


 <https://doi.org/10.31651/2524-2660-2026-1-12-26>

 <https://orcid.org/0000-0003-4047-1301>

СКВОРЦОВА Світлана


докторка педагогічних наук, професорка, професорка кафедри математики та методики її навчання,
Південноукраїнський національний педагогічний університет імені К.Д. Ушинського


 skvo08@i.ua

 <https://orcid.org/0000-0001-5963-0451>

СИМОНЕНКО Тетяна


докторка педагогічних наук, професорка,
професорка кафедри методики навчання, стилістики та культури української мови,
Черкаський національний університет імені Б. Хмельницького;
директорка Центру українських дослідників у Австрії,
афілійована професорка департаменту стратегічних комунікацій,
Віденський університет економіки та бізнесу


 tetianasymonenko4@gmail.com

 <https://orcid.org/0000-0002-5226-840X>

ГНЕЗДІЛОВА Кіра

докторка педагогічних наук, професорка, професорка кафедри початкової і спеціальної освіти,
Черкаський національний університет імені Б. Хмельницького

 kiragnez@gmail.com

 <https://orcid.org/0000-0003-1092-2116>

НЕДЯЛКОВА Катерина

кандидатка педагогічних наук, доцентка кафедри математики і методики її навчання,
Південноукраїнський національний педагогічний університет імені К. Д. Ушинського

 Niedalkova.KV@pdpu.edu.ua

УДК 378.091.227.7:[004.738.1Peersch0lar:004.8](045)

**ПЛАТФОРМИ ВЗАЄМООЦІНЮВАННЯ З ПІДТРИМКОЮ ШТУЧНОГО ІНТЕЛЕКТУ
У МЕТОДИЧНІЙ ПІДГОТОВЦІ МАЙБУТНІХ УЧИТЕЛІВ МАТЕМАТИКИ:
ПОРІВНЯЛЬНИЙ АНАЛІЗ ТА АПРОБАЦІЯ PEERSCHOLAR**

Мета дослідження – систематичний порівняльний аналіз платформ взаємооцінювання з підтримкою штучного інтелекту та оцінювання ефективності PeerScholar у розвитку оцінювальної компетентності майбутніх учителів математики в умовах української вищої освіти.

Використано змішаний дизайн дослідження, що поєднує порівняльний аналіз семи III-платформ взаємооцінювання та емпіричну апробацію PeerScholar з участю 56 студентів спеціальності «Середня освіта (математика)» протягом осіннього семестру 2024/2025 н.р. Збір даних здійснювався через структуроване анкетування (32 питання за шкалою Лайкерта), напівструктуровані інтерв'ю (n=6) та аналіз об'єктивних метрик платформи. Статистична обробка включала описову статистику, кореляційний аналіз (р Спірмена) та оцінку внутрішньої узгодженості (а Кронбаха=0.89).

Порівняльний аналіз платформ взаємооцінювання виявив диференційовані рішення EdTech від адаптивних систем (ALEKS) до комплексних рефлексивних екосистем (PeerScholar, RiPPL), з критичною роллю економічної доступності для українських ЗВО.

Апробація PeerScholar засвідчила статистично значуще позитивне сприйняття студентами методичного (M=4.36, SD=0.61) та професійно-рефлексивного (M=4.35, SD=0.65) аспектів платформи.

Виявлено сильні кореляції між виявленням альтернативних підходів та цінністю формування зворотного зв'язку ($r_s=0.913$, $p<0.01$), а також між розвитком методичних та рефлексивних навичок ($r_s=0.713$, $p<0.05$).

Міжрейтерська надійність зросла від $r=0.51$ до $r=0.71$ протягом семестру, підтверджуючи поступовий розвиток оцінювальної компетентності. Студенти виявили вищу зацікавленість у конструктивному зворотному зв'язку (M=4.81) порівняно з отриманням балів (M=4.42), що свідчить про запит на професійну комунікацію.

Отже, інтеграція PeerScholar створює якісне середовище для професійного діалогу, де взаємооцінювання стимулює перехід від рутинного розв'язування задач до глибокої методичної рефлексії. Виявлено послідовність: виявлення альтернативних підходів → критичний аналіз → запит на якісний зворотний зв'язок. Часові витрати (M=3.40) вимагають коригування планування, а технологічний бар'єр англійської інтерфейсу знижує ефективність для частини студентів.

Дослідження пропонує практичні рекомендації щодо вибору платформ для закладів з обмеженими ресурсами та впровадження структурованого взаємооцінювання у підготовці вчителів.

Ключові слова: взаємооцінювання студентів; штучний інтелект в освіті; PeerScholar; оцінювальна компетентність; підготовка вчителів математики; цифрова трансформація оцінювання; рефлексивне навчання; peer assessment; методична компетентність.

Постановка проблеми. Сучасна вища освіта перебуває на етапі фундаментальної перебудови оціночних парадигм, де традиційне оцінювання за принципом «зверху

вниз» дедалі частіше поступається місцем партисипативним – учасницьким моделям, таким як взаємооцінювання (peer assessment) (Topping et al., 2025; Babik et al., 2024). Проблема взаємооцінювання в університетському середовищі є багатогранною, оскільки вона охоплює не лише технічні аспекти організації перевірки робіт, а й глибинні психолого-педагогічні виклики, пов'язані з надійністю оцінок, якістю зворотного зв'язку та розвитком оціночної компетентності студентів. Взаємооцінювання визначається як освітній механізм, за якого студенти оцінюють якість або успішність результатів навчання своїх колег, що мають аналогічний статус (Topping et al., 2025). Ця практика є надзвичайно важливою для формування навичок саморегульованого навчання (SRL), оскільки вона змушує студента не просто споживати контент, а й виступати в ролі критика та експерта (Недялкова, 2025; Pare, Joordens, 2008). Проте впровадження цієї практики стикається з фундаментальними викликами: суб'єктивністю балів, когнітивним навантаженням на викладача та проблемою «порожнього оберненого зв'язку», коли студенти обмежуються загальними фразами без методичної глибини (Topping et al., 2025). Ці обмеження історично стримували масштабованість та довіру до оцінювання за участю однокласників як інструменту підсумкової оцінки.

Поява технологій штучного інтелекту (ШІ), зокрема обробки природної мови (NLP) та алгоритмів машинного навчання, пропонує трансформаційний потенціал для подолання фундаментальних викликів взаємооцінювання за (Guo, 2024; Liu et al., 2024). Сучасні платформи, вдосконалені ШІ, виходять за межі простої автоматизації розподілу робіт (peer-grading – процесу, коли студенти або учні оцінюють роботи одне одного за заздалегідь визначеними критеріями), створюючи всеосяжні інтелектуальні екосистеми, що підтримують численні функції оцінювання:

1. Підвищення якості зворотного зв'язку. ШІ структурує роботу студентів-рецензентів через реальний час аналіз чернеток коментарів, спонукаючи до конкретики, відповідного тону та дієвих пропозицій (Guo, 2024). Системи на кшталт EvaluMate застосовують великі мовні моделі (LLMs) для виявлення «порожнього зворотного зв'язку» – загальних фраз на зразок «good work – гарна робота» – та спрямовують студентів до конструктивної, обґрунтованої доказами критики.

