

О. Ф. Шевчук¹
А. А. Яровий¹
Ю. М. Паночишин¹
С. І. Петришин¹
О. А. Козловський¹

МОДЕЛЮВАННЯ АДАПТИВНОГО ТЕСТУВАННЯ ЗНАНЬ: ПОРІГ ЕФЕКТИВНОСТІ, РІВЕНЬ СКЛАДНОСТІ ТА ЧАС ВИКОНАННЯ ЗАВДАНЬ

¹Вінницький національний технічний університет

Проведено комплексний інформаційно-аналітичний аналіз можливості впровадження адаптивного комп'ютерного тестування знань з окремих навчальних дисциплін у закладах освіти. Вказано на недоліки традиційного підходу, який передбачає встановлення фіксованих часових обмежень на виконання тесту, що не враховує індивідуальні особливості здобувачів освіти та може викликати негативну реакцію учасників тестування. Натомість розглянуто інтегральну оцінку тестування, яка враховує як рівень складності, так і час виконання окремого завдання. Запропоновано адаптивний алгоритм, який базується на порозі ефективності q , що визначає зміну рівня складності наступного завдання залежно від результату інтегральної оцінки попереднього. Для перевірки ефективності запропонованого підходу проведено імітаційне моделювання з використанням мови програмування Python. Створено тестову вибірку із завдань трьох рівнів складності, час виконання яких змодельовано за законом нормального розподілу. Проведений аналіз показав, що за значної різниці у рівнях складності завдань виникає потреба у встановленні окремого порогу ефективності для кожної категорії питань, тоді як у разі незначних відмінностей допустимо застосувати єдиний поріг для всіх завдань тесту. В межах тестової вибірки проведено налаштування параметрів інтегральної оцінки та досліджено ефективність запропонованого підходу. Зазначено, що отримані коефіцієнти інтегральної оцінки можуть виступати базовими на початковому етапі реалізації системи, з подальшою можливістю їхньої оптимізації на основі результатів навчання моделі в процесі тренувальних тестувань. Описана методика демонструє гнучкість і простоту впровадження, забезпечуючи можливість налаштування параметрів моделі та ефективну адаптацію як до індивідуальних особливостей здобувачів освіти, так і з урахуванням специфіки окремої дисципліни. До того ж фіксація часу виконання завдань може слугувати додатковим інструментом оцінювання якості тестових завдань.

Ключові слова: моделювання, адаптивне тестування, інтегральна оцінка, час виконання завдань, рівень складності.

Вступ

Сучасний освітній процес потребує інноваційних підходів до оцінювання знань, які забезпечують об'єктивність, гнучкість та зручність у використанні. Традиційні методи тестування не завжди дозволяють повною мірою врахувати індивідуальні особливості здобувачів освіти, такі як їхній рівень підготовки, темп засвоєння матеріалу та здатність до самостійного навчання.

У цьому контексті впровадження інформаційних технологій дозволяє не лише автоматизувати сам процес оцінювання знань, а й створити нові можливості для налаштування тестових систем таким чином, щоб вони динамічно адаптувались до індивідуальних потреб здобувача освіти.

Більше того, адаптивні системи тестування забезпечуючи точнішу оцінку рівня знань, виступають й додатковим мотиваційним фактором та сприяють досягненню ліпших результатів у порівнянні з традиційними методами тестування [1]—[3].

Наразі алгоритми адаптивного тестування широко використовуються у сучасних освітніх плат-

формах, таких як Coursera, Duolingo, Khan Academy та інші. Ці платформи демонструють ефективність адаптивних технологій у підвищенні якості навчання завдяки індивідуалізованому підходу, оптимізації навчального процесу та інтерактивності [4], [5].

Проте на рівні закладів освіти, в процесі викладання окремих курсів, застосування адаптивного підходу тестування знань залишається доволі обмеженим і здебільшого залежить від ініціативи окремих новаторів-науковців. За їхньої підтримки створені методичні рекомендації [6] та інструкції [7], [8] щодо покрокової розробки тестових завдань з поступовим підвищенням рівня складності на основі платформ Google Forms, LMS Moodle тощо. Але, незважаючи на це, практичне впровадження адаптивного тестування однаково залишається доволі громіздким процесом, що вимагає значних часових та ресурсних витрат. До того ж необхідність ручного налаштування кожного тесту з урахуванням логіки їхньої послідовності значно знижує ефективність використання таких систем.

Натомість автоматизовані системи тестування з інтегрованими алгоритмами адаптивного тестування здатні оптимізувати цей процес, значно зменшуючи навантаження на викладачів і водночас забезпечуючи індивідуалізований підхід до кожного студента. Завдяки автоматичному аналізу результатів система може динамічно змінювати рівень складності завдань відповідно до рівня знань студента [3], [9]. Проте у цьому контексті постає проблема вибору оптимальної методики адаптації складності завдань та її відповідного калібрування.

На сьогодні існує широкий спектр методичних підходів до адаптації рівня складності тестових завдань у системах комп'ютерного тестування [10]—[15]. Вони варіюються від базових алгоритмів, які орієнтуються на попередні результати оцінювання, до комплексних математичних моделей. Зокрема, у найпростіших моделях, підвищення рівня складності наступного завдання відбувається лише за умови правильної відповіді на поточне питання або серії правильних відповідей поспіль.

