



ФІЗИКО-МАТЕМАТИЧНА ОСВІТА (PHYSICAL AND MATHEMATICAL EDUCATION)

p-ISSN: 2413-1571, e-ISSN: 2413-158X

2026, 41(3), <https://doi.org/10.31110/fmo2026.v41i3-05>

ВІД САМОПЕРЕВІРКИ ДО ВЗАЄМНОГО РЕЦЕНЗУВАННЯ: РОЗВИТОК КОГНІТИВНО- ВЕРИФІКАЦІЙНОГО ПІДХОДУ ДО ОЦІНЮВАННЯ МАТЕМАТИЧНИХ ЗАВДАНЬ В УМОВАХ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ

Ярослав ЧКАНА ✉

Сумський державний педагогічний університет
імені А.С.Макаренка, Україна
ya.chkana_76@sspu.edu.ua
<https://orcid.org/0000-0003-3667-3584>

FROM SELF-ASSESSMENT TO PEER REVIEW: THE DEVELOPMENT OF A COGNITIVE-VERIFICATION APPROACH TO THE EVALUATION OF MATHEMATICAL PROBLEMS IN THE CONTEXT OF GENERATIVE ARTIFICIAL INTELLIGENCE

Yaroslav CHKANA ✉

Sumy State Pedagogical University
named after A. S. Makarenko, Ukraine
ya.chkana_76@sspu.edu.ua
<https://orcid.org/0000-0003-3667-3584>

АНОТАЦІЯ

Формулювання проблеми. Генеративний штучний інтелект підриває діагностичну надійність домашніх завдань у дистанційній математичній освіті: зовнішньо правильний результат більше не є достатнім доказом самостійного досягнення результатів навчання. Ключовим ризиком є метакогнітивна пасивність студента — делегування технології аналізу умови, вибору методу та перевірки результату. Стаття розвиває когнітивно-верифікаційний підхід автора, переносючи верифікаційну діяльність з індивідуальної самоперевірки на аналіз чужого математичного міркування через взаємне рецензування.

Матеріали і методи. Дослідження є теоретичним із пілотною апробацією. Теоретичну основу становлять концепція валідності оцінювання Мезіка, поняття метакогнітивної пасивності та концепція «assessment twins». На цій базі сконструйовано п'ятикомпонентну ризик-орієнтовану модель оцінювання. Пілот передбачав одну академічну групу (10 студентів), одне завдання з математичного аналізу та якісний описовий аналіз результатів.

Результати. Розроблено п'ятикомпонентну модель: когнітивно-верифікаційне розв'язання з контрольними точками; структуроване рецензування аналогічної роботи за критеріями; обов'язкова відповідь здобувача на рецензію; коротка саморефлексія; вибірковий аудит викладача (20–30 % робіт). Для курсу математичного аналізу розроблено приклади реалізації моделі (границя функції двох змінних, екстремум, подвійний інтеграл) та три варіанти підсумкового оцінювання.

Висновки. Модель переміщує проблему з площини виявлення використання штучного інтелекту в площину валідності оцінювання, формуючи кілька взаємодоповнювальних джерел доказів компетентності. Вибірковий аудит є масштабованішою альтернативою суцільному індивідуальному захисту робіт. Модель особливо перспективна для підготовки вчителів математики. Подальші дослідження мають емпірично перевірити вплив на метакогнітивну активність та оцінити співвідношення між валідністю й ресурсними витратами.

КЛЮЧОВІ СЛОВА: генеративний штучний інтелект; взаємне рецензування; валідність оцінювання; метакогнітивна пасивність; когнітивно-верифікаційний підхід; математичний аналіз; дистанційне навчання; ризик-орієнтований аудит.

ABSTRACT

Problem statement. Generative artificial intelligence undermines the diagnostic reliability of homework assignments in distance mathematics education: an externally correct result is no longer sufficient evidence of a student's independent achievement of learning outcomes. The key risk is students' metacognitive passivity, understood as delegating to technology the analysis of the problem conditions, the choice of method, and the verification of the result. This article extends the author's cognitive-verification approach by shifting verification activity from individual self-checking to the analysis of another student's mathematical reasoning through peer review.

Materials and methods. The study is theoretical, with a pilot trial. The theoretical framework is based on Messick's concept of assessment validity, the notion of metacognitive passivity, and the concept of assessment twins. On this basis, a five-component risk-oriented assessment model was constructed. The pilot involved one academic group of 10 students, one mathematical analysis task, and a qualitative descriptive analysis of the results.

Results. A five-component model was developed: a cognitive-verification solution with checkpoints; structured review of an analogous work according to a rubric; a mandatory author's response to the review; brief self-reflection; and selective instructor audit of 20–30% of the submitted works. Examples of model implementation for a mathematical analysis course were designed, including the limit of a function of two variables, extrema, and a double integral, as well as three variants of final assessment.

Conclusions. The model shifts the problem from detecting the use of artificial intelligence to ensuring assessment validity by generating several complementary sources of evidence of competence. Selective audit is a more scalable alternative to the comprehensive individual defence of all submitted works. The model is particularly promising for the training of mathematics teachers. Further research should empirically test its impact on metacognitive activity and evaluate the balance between validity gains and resource costs.

KEYWORDS: generative artificial intelligence; peer review; assessment validity; metacognitive passivity; cognitive-verification approach; mathematical analysis; distance learning; risk-oriented audit.

