказує на великий відсоток високих балів. Тобто, для рівня підготовленості студентів, які писали тест, він виявився не складним.

Емпіричні дані		Додаткові розрахунки			
i	x_i	$(x_i - x_j)$	$(x_i - x_j)^2$	$(x_i - x_j)^3$	$(x_i - x_j)^4$
1	13	3,192307692	10,1908284	32,5322599	103,8529835
2	9	-0,807692308	0,652366864	-0,526911698	0,425582525
3	8	-1,807692308	3,267751479	-5,907089213	10,67819973
4	6	-3,807692308	14,49852071	-55,20590578	210,2071028
5	14	4,192307692	17,57544379	73,68166818	308,8962243
6	10	0,192307692	0,036982249	0,007111971	0,001367687
7	10	0,192307692	0,036982249	0,007111971	0,001367687
8	5	-4,807692308	23,11390533	-111,1245448	534,2526194
9	8	-1,807692308	3,267751479	-5,907089213	10,67819973
10	5	-4,807692308	23,11390533	-111,1245448	534,2526194
11	15	5,192307692	26,96005917	139,9849226	726,8447905
12	11	1,192307692	1,421597633	1,694981793	2,020939831
13	10	0,192307692	0,036982249	0,007111971	0,001367687
14	13	3,192307692	10,1908284	32,5322599	103,8529835
15	12	2,192307692	4,806213018	10,53669777	23,09968357
16	8	-1,807692308	3,267751479	-5,907089213	10,67819973
17	10	0,192307692	0,036982249	0,007111971	0,001367687
18	6	-3,807692308	14,49852071	-55,20590578	210,2071028
19	15	5,192307692	26,96005917	139,9849226	726,8447905
20	10	0,192307692	0,036982249	0,007111971	0,001367687
21	14	4,192307692	17,57544379	73,68166818	308,8962243
22	11	1,192307692	1,421597633	1,694981793	2,020939831
23	6	-3,807692308	14,49852071	-55,20590578	210,2071028
24	5	-4,807692308	23,11390533	-111,1245448	534,2526194
25	13	3,192307692	10,1908284	32,5322599	103,8529835
26	8	-1,807692308	3,267751479	-5,907089213	10,67819973
	255	0,00	254,0384615	15,74556213	4686,70693
МЦТ					
Обсяг вибірки	26				
Середнє (\bar{x}_j)	9,807692308				
Мода(M_o)	10				
Медіана(M_e)	10				
ММ					
Дисперсія	10,16153846				
Ст. відхилення	3,18771681				
Асиметрія	0,021064011				
Екссес	-1,087829729				

Рисунок 4 – Результати розрахунку МЦТ та ММ

Наступним кроком, знайдено коефіцієнт надійності Кронбаха (альфу Кронбаха), за допомогою двофакторного дисперсійного аналізу з пакету «аналіз даних» MS Excel. Результати аналізу подано на (рис. 5-6).