Складніші алгоритми базуються на моделях глибокого навчання (Deep Knowledge Tracing) [12], [13] та ймовірнісних підходах на основі Байєсової оцінки (Bayesian Knowledge Tracing), яка дозволяє достатньо точно визначати рівень знань здобувача та підбирати завдання відповідно до його поточних можливостей [14], [15]

$$P(s_j|R, q_i) = \frac{P(q_i|s_j) \cdot P(s_j|R)}{P(q_i)}, \quad (1)$$

де $P(s_j|R, q_i)$ — ймовірність того, що користувач володіє навичкою s_j , з урахуванням його попередніх відповідей (R) і результату відповіді на поточне питання q_i ; $P(q_i|s_j)$ — ймовірність того, що користувач з рівнем знань s_j правильно відповість на питання q_i ; $P(s_j|R)$ — ймовірність того, що тестований володіє навичкою s_j , на основі попередніх відповідей (R); $P(q_i)$ — ймовірність того, що питання q_i буде запропоновано у тесті.

Така Байєсова модель адаптивного тестування є однією з найефективніших та наразі широко застосовуваних у сучасних інтелектуальних навчальних системах. Вона успішно інтегрована в платформи для вивчення мов і оцінювання рівня знань, такі як Duolingo та інші. Проте її практичне впровадження у сфері тестування знань з окремих навчальних дисциплін у закладах освіти ускладнюється через потребу значних ресурсів, спеціальної технічної підготовки та точного калібрування завдань. А тому в цьому випадку доцільніше використовувати менш ресурсоємні алгоритми адаптивного тестування, які зокрема можуть бути розширені за рахунок введення додаткових параметрів оцінювання.

Одним із таких параметрів може слугувати час виконання завдань, інтеграція якого у формулу розрахунку підсумкової оцінки тесту дозволить системі динамічно коригувати рівень складності наступних запитань [16]—[17]. Такий підхід дозволить значно спростити роль розробника тестів, оскільки йому потрібно буде лише встановити рівень складності кожного тестового завдання або вказати його бальну шкалу оцінювання, тоді як решта параметрів можуть бути визначені за допомогою імітаційного моделювання.

Метою статті є побудова вдосконаленої моделі тестування знань на основі інтегральної оцінки з урахуванням рівня складності завдань та часу їхнього виконання, що забезпечує адаптивність процесу оцінювання до індивідуальних когнітивних особливостей здобувача освіти. Зокрема завдання статті передбачають розробку відповідної моделі оцінювання, алгоритму адаптивного тестування знань та проведення імітаційного моделювання для налаштування оптимальних параметрів системи тестування та визначення її загальної ефективності.

Результати дослідження

Зазвичай, у комп'ютерному тестуванні, час, відведений на виконання окремого завдання, визначається як граничний параметр, після перевищення якого відповідь системою не зраховується. Проте встановлення такого нормативу доволі часто є суб'єктивним, а інколи й помилковим, що до того ж не враховує індивідуальних особливостей здобувачів освіти. Як наслідок це може викликати певну негативну реакцію серед учасників тестування, знижуючи їхню мотивацію та впливаючи на об'єктивність і достовірність отриманих результатів.

А тому параметр, такий як час виконання тестового завдання, доцільно розглядати не як фіксовану величину, а як динамічний показник, інтегрований у загальну модель оцінювання знань. У цьому випадку, за умови правильності відповіді, інтегральна оцінка виконання окремого тесту може бути розрахована на основі адитивної або мультиплікативної моделі, що враховує рівень складності завдання та ефективність виконання за часом.

Для адитивної моделі інтегральна оцінка (I) визначається на основі зваженої суми двох компонент, одна з яких визначає рівень складності тесту S_k , а інша — ефективність виконання тесту за часом T_k

$$I = w_1 S_k + w_2 T_k, \quad (2)$$

де w_1, w_2 — вагові коефіцієнти, що визначають відносну важливість складності завдання та часу виконання ($w_1 + w_2 = 1$); k — рівень складності тестового питання; S_k — оцінка тесту рівня складності k ; T_k — ефективність виконання тестового завдання рівня складності k за часом ($0 < T_k \leq 1$).

При цьому час виконання тестового завдання, на основі якого розраховується T_k , регулює інтегральну оцінку лише в межах відповідного вагового коефіцієнта (w_2)

$$(w_1 S_k) < I \leq (w_1 S_k + w_2).$$

З одного боку, це дозволяє коригувати підсумковий результат правильної відповіді залежно від часу виконання тесту, а з іншого — встановлює межі такого впливу, запобігаючи повному анулюванню позитивної відповіді учасника тестування.

Натомість у мультиплікативній моделі оцінювання результатів тестування, додаткові вагові коефіцієнти не застосовуються

$$I = S_k \cdot T_k. \quad (3)$$

Але у такому випадку інтегральна оцінка в цілому залежить від значення показника ефективності виконання тестового завдання за часом (T_k) і може наблизитися до нуля, навіть якщо учасник тестування дав правильну відповідь

$$0 < I \leq S_k.$$

Отже такий підхід підкреслює важливість швидкості відповіді, проте він може надмірно знижувати підсумкову оцінку для учасників, які правильно виконали завдання, але витратили більше часу. А тому це накладає певні обмеження на застосування цього методу та потребує ретельнішого обґрунтування вибору методики розрахунку показника T_k на основі фактичного часу виконання тестового завдання. До того ж дослідження у сфері психометрії [18]—[21] вказують і на досить складний нелінійний зв'язок між швидкістю та точністю відповідей, який також може залежати як від впливу зовнішніх чинників, так і від когнітивних особливостей учасника тестування.