ДЛЯ ЦИТУВАННЯ: Чкана Я. Від самоперевірки до взаємного рецензування: розвиток когнітивно-верифікаційного підходу до оцінювання математичних завдань в умовах генеративного штучного інтелекту. *Фізико-математична освіта*, 2026. Том 41. № 3. С. 37-44. <https://doi.org/10.31110/fmo2026.v41i3-05>.

FOR CITATION: Chkana, Ya. (2026). From self-assessment to peer review: the development of a cognitive-verification approach to the evaluation of mathematical problems in the context of generative artificial intelligence. *Physical and Mathematical Education*, 41(3), 37-44. <https://doi.org/10.31110/fmo2026.v41i3-05>.

ВСТУП

Поява загальнодоступних систем генеративного штучного інтелекту стала одним із тих технологічних зрушень, які змушують переглядати усталені педагогічні практики. Сьогодні викладач дедалі частіше отримує від студента не стільки свідчення його власної навчальної діяльності, скільки готовий цифровий артефакт — текст, розв'язання, пояснення, — чие походження й когнітивна історія залишаються невідомими. Особливо гостро ця проблема виявляється в дистанційному навчанні, де викладач позбавлений можливості безпосередньо спостерігати за процесом роботи здобувача, а отже, не може з певністю розрізнити самостійне міркування й відповідь, отриману за допомогою цифрового інструменту.

Для математичної освіти ця ситуація є принциповою, оскільки навчальний результат тут не зводиться до правильної числової або символічної відповіді. Справжнє оволодіння математичним аналізом передбачає вміння аналізувати умову задачі, обирати адекватний метод, перевіряти умови застосування теорем, будувати логічний ланцюг міркувань, інтерпретувати отриманий результат і, що особливо важливо, помічати й виправляти помилки — як власні, так і чужі. Генеративний штучний інтелект здатен швидко створити зовнішньо переконливе розв'язання, однак наявність такого розв'язання аж ніяк не гарантує, що студент справді опанував відповідні компетентності.

У попередньому дослідженні автора було запропоновано когнітивно-верифікаційний підхід до трансформації домашніх завдань із математичного аналізу (Chkana, 2026). Його основна ідея полягала в тому, що завдання має містити не лише обчислювальний, а й обов'язкові пояснювальні та перевірочні елементи — вибір методу, обґрунтування ключових переходів, перевірку результату, аналіз можливих помилок, змістову інтерпретацію відповіді. Така структура, як було показано, підвищує когнітивну складність діяльності та зменшує можливість формальне подання готової відповіді без її осмислення. Проте включення цих компонентів не усуває повністю проблеми оцінювання, адже генеративний штучний інтелект здатен створити не лише обчислення, а й пояснення вибору методу, перевірку результату й навіть рефлексивний коментар.

Звідси виникає подвійна суперечність. З одного боку, для підвищення надійності оцінювання потрібно отримувати додаткові докази розуміння студента — наприклад, через індивідуальну усну співбесіду щодо кожної зданої роботи. З іншого боку, в умовах великої кількості студентів і жорстких часових обмежень така процедура стає організаційно нереалістичною, а додаткове навантаження на викладача здебільшого не передбачене в індивідуальному плані роботи. Отже, потрібна модель, яка зберігає діагностичну цінність оцінювання, але не перекладає весь тягар додаткового контролю на викладача.

Саме тому ми пропонуємо розвинути когнітивно-верифікаційний підхід у напрямі від індивідуальної самоперевірки до комплементарного взаємного рецензування. Ідея полягає в тому, щоб зробити аналіз і оцінювання математичного міркування не лише прерогативою викладача, а й повноцінною навчальною діяльністю самих здобувачів. Рецензуючи роботу іншого, здобувач змушений застосовувати ті самі критерії правильності, що й викладач, але в іншій когнітивній ситуації — коли йому невідома логіка автора, коли він має реконструювати чуже міркування, виявити слабкі місця й аргументовано сформулювати свої зауваження. Взаємне рецензування, таким чином, стає не просто засобом розвантаження викладача, а самостійним механізмом розвитку критичного мислення, рефлексії й метакогнітивної активності.

Метою цієї статті є теоретичне обґрунтування ризик-орієнтованої моделі оцінювання математичних завдань, яка поєднує власне когнітивно-верифікаційне виконання, структуроване взаємне рецензування, відповідь здобувача на рецензію та вибірковий аудит викладача. Така модель, на нашу думку, здатна підвищити валідність оцінювання в умовах генеративного штучного інтелекту, розвивати в здобувачів професійно важливі вміння аналізувати математичні міркування й водночас зберігати реалістичне навантаження на викладача.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Сучасний міжнародний дискурс дедалі виразніше зміщує акцент із питання «Чи використовував студент штучний інтелект?» на питання «Чи дозволяє процедура оцінювання зробити обґрунтований висновок про досягнення результатів навчання?» Цей перехід є надзвичайно важливим, оскільки він переносить проблему з площини виключно академічної доброчесності у площину валідності оцінювальних процедур. Як слушно зауважують Dawson та його колеги (Dawson et al., 2024), навіть якщо неможливо остаточно встановити, чи користувався здобувач генеративним штучним інтелектом, викладач і заклад освіти повинні мати достатні підстави стверджувати, що здобувач досяг відповідних результатів навчання. Автоматичні детектори текстів, створених штучним інтелектом, як свідчать емпіричні дослідження (Weber-Wulff et al., 2023; Perkins et al., 2024; Sun et al., 2026), не забезпечують необхідної надійності й можуть бути легко обійдені редагуванням або перефразуванням; до того ж вони створюють ризик хибних звинувачень. Тому покладатися на них як на основний інструмент контролю було б помилкою.