2. Агрегація та калібрування оцінок. Платформи, такі як Peereptiv, використовують складні алгоритми для розрахунку надійності рецензентів на основі історичних

показників та узгодженості з експертними судженнями, що дає змогу зваженої агрегації, де оцінки надійних рецензентів мають більшу вагу (Li, 2023). Це вирішує фундаментальну проблему варіативної якості рецензій.

3. Мета-зворотний зв'язок щодо якості зворотного зв'язку. Замість того, щоб викладачі перевіряли кожен коментар здобувачів освіти, агенти ШІ забезпечують миттєву оцінку якості самого зворотного зв'язку, допомагаючи студентам розвивати навички професійної критики ще до того, як зворотний зв'язок дійде до авторів робіт (Guo, 2024; Liu et al., 2024).

4. Підтримка викладача через аналітику. ШІ виявляє аномалії – значні розбіжності в оцінках або конфлікти в групах – що дає змогу втручатися цільове лише там, де це необхідно (Topping et al., 2025). Це змінює роль викладача з рутинної модерації на стратегічну підтримку студентів, які відстають.

Українські заклади вищої освіти зазнають унікальних викликів впровадження освітніх технологій. По-перше, це обмежені інституційні бюджети, застарілі системи управління навчанням (LMS), мовні бар'єри з переважно англійськими платформами та відсутністю комплексної інфраструктури підтримки цифрової грамотності (Недялкова, 2025). Ці обмеження вимагають ретельної оцінки доступності платформ, зокрема щодо наявності безкоштовних версій, підтримки української мови та складності інтеграції. Попри ці виклики, програми підготовки вчителів математики в Україні потребують інноваційних підходів до оцінювання. Традиційні методи – оцінка планів уроків та методичних аналізів виключно викладачем не розвивають компетентності в оцінюванні, які є невід'ємними для професійної практики майбутніх учителів (Недялкова, 2025). Взаємооцінювання пропонує подвійні переваги: студенти розвивають як предметну експертизу, так і компетентність в оцінюванні, необхідну для їхньої педагогічної кар'єри.

Хоча багато досліджень вивчають ефективність взаємооцінювання у західних університетах (Babik et al., 2024; Topping et al., 2025), емпіричні дані з пострадянських освітніх систем залишаються обмеженими. Крім того, систематичні порівняння платформ, вдосконалених ШІ, з акцентом на доступність для закладів з обмеженими ресурсами, є рідкісними. Більшість наявних оглядів платформ віддають пріоритет функціональності над економічною доступністю, залишаючи освітян у країнах, що розвиваються, без практичних рекомендацій. Дослідження заповнює ці прогалини через

подвійний внесок: 1) систематичний порівняльний аналіз платформ, вдосконалених ІІІ, з ідентифікацією доступних варіантів для закладів з обмеженим бюджетом; 2) емпіричне вивчення впровадження платформи PeerScholar у підготовці вчителів математики в Україні, поєднуючи кількісну оцінку сприйняття студентів з якісним аналізом викликів впровадження.

Мета дослідження полягає у проведенні систематичного порівняльного аналізу платформ взаємооцінювання, що підтримуються ІІІ, та оцінці ефективності платформи PeerScholar у розвитку компетентностей оцінювання серед майбутніх учителів математики в українській вищій освіті.

Дослідницькі питання:

1. Які платформи взаємооцінювання, вдосконалені ІІІ, пропонують доступ до безкоштовних версій, придатні для українських закладів вищої освіти, та як порівнюються їхні функціональні можливості?

2. Як студенти спеціальності «Середня освіта (математика)» сприймають вплив платформи PeerScholar на методичні, організаційні, мотиваційні, технологічні та професійно-рефлексивні аспекти їхнього навчання?

3. Які зв'язки існують між різними аспектами сприйняття платформи та яким чином академічна успішність визначає варіативність цього сприйняття?

4. Які виклики впровадження та фактори успіху характеризують розгортання платформи PeerScholar у контекстах підготовки вчителів в Україні?

Дослідження сприяє як теоретичному розумінню, так і практичному впровадженню взаємооцінювання, вдосконаленого ІІІ. Теоретично воно розширює проблему взаємооцінювання на пострадянські освітні контексти, вивчаючи, як культурні та інституційні фактори впливають на прийняття технологій. Практично воно надає обґрунтовані докази і рекомендації для освітян в умовах обмежених ресурсів, ідентифікуючи доступні платформи та документуючи автентичні досвіди впровадження, виклики та адаптації.

Для підготовки вчителів математики зокрема дослідження демонструє, як взаємооцінювання за участю інших студентів (одногогрупників) розвиває не лише їх предметну експертизу, але й професійні компетентності в оцінюванні – важливі, але часто занедбані виміри підготовки педагогів. Детальний опис впровадження та емпіричні результати пропонують модель, яку можна відтворити в інших українських та міжнародних закладах, що зазнають подібних обмежень.

Виклад основного матеріалу. Взаємооцінювання еволюціонувало через чітко виражені фази розвитку, кожна з яких характеризувалася дедалі складнішим технологічним інтегруванням та педагогічним розумінням. Ранні реалізації – перша фаза (до 2000-х років) зосереджувалися переважно на peer-grading – механічному розподілі завдань оцінювання – з мінімальною технологічною підтримкою понад базові рубрики (Topping et al., 2025; Pare, Joordens, 2008; Joordens et al., 2009). Ці системи трактували взаємооцінювання інструментально як механізм зменшення навантаження, а не як можливість для навчання. Друга фаза (від 2000 до 2015 року) ввела систематичні підходи калібрування. Платформи на кшталт Peerceptiv та CrowdGrader впроваджували вправи калібрування, де студенти спочатку оцінювали зразкові роботи з відомими рівнями якості перед оцінюванням однокласників, з подальшим зважуванням надійності окремих рецензентів (Li, 2023). Дослідження цього періоду встановили потенціал взаємооцінювання для розвитку критичного мислення та метакогнітивної свідомості (Pare, Joordens, 2008). Сучасна фаза (2016 року й дотепер) представляє екосистему взаємооцінювання, вдосконалені ІІІ.

Сучасні платформи інтегрують численні функції ІІІ: обробку природної мови для аналізу якості зворотного зв'язку (Guo, 2024), алгоритми машинного навчання для моделювання надійності рецензентів (Li, 2023), генеративний ІІІ для структурування формулювання зворотного зв'язку (Guo, 2024) та панелі аналітики навчання для нагляду викладача (Topping et al., 2025). Ця еволюція відображає фундаментальне пересмислення: від взаємооцінювання (як заміника оцінки) до потужного механізму навчання, вдосконаленого інтелектуальними технологіями.