З огляду на це, на початковому етапі доцільніше застосувати спрощений підхід до інтегральної оцінки під час розробки комп'ютерної адаптивної моделі тестування знань, яка згодом може бути удосконалена з урахуванням сучасних психометричних підходів. Таким чином, у межах цього дослідження буде використано адитивну інтегральну модель оцінювання (2), в якій ефективність T_k визначатиметься як обернена величина до фактичного часу виконання тестового завдання

$$T_k = \begin{cases} 1, & t \leq Tn_k, \\ \frac{Tn_k}{t}, & t > Tn_k, \end{cases} \quad (4)$$

Tn_k — нормативний час, відведений на виконання завдання складності k ; t — фактичний час виконання поточного питання.

В такому випадку адаптивний механізм вибору подальшого питання базуватиметься на введенні показника порогу ефективності відповіді q . Зокрема, рівень складності подальшого завдання

підвищуватиметься за умови досягнення або перевищення визначеного значення порогу ефективності q . У разі ж недосягнення цього рівня складність наступного завдання, навпаки, знижуватиметься

$$k = \begin{cases} k + 1, & I \geq q; \\ k - 1, & I < q. \end{cases} \quad (5)$$

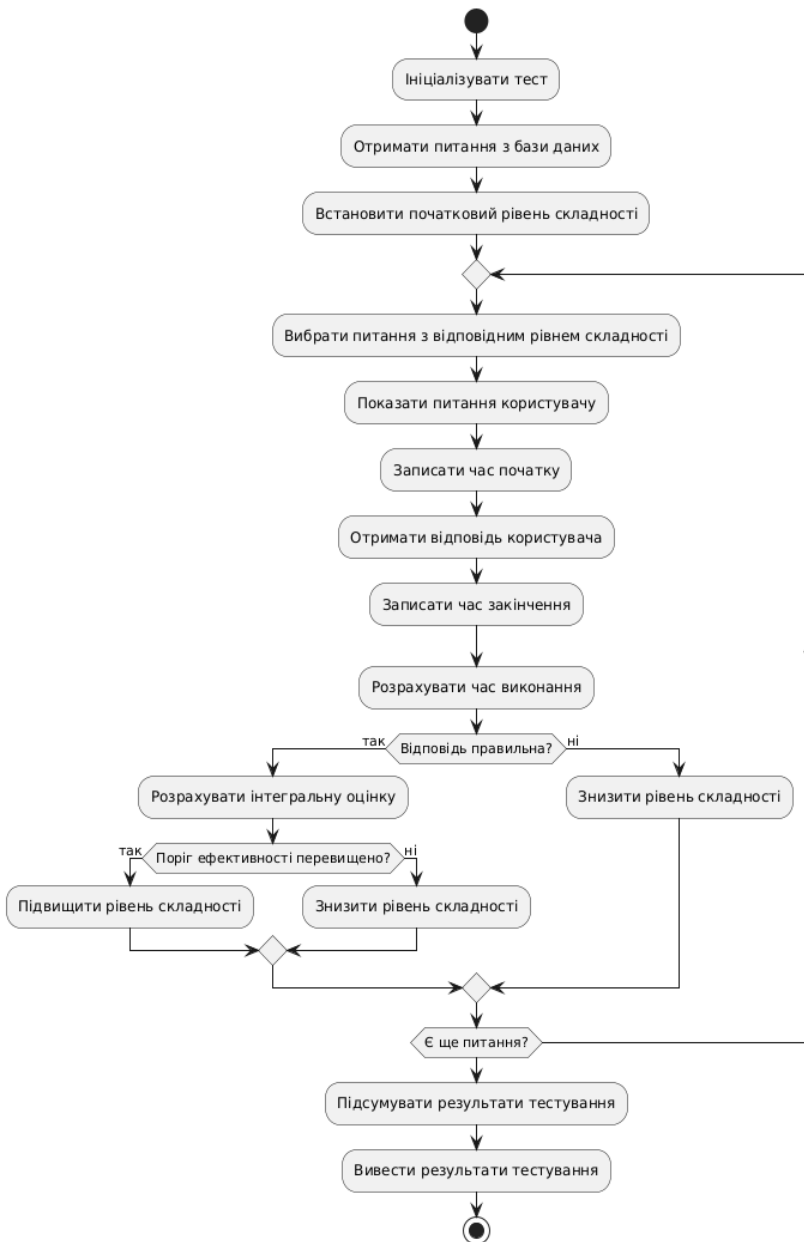


Рис. 1. Діаграма діяльності адаптивного тестування знань на основі інтегральної оцінки

На рис. 1 показано діаграму діяльності процесу адаптивного тестування знань, що реалізує описаний вище підхід.

Зазначимо також певні особливості та переваги наведеної моделі адаптивного тестування знань. Насамперед, вона є доволі простою та гнучкою у налаштуванні параметрів оцінювання. До прикладу, регулювання коефіцієнтів w_1 та w_2 дозволяє змінювати вплив часу виконання завдання на результат інтегральної оцінки. Це, і собі, дає змогу адаптувати модель до особливостей конкретних дисциплін. Зокрема, під час вивчення іноземних мов час відповіді на питання може бути одним з ключових показників, оскільки швидкість реагування відображає рівень компетентності навичок мовлення чи розуміння. Водночас у дисциплінах, що вимагають ретельного аналізу чи розрахунків (наприклад, у математиці чи фізиці), час виконання тестового завдання може мати менший вплив на кінцевий результат тестування.