Натомість перспективнішим видається підхід, який спирається на концепцію валідності, розроблену в класичних працях Messick (Messick, 1989, 1993). Ця концепція передбачає, що валідність оцінювання — це не властивість окремого тесту чи завдання, а ступінь, до якого наявні докази й теоретичні міркування підтримують інтерпретацію результатів оцінювання як показників певного конструкту. У нашому випадку конструктом є математична компетентність здобувача,

а доказом — виконане домашнє завдання. Однак в умовах генеративного штучного інтелекту цей доказ стає неповним або й сумнівним, якщо ми не маємо додаткових джерел інформації про те, як саме здобувач дійшов до свого розв'язання.

Важливим теоретичним поняттям для опису цієї ситуації є метакогнітивна пасивність, яку дослідники (Fan et al., 2025) визначають як схильність здобувача делегувати зовнішньому інструментові (зокрема генеративному штучному інтелекту) не лише технічне виконання, а й планування, моніторинг, оцінювання й корекцію власної роботи. У математичному навчанні це означає, що здобувач може отримати правильний кінцевий результат, але при цьому не виконати тих когнітивних дій, які власне й формують математичне мислення: не проаналізувати умову, не обрати метод самостійно, не перевірити умови застосування теореми, не оцінити правдоподібність відповіді. Таким чином, метакогнітивна пасивність є не просто проблемою академічної доброчесності, а фундаментальним ризиком для самого навчання, адже здобувач позбавляється можливості розвивати ті навички, які становлять суть математичної освіти.

Саме тому оцінювання в умовах генеративного штучного інтелекту має ґрунтуватися не на одному артефакті, а на сукупності взаємопов'язаних доказів. Оригінальну ідею в цьому напрямі пропонують Roe, Perkins і Giray, які вводять концепцію *assessment twins* — двох спеціально розроблених взаємозалежних компонентів оцінювання, спрямованих на ті самі результати навчання, але використовують різні способи отримання доказів і виконуються достатньо близько в часі (Roe, Perkins, & Giray, 2026). Така пара дозволяє перехресно перевірити результати й знизити відому вразливість, зокрема можливість виконання завдання за допомогою генеративного штучного інтелекту. Автори концепції розглядають широкий спектр можливих пар: письмова робота й усна співбесіда, домашнє завдання й контрольоване практичне виконання, звіт і групове обговорення тощо.

Наша модель посідає особливе місце в цьому контексті. Взаємне рецензування, яке ми пропонуємо як комплементарний компонент, не є просто ще одним завданням на ту саму тему. Воно оцінює ті самі результати навчання (здатність аналізувати, застосовувати критерії, виявляти помилки, обґрунтовувати висновки), але робить це через інший вид діяльності — аналіз чужого міркування. Така діяльність, як показують систематичні огляди та метааналізи (Kerman et al., 2024; Zhan et al., 2023), здатна підтримувати мислення вищого порядку, оскільки вона вимагає не простого відтворення знань, а застосування критеріїв, порівняння, оцінювання й аргументації. Особливо цінним є те, що в математичній освіті рецензування має професійну автентичність, особливо для майбутніх учителів математики, адже вміння аналізувати учнівські помилки, ставити діагностичні запитання й пропонувати корекцію є невід'ємною частиною їхньої майбутньої діяльності (de-Armas-González et al., 2023; Biton, 2025).

МЕТОДИ ДОСЛІДЖЕННЯ

Для досягнення поставленої мети — теоретичного обґрунтування ризик-орієнтованої моделі оцінювання — у роботі використано комплекс загальнонаукових і педагогічних методів.

На теоретичному рівні застосовано метод критичного аналізу наукової літератури (зокрема, праць із валідності оцінювання, метакогнітивної пасивності, генеративного штучного інтелекту в освіті, а також систематичних оглядів із взаємного рецензування). Це дало змогу визначити ключові ризики традиційних форм контролю та обґрунтувати необхідність переходу від індивідуальної самоперевірки до комплементарного рецензування. Використано метод концептуального моделювання, за допомогою якого синтезовано чотири обов'язкові та один вибірковий компоненти в цілісну процедуру; побудовано структурно-логічну схему (рис. 1) та описано різні варіанти адитивного, порогового та підтверджувального оцінювання. Метод аналогії та узагальнення залучено для перенесення принципів перехресної верифікації (Roe, Perkins, & Giray, 2026) у контекст математичних завдань, а також для розробки типових прикладів реалізації моделі в курсі математичного аналізу.

На емпіричному рівні проведено пілотну апробацію моделі в одній академічній групі (N=10) дистанційної форми навчання в межах одного завдання з теми «Екстремуми функції двох змінних». Пілот мав розвідувальний характер і був спрямований на перевірку принципової працездатності процедури, а не на статистичне підтвердження гіпотез. Збір даних здійснено за допомогою трьох інструментів:

аналіз виконання рецензій — оцінено, чи виявляють студенти закладені типові помилки в запропонованих викладачем рецензентам розв'язаннях (критерій — наявність аргументованого зазначення пропущеної достатньої умови, неправильного застосування критерію або пропущеного якобіана);

коротке анонімне опитування з відкритими питаннями — два питання: «Чи змусило вас рецензування чужої роботи уважніше поставитися до власного розв'язання?» та «Із якими труднощами ви зіткнулися під час рецензування?»;

спостереження викладача за діалогом «автор — рецензент» — фіксувалося, чи наводить автор математичні аргументи у відповіді на рецензію, чи обмежується формальною згодою.