Роль ІІІ в подоланні викликів взаємооцінювання студентів. Дослідники (Topping et al., 2025) запропонували комплексну рамку, що ідентифікує шість процесуальних зон, де ІІІ вдосконалює процеси оцінювання за участю студентів:

Зона 1. Призначення рецензентів (Assigning Peer Assessors). Традиційне випадкове призначення часто зумовлює невідповідність рівнів експертизи. Алгоритми ІІІ оптимізують підбір на основі метрик довіри, попередніх показників та доменної експертизи. Платформа RiPPLE (UQ-розробка) як кейс-стаді з AI для призначення та трістої діяльності (створення – creation, перевірка – review, практика – practice); платформа RiPPLE фокусується на знаннях та оцінному судженні студентів (student wisdom) та

автоматизованому зворотному зв'язку AI-feedback (Topping et al., 2025).

Зона 2. Підтримка індивідуальних рецензій (Enhancing Individual Reviews). III структурує процес рецензування через реальний час аналізу та підказки. Чат-бот Eva платформи EvaluMate аналізує чернетки коментарів щодо конкретики, конструктивності та дієвості, надаючи миттєві пропозиції вдосконалення (Guo, 2024). Це вирішує проблему «порожнього зворотного зв'язку», коли студенти надають загальні, некорисні коментарі.

Зона 3. Виведення остаточних балів / зворотного зв'язку (Deriving Peer Grades/Feedback). Поза простим усередненням III застосовує складні методи агрегації. Байєсівські моделі оцінюють істинні бали з урахуванням упередженості та надійності рецензентів (Li, 2023). Підходи теорії відповіді на елемент (IRT) моделюють як якість роботи, так і характеристики рецензентів одночасно, створюючи точніші підсумкові оцінки.

Зона 4. Аналіз корпусу зворотного зв'язку (Analyzing Student Feedback). Методи NLP (Natural Language Processing – обробка природної мови) ідентифікують патерни в наборах зворотного зв'язку, виявляючи поширені помилкові уявлення, часто згадувані сильні/слабкі сторони та тенденції настроїв (Wei et al., 2020; Xiong et al., 2012). Цей метааналіз надає цінні інсайти для викладачів та студентів.

Зона 5. Підтримка викладача (Facilitating Instructor Oversight). Панелі, керовані III, виділяють аномалії – крайні розбіжності оцінок, потенційну змову, студентів, які відстають – що дає змогу цільового втручання (Topping et al., 2025). Автоматизоване виявлення плагиату та перевірки якості зменшують рутинне навантаження модераторів.

Зона 6. Інтегровані системи взаємооцінювання (Peer Assessment Systems). Комплексні платформи на кшталт RiPPLЕ інтегрують створення, оцінювання та персоналізовані практичні активності, з III, що координує всі виміри для оптимізації навчальних результатів (Topping et al., 2025).

Останні емпіричні дослідження демонструють педагогічну цінність вдосконалення III за кількома вимірами:

1. Покращення якості зворотного зв'язку. Експериментальне дослідження К. Guo (Guo, 2024) з EvaluMate (онлайн-системою для підтримки взаємооцінювання письмових робіт студентів) показало, що студенти, яких підтримує III, надавали значно більш конкретний, конструктивний та орієнтований на ревізію зворотний зв'язок порівняно з традиційним взаємооцінюванням ($d=1.28$, $p<0.7$), підтримуючи

життєздатність взаємооцінювання для підсумкової оцінки у великих курсах.

2. Розвиток метакогніції. Лонгітудне дослідження використання платформи PeerScholar (Pare, 2020) задокументувало прогресивне покращення здатності студентів розрізняти якісну роботу від слабкої протягом кількох циклів рецензування. Цей метакогнітивний розвиток представляє навчальний результат, незалежний від оволодіння предметним змістом.

3. Масштабованість без компромісів якості. Впровадження платформи RiPPLЕ продемонструвало масштабованість взаємооцінювання, вдосконаленого III, при збереженні стандартів якості (Topping et al., 2025). Узгодженість між рецензентами покращилася від $r=0.51$ до $r=0.71$ протягом циклів оцінювання, вказуючи на систематичний розвиток компетентності в оцінюванні.

Однак дослідження також ідентифікують обмеження та проблеми. Зокрема, надмірна залежність від пропозицій III. К. Guo (Guo, 2024) застерігав, що надмірна залежність від підказок III може зменшити розвиток автономного критичного мислення. Студенти мусять навчатися оцінювати та адаптувати рекомендації III, а не приймати їх некритично. Варіативна ефективність у різних контекстах. Ефективність варіюється залежно від дисципліни, типу завдання та контингенту студентів (Babik et al., 2024). Добре визначені завдання з чіткими критеріями якості показують кращі результати, ніж відкриті творчі завдання. Вимоги початкових інвестицій. Успішне впровадження вимагає значних попередніх інвестицій: розробки якісних рубрик, навчання студентів конструктивному зворотному зв'язку, калібрування алгоритмів III та постійного моніторингу якості (Topping et al., 2025).

Взаємооцінювання у підготовці вчителів. Підготовка вчителів представляє можливість для впровадження взаємооцінювання студентів. Поза навчанням предметного змісту, майбутні вчителі мають розвивати компетентність в оцінюванні – розуміння цілей оцінювання, методів, критеріїв якості та надання зворотного зв'язку (Недялкова, 2025). Взаємооцінювання надає автентичну практику оцінювання студентських робіт, формулювання конструктивного зворотного зв'язку та розпізнавання поширених помилок – педагогічних компетентностей. Дослідження К. Недялкової (Недялкова, 2025), проведене з українськими майбутніми вчителями математики, продемонструвало, що структуроване оцінювання за участю одногрупників підвищувало залученість, відповідальність та розвиток методичної компетентності. Студенти виконували

оцінювання за участю однокласників у трьох формах: формативній (вербальна оцінка), кількісній (балова оцінка) та конструктивній (формат порад), що давало змогу розвивати професійні навички оцінювання. Підтримка платформи для завантаження фотографій рукописних робіт давала змогу автентичного оцінювання математичних задач. Однак дослідження взаємооцінювання у підготовці вчителів переважно походять із західних контекстів. Авторка зазначає, що культурні фактори – ставлення до критики, колективізм проти індивідуалізму, дистанція влади – можуть впливати на динаміку оцінювання за участю однокласників у пострадянських освітніх системах (Недялкова, 2025). Це дослідження сприяє розумінню цих контекстуальних варіацій.