Решта параметрів моделі, таких як нормативний час виконання Tn_k та оцінка S_k питання складності k , можуть бути визначені як експертним методом, так і шляхом навчання моделі на основі даних, отриманих під час тренувальних або пробних тестувань. У першому випадку експерти задають значення параме-

трів, враховуючи специфіку дисципліни та досвід практичного застосування. У другому випадку модель адаптується до особливостей користувачів шляхом аналізу результатів тестування.

При цьому критерії для встановлення нормативів можуть базуватися на стандартах оцінювання, наприклад, системи ECTS, що дозволяє забезпечити узгодженість моделі з загальноприйнятими вимогами до якості освітнього процесу.

Для перевірки ефективності запропонованого підходу до адаптивного тестування знань проведено імітаційне моделювання, реалізоване мовою програмування Python у середовищі Jupyter Notebook з використанням бібліотек numpy, pandas та matplotlib.

На початковому етапі створено тестову вибірку, що містить завдання трьох рівнів складності:

легкий, середній та складний. Час виконання завдань для кожного рівня згенеровано відповідно до нормального розподілу з параметрами T_k та σ_k , що відповідають їхній складності. Для наближення до реальних умов у вибірку також додано аномальні значення, які утворювалися шляхом випадкового множення часу виконання окремих завдань на заданий коефіцієнт аномалії (рис. 2). На основі сформованого датасету розраховано ефективність виконання кожного тестового завдання та його відповідна інтегральна оцінка (рис. 3).

```
def generate_time_data(mean, std_dev, size, anomaly_prob=0.05, anomaly_multiplier=3):
    times = np.random.normal(mean, std_dev, size)
    anomalies = np.random.rand(size) < anomaly_prob
    times[anomalies] *= anomaly_multiplier
    return times

np.random.seed(42)
simple_times = generate_time_data(4, 0.8, 500)
medium_times = generate_time_data(7.5, 2.5, 500)
hard_times = generate_time_data(13, 4, 500)

data = pd.DataFrame({
    'Category': ['Simple']*500 + ['Medium']*500 + ['Hard']*500,
    'Time': np.concatenate([simple_times, medium_times, hard_times])
})

simple_threshold = np.percentile(data[data['Category'] == 'Simple']['Time'], 56)
medium_threshold = np.percentile(data[data['Category'] == 'Medium']['Time'], 50)
hard_threshold = np.percentile(data[data['Category'] == 'Hard']['Time'], 47)

def calculate_efficiency(time, threshold):
    return 1 if time <= threshold else threshold / time

data['Efficiency'] = data.apply(lambda row: calculate_efficiency(
    row['Time'],
    simple_threshold if row['Category'] == 'Simple' else
    medium_threshold if row['Category'] == 'Medium' else
    hard_threshold
), axis=1)
```

Рис. 2. Фрагмент програмного коду створення тестової вибірки

Category	Time	Efficiency	Weight	Score
0	Simple	4.397371	0.948572	0.8 0.859429
1	Simple	3.889389	1.000000	0.8 0.880000
2	Simple	4.518151	0.923215	0.8 0.849286
3	Simple	5.218424	0.799326	0.8 0.799731
4	Simple	3.812677	1.000000	0.8 0.880000
...
1495	Hard	30.869224	0.417703	1.0 0.767081
1496	Hard	17.776438	0.725351	1.0 0.890140
1497	Hard	9.075338	1.000000	1.0 1.000000
1498	Hard	11.142383	1.000000	1.0 1.000000
1499	Hard	44.544731	0.289465	1.0 0.715786

1500 rows x 5 columns

Рис. 3. Структура датасету

На рис. 4 показано приклади обчислення інтегральної оцінки (2) для згенерованих тестових завдань, за використання параметрів $w_1 = 0,6$, $w_2 = 0,4$. Значення параметра складності S_k змінювалися таким чином: у першому випадку $S_1 = 0,6$; $S_2 = 0,8$; $S_3 = 1,0$; у другому — $S_1 = 0,8$; $S_2 = 0,9$; $S_3 = 1,0$. На основі проведених розрахунків визначено порogi ефективності $q = 0,7$ та $q = 0,85$ відповідно.