Обробка отриманих даних проводилася якісними методами: контент-аналіз відповідей на відкриті запитання, порівняльний аналіз рецензій із еталонними викладацькими очікуваннями, тематичне кодування труднощів, що виникали в студентів. Кількісні показники не обчислювалися через малий об'єм вибірки та розвідувальний характер пілоту. Отримані якісні результати використано для корекції критеріїв і уточнення процедури аудиту перед наступними етапами дослідження.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Когнітивно-верифікаційний підхід, запропонований у попередній роботі (Chkana, 2026), спирався на ідею, що домашнє завдання має стати не лише тренувальною вправою, а й засобом формування самоконтролю та рефлексії. Здобувач, виконуючи таке завдання, не просто отримує відповідь, а й пояснює вибір методу, обґрунтовує ключові кроки, перевіряє результат, аналізує можливі помилки й інтерпретує відповідь у змістовому контексті. Однак, як уже зазначалося, генеративний штучний інтелект здатен запропонувати не тільки розв'язання, але й його перевірку та пояснення. Тому

наступним логічним кроком є перенесення верифікаційної діяльності на новий об'єкт — математичне міркування іншої особи.

Такий перехід є не просто технічним ускладненням, а якісною зміною когнітивної ситуації. Коли здобувач перевіряє власну роботу, він часто спирається на пам'ять про свої дії, на інтуїцію, на знання того, «як мало бути». Коли ж він рецензує чужу роботу, він не має цих підказок; він змушений реконструювати логіку невідомого автора, зіставляти її з математичними критеріями й формулювати зовнішнє аргументоване судження. Це значно складніше завдання, яке активізує саме ті метакогнітивні процеси, які часто залишаються пасивними при використанні генеративного штучного інтелекту (Fan et al., 2025).

Варто підкреслити, що запропонована модель не скасовує й не применшує значення попереднього етапу. Навпаки, якісне рецензування неможливе без сформованих умінь пояснювати й перевіряти власне розв'язання. Здобувач, який не навчився верифікувати власну роботу, навряд чи зможе аргументовано оцінити чужу. Отже, два етапи — самоперевірка й перехресне рецензування — органічно доповнюють один одного, утворюючи єдину систему формування математичної компетентності.

Запропонована нами модель складається з чотирьох обов'язкових компонентів і одного вибіркового, які разом утворюють цілісну процедуру, здатну забезпечити надійні докази досягнення результатів навчання навіть в умовах широкої доступності генеративного штучного інтелекту.

Першим компонентом є власне *когнітивно-верифікаційне розв'язання*. Здобувач отримує математичне завдання, яке містить не лише вимогу отримати числову або символічну відповідь, а й низку контрольних точок, що фіксують логіку міркування. Залежно від теми, це можуть бути: визначення типу задачі, вибір методу, пояснення умов застосування теореми, виконання розв'язання з проміжними викладками, проведення перевірки, формулювання змістового висновку тощо. Наприклад, у задачі на дослідження екстремуму функції двох змінних здобувач повинен подати частинні похідні першого порядку, критичні точки, другі похідні, дискримінант другого диференціала, висновок про характер точки та коротку геометричну інтерпретацію. Така структура робить завдання значно менш вразливим до простого копіювання згенерованої відповіді, оскільки воно вимагає осмисленого членування процесу.

Другий компонент — це *комплементарне рецензування*. Здобувач отримує не довільну роботу з будь-якої теми, а розв'язання, структурно аналогічне його власному завданню й спрямоване на ті самі результати навчання. Це принципова умова: якщо здобувач розв'язував задачу на умовний екстремум, а рецензує задачу на криволінійний інтеграл, така діяльність може бути навчальною корисною, але вона не створює комплементарного доказу того самого результату. Рецензування виконується за чітким критерієм, яка спрямовує увагу здобувача на ключові аспекти математичної правильності: розуміння умови, відповідність методу, перевірку передумов, логічну коректність, обчислювальну точність, повноту, наявність перевірки результату, обґрунтованість висновку. Важливо, що рецензент не виставляє остаточної оцінки здобувачу; його завдання — надати аргументований аналіз і, за можливості, сформулювати уточнювальне запитання та конкретну рекомендацію щодо покращення.

Третій компонент — *відповідь здобувача на отриману рецензію*. Здобувач має коротко, але змістовно відреагувати на зауваження: вказати, з якими з них він погоджується й які виправлення вносить, а з якими не погоджується й чому, наводячи математичні аргументи. Цей етап є критично важливим з двох причин. По-перше, він показує, чи здатен здобувач оцінити якість отриманого зворотного зв'язку, тобто чи володіє він критичним ставленням не лише до власної роботи, а й до чужих суджень. По-друге, він запобігає ситуації, коли помилкова або поверхнева рецензія автоматично сприймається як істина. Діалог між здобувачем і рецензентом перетворює процедуру на живу комунікацію, а не на одноразове коментування.