Порівняльний аналіз платформ для організації взаємооцінювання студентів. Систематичний огляд групою дослідників (Babik et al., 2024) 57 онлайн-систем рецензування та оцінювання (1995–2015) пропонує комплексну дослідницьку рамку з чотирма ієрархічними рівнями: варіанти дизайну, функціональні можливості, сценарії використання та потреби користувачів. Аналіз виявив ключову напруженість між орієнтацією на структуровані завдання (з чіткими рубриками та навичками) та дослідницьке навчання (відкриті проекти без «правильних» рішень). Зрілі платформи – Peerceptiv, SWoRD, Expertiza, PeerScholar, SPARKPlus – еволюціонували роками, пропонуючи розширений, емпірично перевірений функціонал. Більш ранні системи обмежуються вузькоспеціалізованими підходами з меншою гнучкістю. Вищезгаданий огляд онлайн-систем рецензування та оцінювання (Babik et al., 2024) передував «буму» III-інтеграції, особливо генеративного III, тому сучасні аналізи потребують врахування складності алгоритмів, приватності даних, прозорості та етичних аспектів, які залишилися поза його полем зору.

Моделі монетизації платформ суттєво впливають на доступність для закладів з обмеженими ресурсами. Ідентифіковані підходи включають:

1. Моделі «студент платить». Наприклад, платформа Kritik стягує зі студентів \$29–49 за курс, залишаючись безкоштовним для викладачів (документація платформи). Це переносить витрати на студентів, потенційно викликаючи питання справедливості.

2. Інституційне ліцензування. Наприклад, платформа Peerceptiv зазвичай веде переговори про ліцензії на рівні закладу з ціноутворенням на основі кількості студентів. Доступ для окремих викладачів може бути обмеженим або забороненим.

3. Freemium-моделі. Платформи Peergrade, Moodle Workshop пропонують безкоштовні базові версії з преміум-функціями, що вимагають оплати. Безкоштовні версії часто накладають обмеження (кількість студентів, функції, підтримка).

4. Відкритий код. Платформи Moodle Workshop є повністю відкритим кодом, хоча вдосконалення III можуть вимагати платних плагінів. Це пропонує максимальну гнучкість, але вимагає технічної експертизи для кастомізації.

5. Дослідницькі/академічні проекти EvaluMate, EduPCR часто доступні безкоштовно для академічного використання як дослідницькі проекти, хоча довгострокова стійкість може бути невизначеною.

Критерії та показники порівняння платформ. Для здійснення систематичного аналізу було виокремлено п'ять груп критеріїв: доступність (наявність безкоштовного плану, підтримка мов), типи оцінювання (peer – оцінювання діяльності одного студента іншими учасниками групи, self – оцінювання студентом власної діяльності, 360 – комплексне оцінювання з кількох джерел одночасно), роль III (калібрування, аналіз оберненого зв'язку), підтримка рефлексії та інтеграція з LMS (Learning Management System – системою управління навчанням). Особлива увага приділялася здатності платформ підтримувати методичну специфіку STEM-дисциплін, зокрема математики (табл. 1).

Для українських ЗВО критичним фактором є вартість впровадження. Платформи з моделлю «студент платить (student-pay)», – наприклад, платформа Kritik, – можуть створювати бар'єри для студентів з низьким рівнем доходу, що ставить під сумнів етичність їх масового впровадження в наших умовах. Більш прийнятними є системи з безкоштовними індивідуальними планами (наприклад, платформи PeerScholar, Gradescope) або повністю відкриті для академічного використання проекти (RiPPL, EvaluMate).

Також слід враховувати «приховані витрати», пов'язані з необхідністю навчання персоналу та технічною інтеграцією. Наприклад, Moodle Workshop є безкоштовним, але вимагає значних зусиль від IT-відділу для налаштування та підтримки III-плагінів. PeerScholar, у свою чергу, пропонує хмарне рішення, яке легко інтегрується з Canvas або Moodle через протокол LTI (Learning Tools Interoperability – інтероперабельність навчальних інструментів, дозволяє інтегрувати зовнішні освітні інструменти та сервіси з LMS (Learning Management System) без потреби дублювати логіни та налаштування), що мінімізує технічний поріг входу для викладача.

Таблиця 1.

**Порівняльна характеристика
провідних ІІІ-платформ для взаємооцінювання**

Платформа	Роль ІІІ та ключові алгоритми	Економічна модель (Accessibility)	Переваги для STEM/Математики
PeerScholar	Автоматизація робочого циклу, NLP-аналіз активності, підтримка ітерацій.	SaaS, обмежений індивідуальний безкоштовний план.	Робота з фотографіями рукописних робіт, фаза рефлексії.
https://www.peerscholar.com/?utm_source=chatgpt.com			
RiPPLE	Графова модель поширення довіри, рекомендаційні ІІІ-моделі.	Безкоштовно для академічних досліджень.	Створення та оцінювання банків тестових завдань.
https://itali.uq.edu.au/digital-learning/learning-analytics/ripple?utm_source=chatgpt.com			
Peerceptiv	ML-моделі надійності рецензентів, калібрування балів за IRT.	Переважно інституційні ліцензії, обмежені пілоти.	Висока надійність оцінок на великих потоках.
https://peerceptiv.com/?utm_source=chatgpt.com			
Kritik	ІІІ-калібрування «grading power», генератор рубрик.	Модель «student-pay» (\$29-49 за курс).	Динамічне зростання точності оцінювання студентів.
https://kritik.io			
EvaluMate	Генеративний ІІІ (LLM) для скафолдингу оберненого зв'язку в реальному часі.	Дослідницький проект, безкоштовно для пілотів.	Глибокий аналіз якості аргументації у відгуках.
https://www.evalumate.org/about			
Gradescope	AI-кластеризація відповідей, OCR для рукописних тестів.	Безкоштовні акаунти для окремих викладачів.	Найшвидша перевірка великих обсягів обчислювальних задач.
https://www.gradescope.eu/?utm_source=chatgpt.com			
ALEKS	Адаптивне навчання, діагностика на основі просторів знань.	Підписка, пробні версії.	Побудова індивідуальних траєкторій у математиці.
https://www.aleks.com/?_s=9969007131129218			

Функціональна специфікація платформи PeerScholar. PeerScholar – це онлайн-інструмент, спеціально розроблений для підтримки складного ітераційного процесу взаємооцінювання. Платформа пропонує три основні режими (Phase Settings):

1. Classic Individual (традиційний індивідуальний режим оцінювання): стандартний режим, де кожен студент працює самостійно.
2. Case Study: унікальний режим, що дозволяє викладачеві призначати різні кейси різним студентам. Студенти можуть оцінювати колег, що працювали над тими ж або іншими кейсами, що стимулює порівняльний аналіз.

3. Group Work (груповою робота): підтримка командних завдань, де групи подають спільну роботу та оцінюють інші групи (Toronto University, 2026).

PeerScholar було обрано для нашого емпіричного дослідження завдяки його унікальній трифазній архітектурі (табл. 2): Create (Створення) – Assess (Оцінювання) – Reflect (Рефлексія). На відміну від систем, що фокусуються лише на фінальному балі, PeerScholar робить акцент на процесі навчання через ревізію (British Columbia University, 2026).

Таблиця 2.