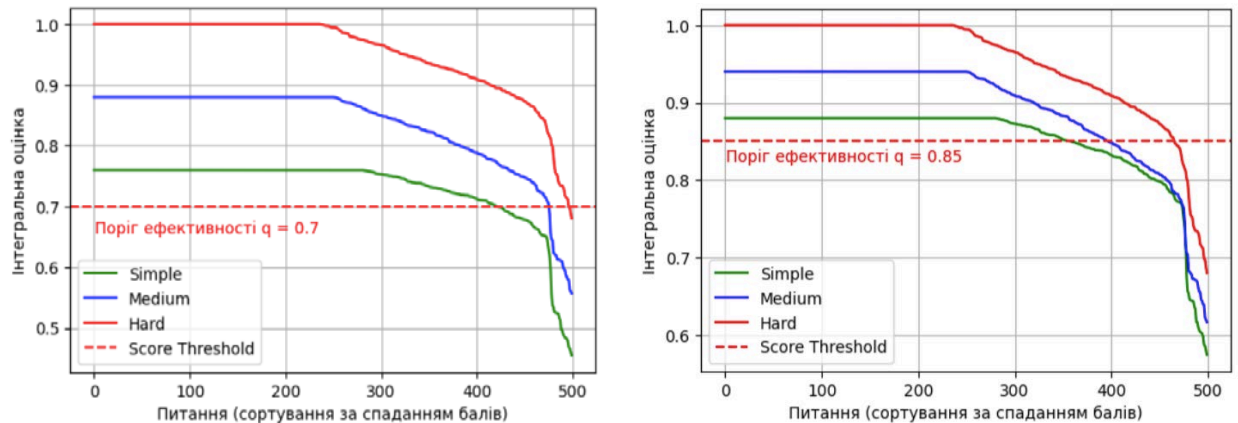


Рис. 4. Розрахунок інтегральної оцінки тесту для різних категорій складності питань

Аналіз графіків на рис. 4 показує, що за значної різниці у рівнях складності завдань, необхідно встановлювати окремий поріг ефективності для кожної категорії питань. Це, насамперед пояснюється тим, що переважна більшість питань найвищої складності долає поріг ефективності, а тому можуть спотворюватися результати адаптації. Натомість, якщо відмінності між рівнями складності завдань є незначними, як у другому випадку, то цілком можливо застосувати єдине значення порогу ефективності для всіх завдань такого тесту.

Далі проведено імітаційне моделювання адаптивного процесу тестування для різної кількості завдань n . При цьому визначено єдиний поріг ефективності для всіх питань тесту ($q = 0,85$), а ймовірність правильної відповіді змінювалася залежно від рівня складності завдань. Решта параметрів

відповідали попередньо розглянутому прикладу ($S_1 = 0,8$, $S_2 = 0,9$, $S_3 = 1,0$, $w_1 = 0,6$, $w_2 = 0,4$). Процес тестування розпочинався із завдання середньої складності, після чого відбувалася зміна рівня на вищий або нижчий, в залежності від розрахованої інтегральної оцінки.

На рис. 5 подано фрагмент програмного коду, що реалізує процес імітаційного моделювання тестування користувачів з динамічним регулюванням рівня складності питань. Основною метою є аналіз часу, витраченого на відповіді, та його вплив на підсумкову оцінку ефективності. У процесі роботи моделі використовується словник `difficulty_settings`, який містить середній час виконання завдань, стандартне відхилення, вагові коефіцієнти та процентильні порогові значення для кожного рівня складності.

```
difficulty_settings = {
    'simple': {'mean_time': 4, 'std_dev': 0.8, 'weight': 0.8, 'threshold_percentile': 56},
    'medium': {'mean_time': 7.5, 'std_dev': 2.5, 'weight': 0.9, 'threshold_percentile': 50},
    'hard': {'mean_time': 13, 'std_dev': 4, 'weight': 1.0, 'threshold_percentile': 47}
}

def generate_time_data(mean, std_dev):
    return np.random.normal(mean, std_dev)

def calculate_efficiency(time, threshold):
    return 1 if time <= threshold else threshold / time

def generate_test_results(username, num_tests=n, initial_difficulty='medium'):
    results = []
    current_difficulty = initial_difficulty

    for question_num in range(1, num_tests + 1):
        settings = difficulty_settings[current_difficulty]
        response_time = generate_time_data(settings['mean_time'], settings['std_dev'])
        threshold = np.percentile(
            np.random.normal(settings['mean_time'], settings['std_dev'], 1000),
            settings['threshold_percentile']
        )

        efficiency = calculate_efficiency(response_time, threshold)
        correctness_prob = {'simple': 0.9, 'medium': 0.8, 'hard': 0.7}[current_difficulty]
        correct = np.random.rand() < correctness_prob
```

Рис. 5. Фрагмент програмного коду, що реалізує процес імітаційного моделювання тестування користувачів з динамічним регулюванням рівня складності питань

Генерація часу відповіді здійснюється функцією `generate_time_data(mean, std_dev)`, яка використовує нормальний розподіл для моделювання варіативності у швидкості виконання тестових завдань. Для кожного питання розраховується порогове значення часу, що визначається емпіричним процентилем на основі випадкової вибірки. Оцінка ефективності проводиться у функції `calculate_efficiency(time, threshold)`, яка визначає ефективність виконання завдання як відношення порогового значення до фактичного часу відповіді або присвоює їй значення одиниці, якщо час не перевищує цей поріг.

Основна логіка тестування реалізована у функції `generate_test_results(username, num_tests, initial_difficulty)`. У цій функції відбувається ітеративне генерування тестових завдань з визначенням часу відповіді та розрахунком ефективності виконання. Визначення правильності здійснюється ймовірнісним методом, де кожному рівню складності відповідає певна ймовірність правильної відповіді на питання тесту. Адаптивний механізм коригує рівень складності наступного завдання відповідно до отриманих результатів, що дозволяє поступово змінювати складність залежно від рівня знань користувача.