Двофакторний дисперсійний аналіз				
Підсумки	Счет	Сумма	Середнє	Дисперсія
Строка 1	15	13	0,866666667	0,123809524
Строка 2	15	9	0,6	0,257142857
Строка 3	15	8	0,533333333	0,266666667
Строка 4	15	6	0,4	0,257142857
Строка 5	15	14	0,933333333	0,066666667
Строка 6	15	10	0,666666667	0,238095238
Строка 7	15	10	0,666666667	0,238095238
Строка 8	15	5	0,333333333	0,238095238
Строка 9	15	8	0,533333333	0,266666667
Строка 10	15	5	0,333333333	0,238095238
Строка 11	15	15	1	0
Строка 12	15	11	0,733333333	0,20952381
Строка 13	15	10	0,666666667	0,238095238
Строка 14	15	13	0,866666667	0,123809524
Строка 15	15	12	0,8	0,171428571
Строка 16	15	8	0,533333333	0,266666667
Строка 17	15	10	0,666666667	0,238095238
Строка 18	15	6	0,4	0,257142857
Строка 19	15	15	1	0
Строка 20	15	10	0,666666667	0,238095238
Строка 21	15	14	0,933333333	0,066666667
Строка 22	15	11	0,733333333	0,20952381
Строка 23	15	6	0,4	0,257142857
Строка 24	15	5	0,333333333	0,238095238
Строка 25	15	13	0,866666667	0,123809524
Строка 26	15	8	0,533333333	0,266666667
Столбец 1	26	16	0,615384615	0,246153846
Столбец 2	26	16	0,615384615	0,246153846
Столбец 3	26	15	0,576923077	0,253846154
Столбец 4	26	19	0,730769231	0,204615385
Столбец 5	26	20	0,769230769	0,184615385
Столбец 6	26	18	0,692307692	0,221538462
Столбец 7	26	16	0,615384615	0,246153846
Столбец 8	26	16	0,615384615	0,246153846
Столбец 9	26	18	0,692307692	0,221538462
Столбец 10	26	20	0,769230769	0,184615385
Столбец 11	26	18	0,692307692	0,221538462
Столбец 12	26	18	0,692307692	0,221538462
Столбец 13	26	13	0,5	0,26
Столбец 14	26	19	0,730769231	0,204615385
Столбец 15	26	13	0,5	0,26

Рисунок 5 – Дані двофакторного дисперсійного аналізу

Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Рядки	16,93589744	25	0,677435897	3,454239821	1,29105E-07	1,53760167
Стовпці	2,692307692	14	0,192307692	0,980575271	0,4726798	1,720112432
Похибка	68,64102564	350	0,196117216			
Разом	88,26923077	389				
Коефіцієнт Кронбаха	0,710500703					

Рисунок 6 – Результати двофакторного дисперсійного аналізу

Коефіцієнт надійності Кронбаха $\alpha = 0,71$ є трохи більшим за нижню межу 0,7. Це вказує на достатню якість та надійність тесту.

На наступному кроці обчислимо індекс дискримінативності завдання, тобто міру, яка використовується для оцінки того, наскільки ефективно конкретне завдання в тесті розподіляє студентів за їхніми результатами.

Для цього розподілимо студентів групи, які проходили тестування на три групи у порядку зростання отриманих балів, та візьмемо результати двох груп: перша група (8 студентів) – ті студенти які отримали найбільший бал за тест та третя група (8 студентів) – ті студенти які отримали найменший бал за тест. Знаходити його будемо за формулою:

$$I_j = \frac{\sum_{i=1}^{n/3} x_{ij} - \sum_{i=2n/3+1}^n x_{ij}}{n/3}$$

За допомогою MS Excel проведено розрахунки результати яких подано на рисунку 8.

Для інтерпретації значення коефіцієнта дискримінативності застосовують критерії подані на рисунку 7.

Значення коефіцієнта	Інтерпретація
від 0,3 до 1,0	висока дискримінативність тесту
від 0,1 до 0,3	завдання потрібно проаналізувати на придатність до використання в тесті (низька диференціальна здатність)
< 0,1	завдання неякісне – краща група відповідає гірше, ніж слабша

Рисунок 7 – критерії оцінювання дискримінативності

З даних, поданих на рисунку 8, видно, що всі питання тесту, окрім питань під номерами 6, 12, 13, мають високу розподільчу здатність студентів. Але питання тесту під номерами 6, 12, 13 потрібно замінити, або доопрацювати їх умови.