**Ключові функціональні зони PeerScholar
у навчальному процесі**

Зона функціоналу	Опис можливостей	Значення для вчителя математики
Фаза Create	Завантаження робіт у різних форматах (текст, фото, відео).	Перевірка методичних планів та рукописних розв'язань.
Фаза Assess	Анонімне рецензування за налаштовуваними рубриками (Matrix, Likert, Comment).	Розвиток навичок ідентифікації типових помилок.
Фаза Reflect	Оцінка корисності отриманого оберненого зв'язку, допрацювання роботи.	Глибока методична рефлексія та управління власних прогалин.
ІІІ-підтримка	Аналіз активності, допомога у формулюванні оберненого зв'язку (через інтерпретацію).	Об'єктивність та зменшення рутинної перевірки.

Алгоритм роботи студента:

Фаза створення (Create): 1) завантаження роботи у різних форматах: текст, зображення, відео та посилання на зовнішні ресурси; 2) детекція плагиату (за потреби).

Фаза оцінювання (Assess): 3) отримання кількох анонімних робіт одногрупників; 4) оцінювання роботи та написання розгорнутого оберненого зв'язку (фідбеку) за допомогою інтерактивних рубрик (Matrix, Point, Likert, Comment); активація функції самооцінювання (за потреби).

Фаза рефлексії (Reflect): 5) перегляд відгуків на свою роботу та оцінювання їх користі («feedback on feedback»); 6) перегляд оцінки інших рецензентів тієї ж роботи, яку він сам оцінював, що створює ефект навчання через порівняння

Алгоритм роботи викладача:

Налаштування: 1) створення рубрик, визначення кількості рецензентів (зазвичай 3–6); 2) розробка графіку автоматичного переходу між фазами.

Модерація: 3) перегляд «дашборду» «Student Progress & Grading», з метою відслідковування прогресу кожного студента, вивчення коментарів та, за необхідності, видалення неадекватних відгуків або зміна балів.

Інтегровані системи: 4) синхронізація списків студентів та груп безпосередньо з LMS (Learn, Canvas, Moodle), а також автоматичне експортування підсумкових оцінок в Gradebook.

Теоретична рамка: шість вимірів досвіду взаємооцінювання. На основі синтезу літератури концептуалізується досвід взаємооцінювання студентів за шістьма взаємопов'язаними вимірами:

1. **Методологічний вимір.** Вплив на розвиток спеціальної компетентності – розуміння критеріїв якості, розпізнавання ефективних/неефективних підходів, глибше концептуальне розуміння через порівняльний аналіз.

2. **Організаційний вимір.** Чіткість структури, вимоги до часу, ефективність робочого процесу, логістичне навантаження проти переваг.

3. **Мотиваційний вимір.** Впливи на мотивацію навчання, відповідальність за якісну роботу, залученість, впевненість в оцінюваному судженні.

4. **Технологічний вимір.** Зручність інтерфейсу, технічна доступність, надійність системи, адекватність функцій.

5. **Професійний-рефлексивний вимір.** Розвиток компетентності в оцінюванні, навичок формулювання зворотного зв'язку, метакогнітивної свідомості, здатності до саморегуляції.

6. **Вимір загального задоволення.** Оцінка досвіду – інтегральне (цілісне) оцінювання суб'єктивного досвіду учасника, яке відображає загальне сприйняття, враження та задоволеність діяльністю взаємооцінювання, готовність рекомендувати, сприймана ефективність відносно традиційних методів.

Ця рамка керувала дизайном та аналізом нашого емпіричного дослідження, даючи змогу комплексної оцінки впровадження PeerScholar за кількома теоретично релевантними вимірами.

Методологія емпіричного дослідження. Дослідження базується на апробації платформи PeerScholar в освітньому процесі педагогічного університету з метою оцінки її впливу на методичну підготовку студентів. Апробація проводилася протягом осіннього семестру 2024/2025 навчального року у межах навчальної дисципліни «Методика навчання математики». Учасниками стали 56 студентів третього і четвертого курсів спеціальності 014.04 Середня освіта (Математика) Державного закладу «Південноукраїнський національний педагогічний університет імені К.Д. Ушинського». Вибірка характеризувалася наступними показниками: 1) середній бал (GPA) з математичних дисциплін: <70 (9%), 70–79 (28%), 80–89 (41%), 90–100 (22%); 2) попередній досвід взаємооцінювання: 38% мали обмежений досвід, 62% не мали досвіду взагалі; 3) частота використання цифрових технологій: постійно (28%), часто (47%), іноді (19%), рідко (6%).

Протокол впровадження та типи завдань. Студенти виконали п'ять комплексних методичних завдань з курсу планіметрії: 1) розробка фрагменту уроку для теореми Піфагора. 2) аналіз поширених студентських помилок у геометричних доведеннях. 3) методичне обґрунтування викладання подібності трикутників. 4) порівняльний аналіз різних підходів до введення координатної геометрії. 5) розробка стратегій диференційованого навчання для просторової геометрії. Кожен цикл взаємооцінювання складався з трьох етапів.

1. **Етап «Створення»:** студенти розробляли фрагмент уроку з розв'язанням задачі підвищеної складності, де потрібно було представити не лише розв'язок, а й методичне обґрунтування вибору методів навчання. Роботи виконувалися рукописно та завантажувалися у вигляді фото.

2. **Етап «Оцінювання»:** кожен студент отримував три анонімні роботи колег. Оцінювання відбувалося за рубрикою, що включала бальну оцінку (правильність розв'язку), формувальний коментар (аналіз

логіки) та конструктивні поради (методичне вдосконалення).

3. Етап «Рефлексія»: студенти аналізували три отримані відгуки, оцінювали їх за шкалою корисності та мали можливість подати фінальну версію своєї роботи з урахуванням зауважень.

Роль викладача полягала у розробці рубрик оцінювання (спочатку з пропозиціями ШІ, потім кастомізованих). А також передбачала: 1) моніторинг прогресу студентів через панель аналітики; 2) виборчу перевірку якості зворотного студентів (приблизно 20% вибірки); 3) надання метазворотного зв'язку щодо особливо сильних або проблемних рецензій; 4) втручання в випадках крайніх розбіжностей оцінок або невідповідних коментарів.

PeerScholar автоматично збирав: 1. Час, витрачений на кожну фазу на студента. 2. Кількість та характер ревізій після зворотного зв'язку. 3. Коефіцієнти узгодженості між рецензентами. 4. Розподіл присвоєних балів. 5. Довжину зворотного зв'язку та лінгвістичні характеристики. 6. Рівні завершення кожної фази. Це типові характеристики для дашбордів сучасних систем, що дозволяють викладачам моніторити прогрес та якість навчання, узгоджуючись з описаною моделлю «Create-Assess-Reflect». Ці метрики відображають ключові аспекти педагогічної ефективності: часові витрати вказують на залученість, коефіцієнти узгодженості (inter-rater reliability) – на надійність, а лінгвістичний аналіз – на глибину оберненого зв'язку, що важливо для метакогніції (Pare, 2020).