Результати проведеного тестування формуються у вигляді відповідного датасету (рис. 6) та візуалізуються графічно (рис. 7), що дозволяє відслідку-

	Username	Question	Difficulty	Correct	Response Time	Efficiency	Time	Score
0	user_123	1	medium	True	8.58	0.88	0.89	
1	user_123	2	hard	True	12.81	1.00	1.00	
2	user_123	3	hard	True	3.78	1.00	1.00	
3	user_123	4	hard	True	14.11	1.00	1.00	
4	user_123	5	hard	False	10.11	1.00	0.00	
5	user_123	6	medium	True	9.86	0.75	0.84	
6	user_123	7	simple	True	5.98	0.58	0.71	
7	user_123	8	simple	True	4.79	0.71	0.76	
8	user_123	9	simple	True	3.31	1.00	0.88	
9	user_123	10	medium	False	5.73	1.00	0.00	

Рис. 6. Структура датасету результатів адаптивного тестування знань

вати покроково весь процес оцінювання знань учасника адаптивного тестування.

Як видно з наведеного прикладу (рис. 7), із 10 запропонованих тестових завдань системою надано 4 складних, 3 середнього рівня та 3 простих питання, на які отримано 8 правильних відповідей в рамках імітаційного моделювання. З урахуванням часу виконання завдань та їхньої складності підсумковий результат становить 70,8 %, що відповідає оцінці D за шкалою ECTS.

На рис. 8 також показано результати роботи адаптивної системи за умови збільшення кількості тестових завдань.

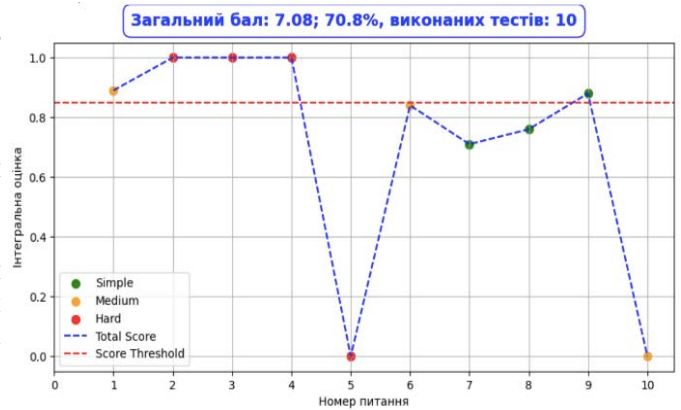


Рис. 7. Візуалізація результатів адаптивного тестування знань

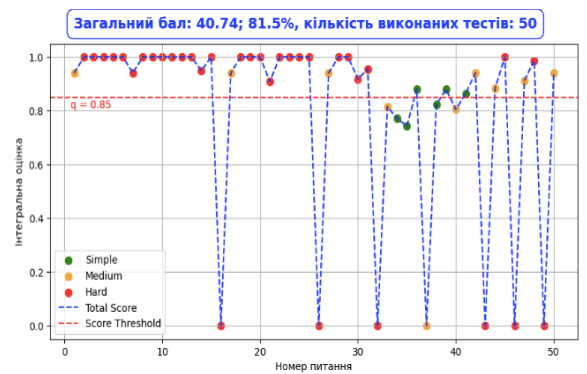
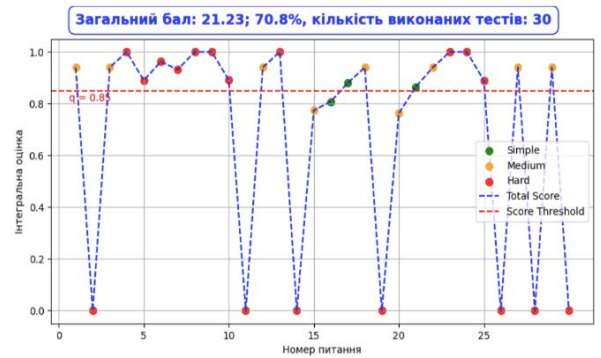
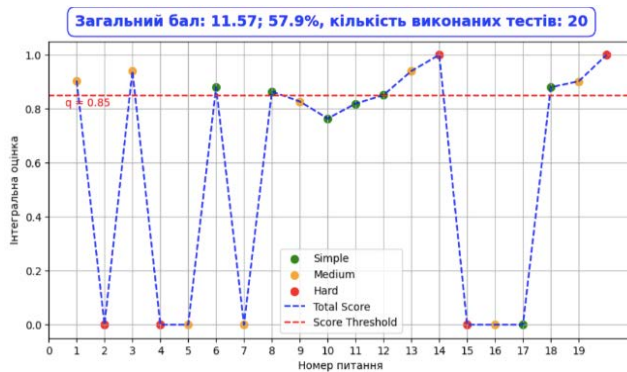


Рис. 8. Імітаційне моделювання адаптивного процесу тестування знань при різній кількості тестових завдань n

Загалом, можна зазначити, що запропонований процес адаптації тестових завдань до рівня знань здобувача освіти виявився достатньо ефективним. До того ж наведені коефіцієнти інтегральної оцінки можуть бути використані на початковому етапі впровадження системи адаптивного тестування. У подальшому їхні значення підлягатимуть корекції на основі результатів навчання системи під час проведення тренувальних тестувань, що загалом дозволить підвищити точність та адаптивність оцінювання.

З іншого боку, розроблена програмна реалізація імітаційного моделювання дозволяє дослідити можливості удосконалення запропонованого підходу до адаптивного тестування знань шляхом модифікації функціональних залежностей розрахунку T_k та I . Це, у свою чергу, відкриває перспективи для подальших досліджень, спрямованих на удосконалення процесу оцінювання знань.