Четвертий компонент — *коротка саморефлексія*, яка для великих груп може бути зведена до кількох запитань: який крок розв'язання був ключовим, де можливе слабе місце, чи використовувався штучний інтелект і для якої операції, що здобувач перевіряв самостійно. Цей блок не повинен перетворюватися на великий текст, який також легко згенерувати; його функція — зафіксувати позицію здобувача перед рецензуванням і створити додаткову точку відліку для викладача (Combrinck & Loubser, 2025).

Нарешті, п'ятий, вибіркового компонента — це *аудит викладача*. Викладач не проводить індивідуальної співбесіди з кожним здобувачем, а здійснює перевірку за комбінованим принципом: випадково обрані роботи, роботи з істотними суперечностями між розв'язанням, рецензією та відповіддю здобувача, рецензії без аргументації, роботи з нетиповим методом, випадки різкого розходження між поточною навчальною активністю й рівнем поданої роботи. Випадковий компонент є необхідним для справедливості: якщо перевіряються лише «підозрілі» здобувачі, процедура може сприйматися як каральна й упереджена. Аудит може мати форму одного-двох коротких запитань, що стосуються ключового переходу, зміненої умови або виправлення конкретної помилки (рис. 1).

Щоб зробити запропоновану модель більш наочною, розглянемо кілька конкретних прикладів із курсу математичного аналізу, які ілюструють, як одне й те саме завдання може бути подане для виконання й для рецензування.

Перший приклад стосується теми границі функції двох змінних. Здобувачеві пропонується дослідити існування границі $\lim_{(x,y) \rightarrow (0;0)} \frac{x^2 - y^2}{x^2 + y^2}$. У розв'язанні він має перевірити принаймні два різні шляхи наближення, зробити висновок

про існування або неіснування границі й пояснити, чому одного шляху недостатньо. Рецензент, у свою чергу, отримує розв'язання, в якому інший здобувач, скажімо, перевіряв лише шлях вздовж прямої $y = x$, отримав нуль і зробив висновок, що границя дорівнює нулю. Завдання рецензента — виявити логічну помилку, запропонувати два інші шляхи (наприклад, $y = 0$ та $y = 2x$), сформулювати правильний висновок про неіснування границі й пояснити, чому одного шляху достатньо для спростування, але недостатньо для доведення. Цей приклад добре демонструє, як рецензування перевіряє не лише знання техніки, а й розуміння логіки доведення.



Рис. 1. Структура ризик-орієнтованої моделі оцінювання (п'ять компонентів)

Джерело: згенеровано за допомогою ШІ (<https://claude.ai>) на основі авторського опису

Другий приклад — дослідження екстремуму функції двох змінних, наприклад, $f(x, y) = x^2 - y^2$. Здобувач має знайти частинні похідні, критичну точку, обчислити другі похідні, застосувати критерій другого порядку й зробити висновок. Рецензент отримує розв'язання, в якому здобувач знайшов частинні похідні, прирівняв їх до нуля й зробив висновок, що в точці $(0,0)$ функція має мінімум, не перевірявши достатніх умов. Рецензент має оцінити достатність умови (рівність нулю частинних похідних першого порядку), обчислити другі похідні, застосувати критерій, пояснити сідловий характер точки й запропонувати геометричну або траєкторну перевірку. Тут перевіряється не техніка диференціювання, а розуміння різниці між необхідною й достатньою умовами екстремуму.

Третій приклад — обчислення подвійного інтеграла $\iint_{x^2+y^2 \leq 1} (x^2 - y^2) dx dy$. Здобувач має виконати перехід до

полярних координат, правильно записати межі інтегрування, не забути про якобіан і виконати обчислення. Рецензент отримує розв'язання, де здобувач перейшов до полярних координат, але записав інтеграл без множника r . Завдання рецензента — знайти пропущений елемент, пояснити роль якобіана, виправити інтеграл, обґрунтувати межі й сформулювати геометричний сенс множника r . Цей приклад ілюструє, як рецензування виявляє не лише обчислювальні помилки, а й концептуальні прогалини.

У кожному з цих випадків рецензування перевіряє ті самі результати навчання, що й власне виконання, але в іншому режимі діяльності — через аналіз чужого міркування, що вимагає глибшого розуміння й критичного підходу.

Підсумкова оцінка за виконану роботу може формуватися на основі трьох компонентів. Найпростіший варіант — адитивна модель, де 40 % ваги припадає на власне розв'язання, 40 % — на якість рецензії чужої роботи й 20 % — на відповідь здобувача на отриману рецензію. Такий розподіл зменшує домінування готового продукту й стимулює здобувача серйозно ставитися до рецензування. Водночас можна застосувати порогову модель, за якої здобувач отримує високу оцінку за власне розв'язання лише за умови досягнення мінімального рівня за рецензування та відповідь на feedback. Це підкреслює обов'язковість аналітичної компетентності. Нарешті, можлива підтверджувальна модель, де рецензія й відповідь здобувача не додають значної кількості балів, але слугують підтвердженням того, що результат первинної роботи можна інтерпретувати як свідчення досягнення результату навчання. Для перших пілотних досліджень доцільно використовувати адитивну модель із порогом для обов'язкового аудиту, що зменшує ризик надмірно жорстких наслідків на етапі, коли здобувачі ще навчаються рецензувати.