Методи збору та аналізу даних. Для всебічної оцінки ефективності використання платформи PeerScholar застосовувався змішаний дизайн дослідження, що поєднував кількісні та якісні методи збору даних. Кількісний компонент включав структуроване анкетування студентів після завершення апробації за спеціально розробленим опитувальником, що містив 32 питання із шкалами Рейнсіса Лайкерта (1–5 балів) та 7 відкритих питань для детальних коментарів. Анкета охоплювала шість аспектів використання платформи: методичний (вплив на розвиток методичної компетентності), організаційний (зручність структури та часові витрати), мотиваційний (вплив на навчальну мотивацію та відповідальність), технологічний (зручність інтерфейсу та технічна доступність), професійно-рефлексивний (розвиток професійних навичок та рефлексивних умінь) та загальну оцінку досвіду.

Якісний компонент дослідження включав глибинні напівструктуровані інтерв'ю з шістьма студентами, відібраними за принципом максимальної варіативності (два

студенти з високим рівнем успішності, два з середнім та два з нижчим за середній), а також аналіз артефактів взаємооцінювання (зразків наданого студентами зворотного зв'язку та ревізій робіт після отримання коментарів від інших студентів). Додатково аналізувалися об'єктивні показники платформи: середній час виконання кожного етапу, кількість внесених ревізій після отримання оберненого зв'язку, ступінь узгодженості оцінок між різними рецензентами (міжрейтерська надійність).

Статистична обробка кількісних даних здійснювалася за допомогою пакету IBM SPSS Statistics 27 та включала дескриптивну статистику (середні значення, стандартні відхилення, частотні розподіли), кореляційний аналіз (коефіцієнт кореляції ρ Спірмена для виявлення зв'язку між різними аспектами сприйняття платформи) та оцінку внутрішньої узгодженості блоків анкети (коефіцієнт α Кронбаха). Якісні дані з відкритих питань та інтерв'ю оброблялися методом тематичного аналізу з виявленням повторюваних категорій, підрахунком частоти згадувань ключових тем та добром ілюстративних цитат.

Кількісні результати апробації. Дескриптивний аналіз результатів анкетування ($n=53$, рівень відгуку 94,6%) засвідчив загалом високе позитивне сприйняття студентами досвіду роботи з платформою PeerScholar. Загальний індекс задоволеності становив $M=4.28$ ($SD=0.62$), що відповідає оцінці «скоріше згоден» з тенденцією до «повністю згоден». Внутрішня узгодженість анкети виявилася високою ($\alpha=0.89$), що підтверджує надійність використаного інструментарію.

Аналіз за окремими аспектами виявив диференційоване сприйняття різних характеристик платформи. Найвище студенти оцінили методичний аспект використання PeerScholar ($M=4.36$, $SD=0.61$), зокрема:

– допомогла усвідомити власні методичні помилки через аналіз робіт одногрупників ($M=4.51$, $SD=0.63$), а також покращила здатність аналізувати методичну правильність розв'язань ($M=4.38$, $SD=0.69$). Такі дані вказують на позитивне сприйняття студентами методичних аспектів платформи: вони одностайно визнають, що аналіз робіт колег допомагає їм краще ідентифікувати власні помилки та критично оцінювати правильність розв'язання задач;

– найбільш значущим результатом є те, що студенти виявили вищу зацікавленість у конструктивному оберненому зв'язку ($M=4.81$, $SD=0.42$), аніж у отриманні балів ($M=4.42$, $SD=0.52$). Отже, отримані дані підтверджують одностайність студентів у

запиті на змістовну професійну комунікацію, що є критично важливим для їх професійного розвитку;

– показник у питанні «глибшого розуміння методики розв'язування математичних задач» ($M=4.21$, $SD=0.92$) вказує на певну диференціацію: для частини студентів автоматизована взаємоперевірка стала інноваційною, тоді як інші адаптувалися до неї складніше. Скоріше це може свідчити про потребу в додатковому інструктажі перед початком роботи з PeerScholar.

Обчислення коефіцієнту кореляції ρ Спірмена дозволило виявити взаємозв'язок між такими складниками методичного аспекту використання PeerScholar:

– між виявленням студентом альтернативних підходів до розв'язування задач, які раніше не розглядалися ним, та покращеною здатністю аналізувати методичну правильність розв'язань задач ($r_s=0.721$, $p<0.05$). Отже, це може свідчити про те, що ідентифікація студентами альтернативних способів розв'язання задач у межах взаємооцінювання є чинником підвищення ефективності їхньої здатності до комплексного аналізу методичної коректності таких рішень;

– між виявленням студентом альтернативних підходів до розв'язування задач, які раніше не розглядалися ним, та методичною цінністю формуальної (словесної) оцінки як форми оцінювання для подальшого професійного розвитку ($r_s=0.913$, $p<0.01$). Такий встановлений зв'язок пояснюється тим, що студентів з методичною новизною альтернативних підходів до розв'язування задач насамперед актуалізує їх потребу в інтерпретації та змістовному роз'ясненні запропонованих рішень (словесній формі);

– між виявленням студентом альтернативних підходів до розв'язування задач, які раніше не розглядалися ним, та методичною цінністю конструктивної (поради) як форми оцінювання для подальшого професійного розвитку ($r_s=0.697$, $p<0.05$). Отже, коли студент знайомиться з альтернативним підходом іншого студента до розв'язання задач, у нього виникає гостра потреба в якісному методичному коментарі (пораді), щоб зрозуміти цей новий досвід;

– між покращенням здатності студента аналізувати методичну правильність розв'язань задач та методичною цінністю формуальної (словесної) оцінки як форми оцінювання для подальшого професійного розвитку ($r_s=0.714$, $p<0.05$). Такі дані демонструють взаємозумовленість аналізу та формуальної (словесного) оцінювання: студенти, які набули практичного досвіду якісно аналізувати роботи своїх одногрупників,

починають вище цінувати саме словесну форму оберненого зв'язку як найбільш інформативну;

– між методичною цінністю формуальної (словесної) оцінки та методичною цінністю конструктивної (поради) як форми оцінювання для подальшого професійного розвитку ($r_s=0.764$, $p<0.05$), що свідчать про конвергенцію різних форм якісного оцінювання.

Наведені вище результати доводять, що використання PeerScholar стимулює перехід від «звичайного» розв'язування задач до проведення студентом – майбутнім учителем глибокої методичної рефлексії. Простіше це можна описати таким чином: бачення альтернативних підходів до розв'язування задач → проведення критичного аналізу → формулювання запиту на якісний обернений зв'язок.

Професійно-рефлексивний аспект також отримав високі оцінки ($M=4.35$, $SD=0.65$). Студенти зазначили розвиток методичних навичок ($M=4.74$, $SD=0.68$), комунікативних навичок формулювання зворотного зв'язку ($M=4.81$, $SD=0.63$) та рефлексивних умінь ($M=4.54$, $SD=0.71$). Важливо, що 87 % студентів погодилися або повністю погодилися з твердженням про доцільність застосування автоматизованої взаємоперевірки у майбутній шкільній практиці ($M=4.19$, $SD=0.63$).