Також варто зазначити, що фіксація часу виконання кожного тестового завдання та його подальший статистичний аналіз можуть бути корисними як додатковий інструмент оцінки якості тестів. З одного боку, цей аналіз дозволить виявити можливі проблеми або недоліки в конструкції тесту, такі як надмірна складність чи некоректність формулювання завдань. З іншого боку, час виконання може виступати індикатором недоброчесності того, що тестується, оскільки надто повільне виконання завдань може свідчити про неналежне ставлення до тестування, що вимагає додаткового контролю та перевірки.

Висновки

Проведене інформаційно-аналітичне дослідження показало, що використання інтегральної оцінки знань, яка враховує рівень складності та час виконання завдань, дозволяє створити ефективну й водночас відносно просту модель адаптивного комп'ютерного тестування. Основні параметри моделі можуть визначатися автоматично за допомогою серії тренувальних тестувань, що сприяє підвищенню точності оцінки та усуненню суб'єктивного впливу. У разі значної різниці у рівнях складності завдань необхідно встановити окремий поріг ефективності для кожної категорії питань, тоді як за незначних відмінностей можливе застосування єдиного порогу ефективності для всіх завдань тесту.

Розроблена програмна реалізація імітаційного моделювання створює основу для подальших досліджень, спрямованих на вдосконалення адаптивних методів оцінювання знань.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] A. Frey, T. Liu, A. Fink, and C. König, "Meta-Analysis of the Effects of Computerized Adaptive Testing on the Motivation and Emotion of Examinees," *Eur. Journal Psychol. Assess.*, vol. 40, no. 5, pp. 427-443, 2024. <https://doi.org/10.1027/1015-5759/a000821>.
- [2] N. Sherkuzyeva, et al., "The comparative effect of computerized dynamic assessment and rater mediated assessment on EFL learners' oral proficiency, writing performance, and test anxiety," *Lang. Test Asia*, vol. 13, no. 15, 2023. <https://doi.org/10.1186/s40468-023-00227-3>.
- [3] P. Gawliczek, V. Krykun, N. Tarasenko, M. Tyshchenko, and O. Shapran, "Computer Adaptive Language Testing According to NATO STANAG 6001 Requirements," *Adv. Educ.*, no. 8 (17), pp. 19-26, 2021. <https://doi.org/10.20535/2410-8286.225018>.
- [4] В. В. Камінський, В. А. Мізюкі, Р. Д. Турчанінов, «Аналіз ефективності штучного інтелекту в адаптивних навчальних платформах для індивідуалізації освітнього процесу», *ZENODO, Груд*, 2024. [Electronic resource]. Available: <https://doi.org/10.5281/zenodo.14562152>.
- [5] A. Trifa, A. Hedhili, and W. L. Chaari, "Knowledge tracing with an intelligent agent, in an e-learning platform," *Educ. Inf. Technol.*, vol. 24, no. 1, pp. 711-741, 2019. <https://doi.org/10.1007/s10639-018-9792-5>.
- [6] Я. Б. Сікора, *Методичні рекомендації до розробки та використання тестових завдань*, Житомир: вид-во ЖДУ ім. Івана Франка, 2024. [Online]. Available: http://eprints.zu.edu.ua/41797/1/metod_test.pdf.
- [7] О. Радкевич, «Адаптивне тестування в контексті використання електронних засобів навчання. Суть, розроблення та оцінювання», *Професійна педагогіка*, т. 1, № 26, с. 58-73, 2023. <https://doi.org/10.32835/2707-3092.2023.26.58-73>.
- [8] С. Загребельний, «Використання комп'ютерного адаптивного тестування у ДДМА на платформі Moodle», *Техн. Електр. Навч.*, no. 3, Лист., 2019.
- [9] Н. Васюкова, В. Крикун, Ю. Грищук, і А. Кравчук, «Експериментальна перевірка результативності методики комп'ютерного адаптивного мовного тестування відповідно до вимог НАТО STANAG 6001», *Людознавчі студії. Серія «Педагогіка»*, № 19 (51), с. 9-17, 2024. <https://doi.org/10.24919/2413-2039.19/51.1>.
- [10] W. J. van der Linden, *Linear models for optimal test design*, Springer, 2005. <https://doi.org/10.1007/0-387-29054-0>.
- [11] С. Л. Загребельний, М. В. Брус, "Адаптивне тестування як один із способів перевірки знань студентів у технічному вузі," *Науковий Вісник ДДМА*, vol. 1, no. 22Е, pp. 155-162, 2017.
- [12] E. H. Am, I. Hidayah, and S. S. Kusumawardani, "A Literature Review of Knowledge Tracing for Student Modeling: Research Trends, Models, Datasets, and Challenges," *Journal. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, 2021. <https://doi.org/10.25126/jitecs.202162344>.
- [13] S. Liu, R. Zou, J. Sun, K. Zhang, L. Jiang, and D. Zhou, "A Hierarchical Memory Network for Knowledge Tracing," *Expert Syst. Appl.*, vol. 177, p. 114935, 2021. <https://doi.org/10.1016/j.eswa.2021.114935>.
- [14] O. Bulut, J. Shin, S. N. Yildirim-Erbasli, G. Gorgun, and Z. A. Pardos, "An Introduction to Bayesian Knowledge Tracing with pyBKT," *Psych.*, vol. 5, no. 3, pp. 770-786, 2023. <https://doi.org/10.3390/psych5030050>.
- [15] F. Liu, X. Hu, C. Bu, and K. Yu, "Fuzzy Bayesian Knowledge Tracing," *IEEE Trans. Fuzzy Syst.*, vol. 6706, pp. 1-15, 2021. <https://doi.org/10.1109/TFUZZ.2021.3083177>.
- [16] W. J. van der Linden, "Modeling response times with latent variables: Principles and applications," *Psychol. Test Assess. Model.*, vol. 53, no. 3, pp. 334-358, 2011.
- [17] W. J. van der Linden and X. Xiong, "Speededness and Adaptive Testing," *J. Educ. Behav. Stat.*, vol. 38, pp. 418-438, 2013.
- [18] W. J. van der Linden, D. J. Scrams, and D. L. Schnipke, "Using Response-Time Constraints to Control for Differential Speededness in Computerized Adaptive Testing," *Appl. Psychol. Meas.*, vol. 23, no. 3, pp. 195-210, 1999. <https://doi.org/10.1177/01466219922031329>.
- [19] H. Sie, M. D. Finkelman, B. Riley, and N. Smits, "Utilizing Response Times in Computerized Classification Testing," *Appl. Psychol. Meas.*, vol. 39, no. 5, pp. 389-405, 2015. <https://doi.org/10.1177/0146621615569504>.
- [20] B. Becker, P. van Rijn, D. Molenaar, and D. Debeer, "Item order and speededness: implications for test fairness in higher educational high-stakes testing," *Assess. Eval. High. Educ.*, vol. 47, no. 7, pp. 1030-1042, 2021. <https://doi.org/10.1080/02602938.2021.1991273>.
- [21] R. H. Klein Entink, G. J. A. Fox, and W. J. van der Linden, "A Box-Cox normal model for response times," *Br. J. Math. Stat. Psychol.*, vol. 62, no. 2, pp. 621-640, 2009. <https://doi.org/10.1348/000711008X354742>.