№	П.І.Б.	Результати тесту															Оцінка
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
6	Демешко Євгеній Андрійович	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15
11	Лаврусенко Андрій Костянтинівич	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15
3	Васильєв Данило Євгенійович	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	14
12	Логвиненко Поліна Олексіївна	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	14
1	Белюсова Валерія Вячеславівна	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	13
8	Зиков Олександр Вікторович	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	13
15	Ніколаско Нікіта Олексійович	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	13
9	Капінос Альона Олександрівна	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	12
...	
17	Панасенко Віталій Євгенійович	0	1	1	0	1	1	0	0	1	0	1	1	1	0	0	8
20	Радомський Дмитро Миколайович	0	1	1	1	0	1	0	1	1	1	0	1	0	0	0	8
13	Мацегора Арсеній Андрійович	1	1	0	1	0	0	0	0	1	0	1	0	0	1	0	6
18	Пилипенко Артемій Валерійович	0	0	0	1	1	1	0	0	0	1	0	1	0	1	0	6
24	Фесенко Владислав Юрійович	0	0	0	1	1	1	0	0	0	1	0	1	0	1	0	6
5	Дадикіна Олександра Костянтинівна	1	1	0	0	0	0	0	0	0	0	1	0	0	1	1	5
14	Назаренко Максим Олександрович	1	0	0	0	1	0	0	0	1	0	1	1	0	0	0	5
21	Сирота Максим Сергійович	0	0	1	0	0	1	1	0	0	0	0	0	1	1	0	5
	Кількість випробуваних в групі студентів, що набрали найвищий бал, що вірно виконали дане завдання	8	8	7	7	7	7	8	8	7	7	8	7	8	7	5	109
	Кількість випробуваних в групі студентів, що набрали найнижчий бал, що вірно виконали дане завдання	3	4	3	4	4	5	1	1	4	3	4	5	2	5	1	49
	Індекс дискримінативності завдання	0,63	0,5	0,5	0,38	0,38	0,25	0,88	0,88	0,38	0,5	0,5	0,25	0,75	0,25	0,5	7,5

Рисунок 8 – Індекс дискримінативності завдання

Розглянемо знайдені коефіцієнти кореляції Пірсона (рис. 9) окремо для кожного завдання тесту, які надано на рисунку 9 і які знайдено за формулою:

$$R_j = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_{ij} - y_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (x_{ij} - y_i)^2}}$$

№ завдання	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Коеф. кореляції Пірсона	0,36	0,51	0,52	0,46	0,38	0,28	0,63	0,76	0,49	0,49	0,47	0,28	0,55	0,18	0,28

Рисунок 9 – Коефіцієнт кореляції Пірсона

Вказані значення коефіцієнтів кореляції Пірсона, другого, третього, сьомого, восьмого та тринадцятого завдання мають значний кореляційний зв'язок. І навпаки чотири завдання (6, 12, 14 та 15) мають дуже слабкий зв'язок, останні шість завдань мають помірний кореляційний зв'язок, тому можна зробити висновок про неправильність та низьку розрізняльну здатність складених завдань.

Таблиця 7 – Характер та глибина кореляційного зв'язку

r	Глибина зв'язку	r	Глибина зв'язку
r = 0	відсутня	0,7 < r < 0,9	сильна
0 < r < 0,3	слабка	0,9 < r < 1	дуже сильна
0,3 < r < 0,5	помірна	r = 1	повна
0,5 < r < 0,7	значна		

Проаналізуємо кореляційну матрицю (рис. 10). Як бачимо, дуже багато коефіцієнтів кореляції близькі до нуля та мають від'ємне значення, але також є завдання, які мають гарний кореляційний зв'язок (вище за 0,3) це вказує на присутність помилок яких допустили під час складання завдань.

В цьому аспекті важливо відзначити, що особливість створення тестів подібного типу полягає в оцінці здатностей та навичок студентів у вирішенні «стандартних» завдань, які вони повинні опанувати після завершення конкретного модулю. В межах цього модулю великий відсоток таких «стандартних» задач відзначається схожістю у методах їх вирішення.

Тому, при оцінці якості тесту важливо дбати про обережний підхід до аналізу значень кореляційної матриці завдань тесту. У висновках та рекомендаціях щодо удосконалення тесту слід урахувувати ці особливості. Зважаючи на те, що це перший статистичний аналіз тестів такого типу, варто провести додаткові дослідження, зокрема на групах студентів із різним рівнем підготовки [14]. Це допоможе розширити і коригувати зроблені висновки, а також розробити підходи та методики аналізу.