Отримані значення ρ Спірмена (0,645–0,745 при $p<0,05$) свідчать про статистично значущі сильні та помірно сильні позитивні взаємозв'язки між ключовими складниками професійно-рефлексивного аспекту використання PeerScholar:

– між розвитком методичних навичок студента (аналіз розв'язань, виявлення помилок, оцінювання правильності) та розвитком його комунікативних навичок (формулювання зворотного зв'язку, конструктивна критика) ($r_s=0.745$, $p<0.05$). Можна припустити, що оцінювання робіт одногрупників вимагає не лише правильного аналізу, а й уміння пояснити, аргументувати та коректно донести свою думку, що природно поєднує методичну й комунікативну підготовку;

– між розвитком методичних навичок студента (аналіз розв'язань, виявлення помилок, оцінювання правильності) та розвитком його рефлексивних навичок (аналіз власних помилок, критичне осмислення отриманого оберненого зв'язку) ($r_s=0.713$, $p<0.05$). Такі дані дозволяють зробити висновок: аналіз розв'язувань задач інших студентів сприяє перенесенню критеріїв оцінювання на власну діяльність, що активізує важливу професійну якість майбутнього вчителя – рефлексивне мислення;

– між розвитком методичних навичок студента (аналіз розв'язань, виявлення помилок, оцінювання правильності) та корисністю проведення рефлексії для навчання (аналіз отриманих відгуків і доопрацювання роботи) ($r_s=0.645$, $p<0.05$). Виявлений взаємозв'язок дозволяє стверджувати, що студенти з вищим рівнем розвитку методичних умінь частіше усвідомлюють навчальну цінність рефлексії як інструменту вдосконалення власної навчальної та професійної діяльності. Отже, студенти, які мають вищий рівень методичної підготовки, схильні сприймати рефлексію не як формальну вимогу, а як засіб їхнього професійного зростання, що є принципово важливим для майбутньої педагогічної діяльності;

– між доцільністю застосування автоматизованої взаємоперевірки (на зразок PeerScholar) у шкільній практиці з математики та розвитком рефлексивних навичок (аналіз власних помилок, критичне осмислення отриманого оберненого зв'язку) ($r_s=0.679$, $p<0.05$). Можна припустити: студенти, у яких розвинуті рефлексивні навички краще усвідомлюють педагогічну цінність використання у шкільній практиці такої платформи автоматизованої взаємоперевірки як PeerScholar.

Отримані результати підтверджують, що використання автоматизованої взаємоперевірки PeerScholar сприяє комплексному розвитку професійно-рефлексивних умінь майбутніх учителів, поєднуючи їх методичну, комунікативну та рефлексивну складові. Така взаємопов'язаність умінь є важливою передумовою формування готовності майбутніх учителів до впровадження інноваційних освітніх технологій у шкільній практиці. Нижче побудована кореляційна плеяда (рис. 1), яка демонструє, що центральним компонентом професійно-рефлексивного аспекту використання PeerScholar є розвиток методичних навичок (5.1), що має сильні позитивні зв'язки з комунікативними (5.2) та рефлексивними навичками (5.3), а також з усвідомленням корисності рефлексії для навчання (5.4), яка пов'язана з усвідомленням використання такої платформи у шкільній практиці (5.6).

Організаційний аспект характеризувався помірними оцінками ($M=3.78$, $SD=0.84$). Студенти визнали структуру роботи зрозумілою ($M=4.40$, $SD=0.699$), проте оцінки достатності часових рамок виявилися нижчими ($M=3.40$, $SD=0.516$), що вказує на необхідність коригування термінів на етапі планування. Цікаво, що студенти не погодилися з твердженням про меншу часозатратність PeerScholar порівняно з традиційним оцінюванням ($M=2.89$, $SD=0.632$), проте визнали, що ці витрати виправдані

навчальними результатами ($M=4.31$, $SD=0.675$), а алгоритм автоматичного розподілу робіт між студентами працював коректно ($M=4.61$, $SD=0.516$).

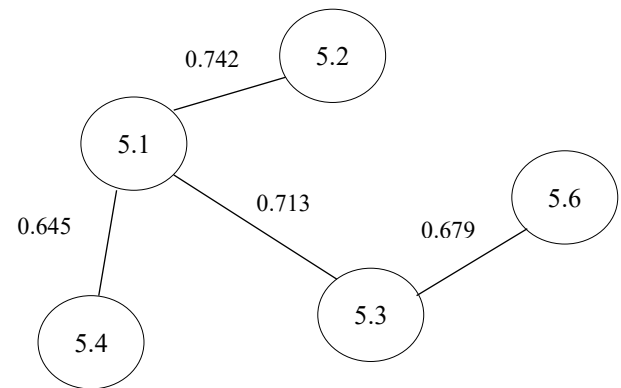


Рис. 1. Кореляційна плеяда складників професійно-рефлексивного аспекту використання PeerScholar

Мотиваційний аспект характеризувався помірно високими оцінками. Студенти підтвердили, що участь в автоматизованій взаємоперевірці підвищила їхню відповідальність за якість власної роботи ($M=4.33$, $SD=0.69$) та навчальну мотивацію ($M=3.97$, $SD=0.84$). Отримання зворотного зв'язку від кількох одногрупників було оцінено як цінніше порівняно з оберненим зв'язком викладача ($M=4.24$, $SD=0.75$). Можливість анонімного оцінювання позитивно вплинула на зниження тривожності студентів ($M=4.16$, $SD=0.88$), хоча рівень упевненості під час оцінювання робіт виявився помірним ($M=3.32$, $SD=0.94$), що вказує на необхідність додаткової підготовки студентів до ролі рецензентів.

Обчислення коефіцієнту кореляції ρ Спірмена дозволило виявити взаємозв'язок між такими складниками мотиваційного аспекту використання PeerScholar: між підвищенням навчальної мотивації студента завдяки участі в автоматизованій взаємоперевірці та цінністю отримання оберненого зв'язку від кількох одногрупників, а не лише від викладача ($r_s=0.707$, $p<0.05$). Отримані результати дозволяють припустити, що орієнтація майбутніх учителів на отримання багатоканального оберненого зв'язку (від одногрупників і викладача) пов'язана з підвищенням їхньої пізнавальної мотивації.

Технологічний аспект отримав задовільні оцінки ($M=3.91$, $SD=0.79$), причому виявлено значну варіативність у сприйнятті зручності інтерфейсу ($M=3.85$, $SD=1.02$). Кореляційний аналіз показав сильний зв'язок між оцінкою технічної зручності платформи та загальною задоволеністю досвідом ($r_s=0.68$, $p<0.001$), що підкреслює критичну важливість технічного аспекту для загального сприйняття. Більшість студентів (79 %)

мали технічно доступні пристрої для роботи, проте 21 % вказали певні труднощі з доступом, особливо під час дистанційного навчання.