Шевчук Олександр Федорович — канд. фіз.-мат. наук, доцент, доцент кафедри комп'ютерних наук, e-mail: shevchuk177@gmail.com ;

Яровий Андрій Анатолійович — д-р техн. наук, професор, завідувач кафедри комп'ютерних наук, e-mail: a.yarovyy@vntu.edu.ua ;

Паночішин Юрій Миколайович — канд. техн. наук, доцент, доцент кафедри комп'ютерних наук, e-mail: y.panochyshyn@vntu.edu.ua ;

Петришин Сергій Іванович — канд. техн. наук, старший викладач кафедри комп'ютерних наук, e-mail: petryshyn@vntu.edu.ua ;

Козловський Олексій Андрійович — студент факультету інтелектуальних інформаційних технологій та автоматизації, e-mail: Oleksii.Kozlovskiy@outlook.com .

Вінницький національний технічний університет, Вінниця

O. F. Shevchuk¹
A. A. Yarovyi¹
Yu. M. Panochyshyn¹
S. I. Petryshyn¹
O. A. Kozlovskiy¹

Modeling of Adaptive Knowledge Testing: Efficiency Threshold, Task Complexity and Completion Time

¹Vinnitsia National Technical University

A comprehensive informational and analytical analysis has been conducted to evaluate the feasibility of implementing adaptive computer-based knowledge testing for specific academic disciplines in educational institutions. The shortcomings of the traditional approach, which imposes fixed time constraints for test completion without considering the individual characteristics of learners and potentially causing negative reactions among test participants, have been identified. Alternatively, an integral assessment approach is proposed, accounting for both task complexity and task completion time. An adaptive algorithm has been developed based on the efficiency threshold q , which determines the adjustment of the difficulty level for subsequent tasks depending on the integral evaluation result of the previous task. Simulation modeling was carried out using Python to verify the effectiveness of the proposed approach. A test dataset comprising tasks of three complexity levels was created, with completion times modeled according to the normal distribution. The analysis revealed that significant differences in task difficulty levels necessitate establishing separate efficiency thresholds for each category of questions, while minor differences allow for a single threshold for all test tasks. Parameter tuning for the integral assessment was performed within the test dataset, and the effectiveness of the proposed method was examined. It was noted that the obtained coefficients of the integral evaluation could serve as baseline values during the initial implementation phase of the system, with further optimization based on model training results during pilot testing. The described methodology demonstrates flexibility and ease of implementation, enabling parameter customization and effective adaptation to both the individual characteristics of learners and the specific requirements of individual disciplines. Furthermore, recording task completion times can serve as an additional tool for assessing the quality of test items.

Keywords: modeling, adaptive testing, integral assessment, task completion time, task complexity.

Shevchuk Oleksandr F. — Cand. Sc. (Phys.-Math.), Associate Professor, Associate Professor of the Chair of Computer Sciences, e-mail: shevchuk177@gmail.com ;

Yarovyi Andrii A. — Dr. Sc. (Eng.), Professor, Head of the Chair of Computer Science, e-mail: a.yarovyy@vntu.edu.ua ;

Panochyshyn Yurii M. — Cand. Sc. (Eng.), Associate Professor, Associate Professor of the Chair of Computer Science, e-mail: y.panochyshyn@vntu.edu.ua ;

Petryshyn Serhii I. — Cand. Sc. (Eng.), Senior Lecturer of the Chair of Computer Science, e-mail: petryshyn@vntu.edu.ua ;

Kozlovskiy Oleksii A. — Student of the Department of Intelligent Information Technologies and Automation, e-mail: Oleksii.Kozlovskiy@outlook.com