№ завдання	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1														
2	0,1875	1													
3	0,123091	0,443129	1												
4	-0,12339	0,233074	0,182262	1											
5	-0,05774	-0,05774	0,08528	0,079156	1										
6	-0,18447	0,329404	0,272475	-0,0289	0,228218	1									
7	0,35	0,025	0,28311	0,233074	0,129904	-0,01318	1								
8	0,025	0,35	0,28311	0,411306	0,317543	0,158114	0,35	1							
9	0,158114	0,329404	0,441149	0,346844	0,030429	-0,08333	0,158114	0,329404	1						
10	-0,05774	-0,05774	0,08528	0,696572	0,35	0,228218	0,317543	0,505181	0,030429	1					
11	0,329404	0,329404	0,272475	0,15897	0,030429	-0,44444	0,158114	0,329404	0,638889	-0,16736	1				
12	-0,01318	-0,18447	0,1038	0,15897	0,228218	-0,08333	0,158114	0,158114	0,277778	0,228218	0,277778	1			
13	0,316228	0,474342	0,389249	-0,08671	0,182574	0,5	0,474342	0,316228	0	0	0	-0,16667	1		
14	0,054841	-0,12339	-0,34427	0,022556	0,079156	0,15897	0,233074	0,233074	-0,21678	0,079156	-0,0289	-0,21678	0,086711	1	
15	0,158114	0	-0,23355	-0,08671	0	-0,16667	0,316228	0,316228	-0,16667	0,182574	0,166667	0	0,076923	0,260133	1

Рисунок 10 – Матриця кореляції

Отже, узагальнюючи весь вище проведений аналіз тематичного тесту, можемо зробити наступні висновки:

– Тематичний тест «Матриці» є унімодальним, на це вказує значення моди яке є близьким до вибіркового середнього разом із інтервальним частотним розподілом і гістограмою.

– Вибіркове середнє частотного розподілу балів дорівнює – 9,8, значення медіани – 10 та значення коефіцієнта асиметрії, яке дорівнює – 0,02 при від'ємному ексцесі, який дорівнює – (-1,08) вказує на середню складність тесту. Другими словами, для рівня підготовленості студентів, які пройшли тест, він виявився не складним.

– коефіцієнт надійності Кронбаха, який дорівнює 0,7 показав надійність тестової контрольної роботи. Але, коефіцієнт кореляції Пірсона показує низьку розрізняльну здатність по всій вибірці. Результати, які отримали показують, що дана тестова контрольна робота є достатньо надійною, але завдання між собою є не дуже узгодженими та мають дуже невисоку розрізняльну здатність.

– підсумовуючи всі отримані результати, можна сказати, що даний тест потребує коригування. Оскільки це перший статистичний аналіз тесту подібного типу, це питання потребує подальших досліджень, зокрема на групах студентів з іншим рівнем підготовки. Це допоможе доповнити і скоригувати отримані висновки, а також розробити підходи та методики аналізу [15].

ВИСНОВКИ

У даній кваліфікаційній роботі було проаналізовано метод тестування як засіб для контролю рівня знань, умінь і навичок здобувачів математичної освіти, розглянуто переваги та недоліки впровадження комп'ютерних систем тестування знань, розглянуто характеристики комп'ютерних систем тестування знань які найчастіше використовуються.

Також у ході виконання кваліфікаційної роботи було зроблено детальний аналіз та дослідження застосування математичних методів та тестових технологій для діагностики навчальних досягнень здобувачів математичної освіти. Було проведено статистичний аналіз результатів тематичного тесту «Матриці» який проводився у студентів першого курсу 2023/2024 н. р. за допомогою он-лайн сервісу Google Forms. Аналіз завдань тесту дозволив з'ясувати взаємозв'язки між різними аспектами математичних знань, що відкриває можливості для подальшої оптимізації завдань та підвищення їхньої діагностичної ефективності.

Важливо відзначити, що дана робота відкриває перспективи для подальших досліджень у галузі використання тестових технологій у навчальному процесі, зокрема для адаптації методів до індивідуальних особливостей студентів та покращення процесу оцінювання їхніх навчальних досягнень.

Загалом, результати роботи підтверджують актуальність і перспективність використання математичних методів та тестових технологій для ефективної діагностики та підвищення якості математичної освіти, сприяючи розвитку новаторських підходів до оцінювання студентських знань у даній галузі.