Щодо практичного впровадження, варто дотримуватися кількох рекомендацій. По-перше, завдання мають містити контрольні точки, а не обмежуватися кінцевою відповіддю. По-друге, рецензія повинна бути обов'язково структурованою — без чітких запитань здобувачі схильні писати формальні коментарі на кшталт «усе правильно». По-третє, якість рецензії має впливати на оцінку, інакше здобувачі не сприйматимуть цей етап серйозно. По-четверте, відповідь здобувача на рецензію є обов'язковою — це перетворює реєв'ю на діалог. По-п'яте, аудит викладача має бути вибірковою, але прозорим: здобувачі повинні знати критерії відбору робіт для перевірки. На початковому етапі доцільно пропонувати здобувачам для рецензування не реальні роботи, а спеціально підготовлені викладачем розв'язання з типовими помилками — це допоможе сформувати культуру математичного рецензування. І нарешті, всі процедури слід чітко прописати в силабусі, щоб здобувачі заздалегідь знали, що оцінюється не лише розв'язання, а й здатність аналізувати й рецензувати математичне міркування.

Нами було проведено попередню (пілотну) апробацію запропонованої моделі в межах курсу математичного аналізу для дистанційної форми навчання. Пілот охоплював одну академічну групу (10 студентів) протягом одного завдання на тему, для якої вже розроблено приклад реалізації моделі (зокрема, дослідження екстремуму функції двох

змінних, описане вище). Цикл містив всі чотири обов'язкові компоненти (власне розв'язання, комплементарне рецензування, відповідь здобувача на рецензію та коротку саморефлексію), а вибірковий аудит викладача був застосований до 30 % робіт за комбінованим принципом, описаним вище.

Оцінювання результатів пілоту мав описовий, якісний характер і не передбачав статистичної перевірки гіпотез, що відповідає масштабу й меті пілотного етапу. Як орієнтовні джерела доказів було використано: (1) аналіз того, чи виявляють рецензенти закладені в розв'язаннях типові помилки (наприклад, неперевірену достатню умову екстремуму) — цей показник прямо відповідає меті моделі; (2) коротке анонімне опитування здобувачів із двома-трьома відкритими питаннями про те, чи змусило рецензування чужої роботи уважніше поставитися до власного розв'язання, і з якими труднощами вони зіткнулися під час рецензування; (3) спостереження викладача за якістю діалогу “здобувач – рецензент” на етапі відповіді на рецензію (зокрема, чи наводить здобувач математичні аргументи, чи просто погоджується або відмовчується).

Очікуваним позитивним результатом пілоту є підтвердження принципової працездатності моделі: здатність хоча б частини рецензентів самостійно виявляти закладені помилки без підказки викладача, а також виникнення змістовного діалогу на етапі відповіді автора замість формального “погоджуюсь”. Такий результат, навіть без статистичного узагальнення, слугуватиме первинним підтвердженням того, що комплементарне рецензування активізує саме ті метакогнітивні дії, на розвиток яких орієнтована модель, і створить підставу для розширення вибірки та переходу до повноцінного емпіричного дослідження з кількісними показниками.

ОБГОВОРЕННЯ ТА ОБМЕЖЕННЯ

Запропонована модель розвиває когнітивно-верифікаційний підхід у трьох суттєвих напрямках. По-перше, вона переходить від одного джерела доказів до кількох: власне розв'язання показує, що здобувач може побудувати математичне міркування, а рецензування — що він може застосувати критерії правильності до іншого випадку; відповідь на feedback демонструє здатність оцінити критику й переглянути власну позицію. По-друге, модель виводить проблему з площини виявлення використання штучного інтелекту в площину валідності: її завданням є не доведення того, хто саме створив кожен рядок первинної роботи, а отримання достатньої сукупності доказів того, що здобувач володіє заявленою математичною компетентністю (Dawson et al., 2024; Roe, Perkins, & Giray, 2026). По-третє, модель пропонує масштабованішу альтернативу суцільному усному захисту: вибірковий аудит не усуває додаткового навантаження, але спрямовує його на випадки, де суперечності між компонентами роблять оцінювальне судження недостатньо обґрунтованим.

Водночас слід чітко усвідомлювати, що взаємне рецензування не є автоматично стійким до генеративного штучного інтелекту. Здобувач може передати чужу роботу системі й отримати готовий feedback. Отже, сильна сторона моделі полягає не в технічній неможливості використання штучного інтелекту, а в багатокomпонентності процедури, структурній відповідності завдань, необхідності реагувати на feedback і можливості вибіркового аудиту. Якщо здобувач використовує генеративний штучний інтелект для написання рецензії, це може виявитися під час аудиту, особливо якщо рецензія містить загальні фрази без конкретних посилань на математичні кроки. Однак повністю запобігти такому використанню неможливо, тому модель має ґрунтуватися на педагогічному дизайні, а не на технічних обмеженнях (Corbin, Dawson, & Liu, 2025; Corbin et al., 2025).

Серед обмежень моделі варто назвати обмежений масштаб емпіричної перевірки — на цьому етапі вона спирається лише на одноразовий пілот без статистичного узагальнення, тоді як повноцінна перевірка на більшій вибірці й з кількісними показниками валідності ще попереду. Якість рецензування залежить від рівня підготовки здобувачів: слабший може не помітити помилок у роботі сильнішого, або, навпаки, необґрунтовано їх приписати. Тому необхідно проводити навчання рецензуванню, використовувати еталонні приклади, починати з підготовлених викладачем розв'язань і забезпечувати вибіркoву модерацію (Topping et al., 2025). Нерівноцінність робіт також створює виклик: майже безпомилкове або хаотичне розв'язання може створювати різні когнітивні можливості для рецензента, тому розподіл робіт має враховувати складність і типи помилок. Крім того, модель потребує ресурсів: викладач має підготувати аналогічні завдання, створити критерії, організувати розподіл, перевірити частину рецензій і розв'язати конфліктні випадки. Ця робота має визнаватися складником педагогічного навантаження. Нарешті, варто враховувати питання справедливості й інклюзивності: здобувачі можуть мати різний досвід отримання критичного зворотного зв'язку, мовну підготовку, рівень тривожності й комунікативні особливості. Необхідними є анонімізація, правила коректної комунікації, альтернативні формати й прозорі критерії.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШОГО ДОСЛІДЖЕННЯ

Генеративний штучний інтелект змінює не лише способи виконання навчальних завдань, а й саму обґрунтованість висновків, які викладач робить на основі готового продукту. У дистанційному математичному навчанні правильне й належно оформлене розв'язання вже не може автоматично вважатися достатнім доказом досягнення результату навчання. Натомість оцінювання має будуватися на сукупності взаємопов'язаних доказів, які в різний спосіб засвідчують здатність здобувача аналізувати, застосовувати, перевіряти й аргументувати математичні міркування.

Запропонована в цій статті модель є послідовним розвитком когнітивно-верифікаційного підходу (Chkana, 2026): верифікаційна діяльність переноситься з власного розв'язання на аналіз структурно аналогічного чужого міркування. Комплементарне взаємне рецензування не є прямим доказом авторства первинної роботи й не замінює викладацького оцінювання. Його значення полягає у створенні другого джерела доказів — здатності здобувача застосовувати математичні критерії, виявляти помилки, обґрунтовувати оцінку й пропонувати виправлення. Поєднання власного когнітивно-верифікаційного завдання, структурованої рецензії, відповіді здобувача та вибіркового ризик-орієнтованого

аудиту потенційно посилює валідність оцінювання й створює масштабованішу альтернативу обов'язковому індивідуальному захисту кожної роботи.

Водночас модель не є універсальним або остаточним розв'язанням проблеми генеративного штучного інтелекту. Її ефективність залежить від узгодження результатів навчання, якості критеріїв, підготовки здобувачів, способу розподілу робіт, прозорості аудиту та інституційного врахування навантаження викладача (Corbin, Dawson, & Liu, 2025). Особливо перспективною вона видається для підготовки майбутніх учителів математики, для яких рецензування чужих розв'язань є професійно автентичною діяльністю (de-Armas-González et al., 2023; Biton, 2025).

Подальші дослідження мають бути спрямовані на повномасштабну емпіричну перевірку моделі — оцінювання її впливу на математичне міркування, метакогнітивну активність здобувачів, а також визначення реального співвідношення між підвищенням валідності та додатковими ресурсними витратами. Лише після такої перевірки можна буде з упевненістю говорити про практичну доцільність і межі застосування цієї моделі в різних освітніх контекстах.

КОНФЛІКТ ІНТЕРЕСІВ

Автор підтверджує відсутність фінансових, особистих чи інших інтересів, що можуть розглядатися як потенційний конфлікт інтересів щодо публікації цієї статті.

ДЖЕРЕЛА ФІНАНСУВАННЯ

Робота виконана за відсутності фінансової підтримки з боку будь-яких організацій.

ДОСТУПНІСТЬ ДАНИХ

Дані можуть бути надані за обґрунтованим запитом відповідному автору.

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ

Засоби ШІ, зокрема, claude.ai, було використано для візуалізації моделі запропонованого в статті когнітивно-верифікаційного підходу та ChatGPT для поліпшення якості мови. Автор критично перевірів і відредагував отриманий контент та несе повну відповідальність за його зміст.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
2. Biton, Y. (2025). Learning mathematics through peer assessment: "How can we assess something that we ourselves don't know how to solve?" *EURASIA Journal of Mathematics, Science and Technology Education*. <https://www.ejmste.com/article/learning-mathematics-through-peer-assessment-how-can-we-assess-something-that-we-ourselves-dont-know-15794>
3. Combrinck, C., & Loubser, N. (2025). Student self-reflection as a tool for managing GenAI use in large class assessment. *Discover Education*, 4, Article 72. <https://link.springer.com/article/10.1007/s44217-025-00461-2>
4. Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2025.2553340>
5. Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, 50(7), 1087–1097. <https://doi.org/10.1080/02602938.2025.2503964>
6. Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005–1016. <https://doi.org/10.1080/02602938.2024.2386662>
7. de-Armas-González, P., Perdomo-Díaz, J., & Sosa-Martín, D. (2023). Peer assessment processes in a problem-solving activity with future teachers. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(4), Article em2245. <https://www.ejmste.com/article/peer-assessment-processes-in-a-problem-solving-activity-with-future-teachers-13057>
8. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
9. Gao, X., Noroozi, O., Gulikers, J., Biemans, H. J. A., & Banihashem, S. K. (2024). A systematic review of the key components of online peer feedback practices in higher education. *Educational Research Review*, 42, 100588. <https://doi.org/10.1016/j.edurev.2024.100588>
10. Sun, Y., Liao, Y., & Ma, X. (2026). Trusting AI to detect AI? A systematic evaluation of the reliability and robustness of current AIGC detection tools for student academic work. *Computers & Education*, 249, 105616. <https://doi.org/10.1016/j.compedu.2026.105616>
11. Heil, J., & Ifenthaler, D. (2023). Online assessment in higher education: A systematic review. *Online Learning*, 27(1), 187–218. <https://olj.onlinelearningconsortium.org/index.php/olj/article/view/3398>
12. Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies*, 29(1), 763–813. <https://doi.org/10.1007/s10639-023-12273-8>
13. Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). *American Council on Education*.
14. Messick, S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. *ETS Research Report Series*, 1993(2), i–18. <https://doi.org/10.1002/j.2333-8504.1993.tb01562.x>
15. Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughan, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education*, 21, Article 53. <https://doi.org/10.1186/s41239-024-00487-w>
16. Roe, J., Perkins, M., & Giray, L. (2026). Assessment twins: An approach for strengthening assessment validity in the age of generative AI. *Journal of Applied Learning & Teaching*, 9(2). <https://doi.org/10.37074/jalt.2026.9.2.3>
17. Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>