ПЕРЕЛІК ПОСИЛАНЬ

1. Алексєєва І. В., Гайдей В. О., Диховичний О. О. Застосування сучасних математичних моделей педагогічного тестування у формуванні та аналізі тестових завдань комплекту «Вища математика» *Дидактика математики: проблеми і дослідження* : міжнародний збірник наукових робіт. №33. Донецьк, ДонНТУ, 2019, С.50– 56.
2. Сергієнко В. П., Малежик М. П., Сіткар Т. В. Комп'ютерні технології в тестуванні : навчальний посібник. Луцьк, 2019. С.290.
3. Конструювання тестів. Курс лекцій : навчальний посібник. / за ред. Кухар Л. О., Сергієнко В. П. Луцьк, 2018.с.182.
4. Алексєєва І. В., Гайдей В. О., Диховичний О. О. Застосування математичних моделей тестів у комплекті дистанційної освіти «Вища математика». *Математичні машини та системи*. 2018. №4. С. 89 – 98.
5. Глибовець М. М., Остапенко О. Ю. Аналіз тестових завдань на основі статистичної обробки результатів тестування. *Вісник Київського університету. Фізико-математичні науки*. 2019. №2. С. 22-24
6. Омельчук Т. С., Орловський І. В. Аналіз якості покрокових тестів з математичних дисциплін за допомогою сучасної теорії тестів. *Сьома всеукраїнська конференція студентів, аспірантів та молодих вчених з математики 19-20 квітня 2018 р.* : тези доповідей, Київ. С.8.
7. Рекомендації щодо впровадження змішаного навчання у закладах фахової перед вищої та вищої освіти. URL : <https://mon.gov.ua/storage/app/media/vishchaosvita/2020/zmyshene%20navchannu/zmishanenavchanniabookletsreads-2.pdf> (дата звернення: 30.08.2023).
8. Щенкова Ю. Г. Школа сьогодні: із досвіду роботи сучасних вчителів. *Всеукраїнська науково-практична конференція 13 січня 2021 р.* Київ. URL : <https://genezum.org/library/vykorystannya-kompyuternogo-testuvannya-dlya-osinyuvannya-navchalnyh-dosyagnen-uchniv> (дата звернення: 18.09.2023).

9. Петрицин І. Використання комп'ютерного тестування в процесі професійної підготовки студентів. *Молодь і ринок*. 2016. №2. С.23–28.

10. Інформаційно-комунікаційні технології в сучасній освіті: досвід, проблеми, перспективи : збірник наукових праць. / за ред. Козяра М. М., Ничкало Н. Г. Львів : ЛДУ БЖД, 2017.

11. Староста В. І., Шикітка Г. М. Тестовий контроль інформаційної компетентності майбутніх фахівців дошкільної та початкової освіти: навчальний посібник для самостійної роботи студентів. Ужгород. :УжНУ «Говерла», 2021.с.84.

12. Брітавська О. П., Дончев І. І., Бондаренко Н. В., Горохов В. В. Особливості комп'ютерних засобів тестування знань. URL : <https://core.ac.uk/download/pdf/158553842> (дата звернення: 3.10.2023).

13. Бакаленко О. А. Комп'ютерне тестування як форма контролю навчальних досягнень студентів в умовах дистанційного навчання. URL : <https://ojs.ukrlogos.in.ua/index.php/logos/article/view/15054/13677> (дата звернення: 10.10.2023)

14. Фетісов В. С. Основні вимоги до комп'ютерних систем тестування знань (КСТЗ) *Наукові записки НДУ ім. М. Гоголя. Психолого-педагогічні науки*. 2018. №10. URL : <https://moodle.ndu.edu.ua/file.php/1/NaykZap2010N11/mtp/mtp18.pdf> (дата звернення: 29.09.2023).

15. Орловський І. В., Тимошенко О. А. Аналіз якості нових методів контролю знань з вищої математики в технічному університеті. *Шоста міжнародна науково-практична 54 конференція «Математика в сучасному технічному університеті»*. 2017. Київ. С. 364-377.