18. Topping, K. J., Gehringer, E., Khosravi, H., Gudipati, S., Jadhav, K., & Susarla, S. (2025). Enhancing peer assessment with artificial intelligence. *International Journal of Educational Technology in Higher Education*, 22, Article 3. <https://doi.org/10.1186/s41239-024-00501-1>
19. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19, Article 26. <https://doi.org/10.1007/s40979-023-00146-z>
20. Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology*, 54(4), 817–835. <https://doi.org/10.1111/bjet.13310>
21. Чкана, Я. (2026). Когнітивно-верифікаційний підхід до трансформації домашніх завдань з математичного аналізу в умовах використання штучного інтелекту. *Фізико-математична освіта*, 41(2), 86–93. <https://doi.org/10.31110/fmo2026.v41i2-08>

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
2. Biton, Y. (2025). Learning mathematics through peer assessment: "How can we assess something that we ourselves don't know how to solve?" *EURASIA Journal of Mathematics, Science and Technology Education*. <https://www.ejmste.com/article/learning-mathematics-through-peer-assessment-how-can-we-assess-something-that-we-ourselves-dont-know-15794>
3. Combrinck, C., & Loubser, N. (2025). Student self-reflection as a tool for managing GenAI use in large class assessment. *Discover Education*, 4, Article 72. <https://link.springer.com/article/10.1007/s44217-025-00461-2>
4. Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*. *Advance online publication*. <https://doi.org/10.1080/02602938.2025.2553340>
5. Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, 50(7), 1087–1097. <https://doi.org/10.1080/02602938.2025.2503964>
6. Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005–1016. <https://doi.org/10.1080/02602938.2024.2386662>
7. de-Armas-González, P., Perdomo-Díaz, J., & Sosa-Martín, D. (2023). Peer assessment processes in a problem-solving activity with future teachers. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(4), Article em2245. <https://www.ejmste.com/article/peer-assessment-processes-in-a-problem-solving-activity-with-future-teachers-13057>
8. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
9. Gao, X., Noroozi, O., Gullikers, J., Biemans, H. J. A., & Banihashem, S. K. (2024). A systematic review of the key components of online peer feedback practices in higher education. *Educational Research Review*, 42, 100588. <https://doi.org/10.1016/j.edurev.2024.100588>
10. Sun, Y., Liao, Y., & Ma, X. (2026). Trusting AI to detect AI? A systematic evaluation of the reliability and robustness of current AIGC detection tools for student academic work. *Computers & Education*, 249, 105616. <https://doi.org/10.1016/j.compedu.2026.105616>
11. Heil, J., & Ifenthaler, D. (2023). Online assessment in higher education: A systematic review. *Online Learning*, 27(1), 187–218. <https://oli.onlinelearningconsortium.org/index.php/oli/article/view/3398>
12. Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies*, 29(1), 763–813. <https://doi.org/10.1007/s10639-023-12273-8>
13. Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). *American Council on Education*.
14. Messick, S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. *ETS Research Report Series*, 1993(2), i–18. <https://doi.org/10.1002/j.2333-8504.1993.tb01562.x>
15. Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughan, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education*, 21, Article 53. <https://doi.org/10.1186/s41239-024-00487-w>
16. Roe, J., Perkins, M., & Giray, L. (2026). Assessment twins: An approach for strengthening assessment validity in the age of generative AI. *Journal of Applied Learning & Teaching*, 9(2). <https://doi.org/10.37074/jalt.2026.9.2.3>
17. Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>
18. Topping, K. J., Gehringer, E., Khosravi, H., Gudipati, S., Jadhav, K., & Susarla, S. (2025). Enhancing peer assessment with artificial intelligence. *International Journal of Educational Technology in Higher Education*, 22, Article 3. <https://doi.org/10.1186/s41239-024-00501-1>
19. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19, Article 26. <https://doi.org/10.1007/s40979-023-00146-z>
20. Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology*, 54(4), 817–835. <https://doi.org/10.1111/bjet.13310>
21. Чкана, Я. (2026). Когнітивно-верифікаційний підхід до трансформації домашніх завдань з математичного аналізу в умовах використання штучного інтелекту [A cognitive-verification approach to transforming homework in mathematical analysis in the context of artificial intelligence use]. *Фізико-математична освіта – Physical and Mathematical Education*, 41(2), 86–93. <https://doi.org/10.31110/fmo2026.v41i2-08> (in Ukrainian).

| Матеріал надійшов до редакції: 01.04.2026 р. | Прийнято до друку: 25.05.2026 р. | Опубліковано: 30.06.2026 р. |



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