Порівняльний аналіз за критерієм успішності виявив, що студенти з вищим середнім балом (≥ 85 , $n=22$) демонстрували статистично значущо вищу оцінку методичної цінності платформи ($M=4.61$ проти $M=4.18$) та професійного розвитку ($M=4.52$ проти $M=4.12$) порівняно зі студентами з нижчим середнім балом (< 85 , $n=31$). Проте різниця в загальній задоволеності не досягла статистичної значущості ($M=4.38$ проти $M=4.19$), що свідчить про універсальну придатність платформи для студентів з різним рівнем підготовки.

Якісні результати апробації. Тематичний аналіз відкритих відповідей та інтерв'ю виявив п'ять основних категорій переваг використання PeerScholar у професійній підготовці майбутніх учителів математики.

Розвиток критичного та методичного мислення (згадано 83 % респондентів) виявився найчастіше артикульованою перевагою. Студенти наголошували, що оцінювання робіт одногрупників змусило їх глибше аналізувати методичні аспекти розв'язань та виявляти власні прогалини у розумінні. Один студент зазначив: «Оцінюючи роботи одногрупників, я вперше помітив типові методичні помилки, які сам часто робив, але не усвідомлював. Це змусило мене переглянути свій підхід до планування уроків геометрії». Інша студентка додала: «Я виявила принаймні три різні підходи до пояснення теореми Піфагора, які раніше не розглядала. Це розширило мій методичний інструментарій».

Об'єктивність та справедливість оцінювання (65 % респондентів) визнавалася важливою перевагою автоматизованої системи. Студенти підкреслювали, що отримання оцінок від трьох незалежних рецензентів знижує суб'єктивність порівняно з оцінюванням одним викладачем. «Анонімність дозволила мені бути чеснішою у коментарях, не боячись образити одногрупників. Я могла вказати на реальні недоліки, що неможливо при відкритому взаємооцінюванні», – відзначила одна з учасниць.

Розвиток комунікативних навичок формування конструктивного оберненого зв'язку (58 % респондентів) виділявся як професійно важлива компетенція. Студенти усвідомлювали, що навчилися формулювати критику у конструктивній формі, що є критично важливим для майбутньої роботи з учнями. «Спочатку мені було важко писати коментарі – я не знала, як вказати на помилки, не образивши автора. Але з

кожним завданням це ставало легше. Тепер я розумію, як надавати зворотний зв'язок учням так, щоб вони сприймали його позитивно», – поділилася своїм досвідом студентка.

Зручність дистанційної роботи та гнучкість (52 % респондентів) виявилася актуальною особливо в контексті змішаного навчання. Студенти цінували можливість виконувати завдання та здійснювати взаємооцінювання у зручний час, не будучи прив'язаними до аудиторних занять.

Етап рефлексії як найцінніший для навчання (47 % респондентів) виокремлювався як унікальна характеристика платформи, що відрізняє її від традиційного оцінювання. «Аналіз отриманих коментарів та можливість доопрацювати роботу після отримання оберненого зв'язку – це те, чого найбільше бракує у звичайних контрольних. Я реально вчилася на своїх помилках, а не просто отримувала оцінку», – зазначила студентка з високим рівнем успішності.

Водночас аналіз виявив і низку труднощів та обмежень. Часові витрати (згадано 71 % респондентів) визнавалися основним недоліком, особливо на етапі освоєння платформи. Технічні проблеми (38 % респондентів) включали складнощі з завантаженням фотографій рукописних робіт, нестабільність інтернет-з'єднання та незрозумілість деяких елементів інтерфейсу англійською мовою. Невпевненість у правильності власного оцінювання (34 % респондентів) відображала потребу в додатковому навчанні принципам взаємооцінювання перед початком роботи. Поверхневість деяких коментарів від одногрупників (29 % респондентів) вказувала на необхідність контролю якості зворотного зв'язку з боку викладача.

Аналіз об'єктивних метрик платформи. Додатковий аналіз даних, автоматично збірних платформою PeerScholar, надав об'єктивну інформацію про стійкі, повторювані способи поведінки студентів (поведінкові патерни). Середній час виконання етапу взаємооцінювання становив 47 хвилин на одну роботу ($SD=18$ хв.), що перевищувало початково закладені викладачем 30 хвилин. Це підтверджує студентські скарги на часові витрати та вказує на необхідність коригування термінів. Водночас 68 % студентів ($n=38$) внесли ревізії до своїх робіт після етапу рефлексії, причому середня кількість змін становила 3.4 виправлення на роботу, що свідчить про активне використання отриманого оберненого зв'язку (фідбеку) для вдосконалення розв'язань.

Міжрейтерська надійність, виміряна як кореляція між оцінками трьох незалежних

рецензентів для однієї роботи, становила $r_s=0.62$ ($p<0.001$), що вказує на помірну узгодженість оцінювання. Це значення нижче за ideale, проте є прийнятним для освітнього контексту, враховуючи початковий рівень підготовки студентів до ролі рецензентів. Аналіз динаміки показав зростання міжрейтерської надійності від першого завдання ($r_s=0.51$) до п'ятого ($r_s=0.71$), що свідчить про поступовий розвиток навичок оцінювання протягом семестру.

Висновки і перспективи подальших досліджень. Результати нашого дослідження підтверджують концепцію «ІІІ як підсилювача», а не заміни викладача. Хоча ІІІ ефективно автоматизує рутинні процеси (розподіл робіт, калібрування), саме «людський» компонент – якісний обернений зв'язок від колеги та наступна рефлексія – забезпечує глибоке навчання. Наші дані про зростання міжрейтерської надійності до $r_s=0.71$ узгоджуються з висновками інших дослідників (Li, 2023; Torring et al., 2025), які вказують на можливість досягнення експертної точності peer-оцінок за умови використання інтелектуальних систем калібрування.

Проте слід враховувати ризики «алгоритмічної упередженості» та прозорості. Студенти в інтерв'ю вказали на невпевненість у правильності власного оцінювання ($M=3,32$, $SD=0,94$), що корелює з дослідженнями K. Guo (Guo, 2024) про можливе зниження автономності студентів при надмірному використанні ІІІ-підказок. Відтак, роль викладача зміщується від перевірки робіт до проектування складних рубрик та модерації тих 10-15% випадків, де система фіксує аномалії в оцінках.

Виявлений нами сильний зв'язок між сприйняттям методичної новизни та цінністю словесного оберненого зв'язку ($r_s=0,913$, $p<0,01$) підкреслює соціально-конструктивістську природу навчання в PeerScholar. Згідно з теорією зони найближчого розвитку (ZPD), оберненого зв'язку від одногрупника, який знаходиться на схожому когнітивному рівні, може бути більш зрозумілим і прийнятним для студента, ніж експертний коментар викладача. Анонімність, яку забезпечує платформа, критично важлива для зниження тривожності ($M=4,16$, $SD=0,88$), проте вона не повинна перетворюватися на безвідповідальність, що запобігається через механізм оцінки якості самого оберненого зв'язку («обернений зв'язок на обернений зв'язок»).

Для математичних дисциплін перевагою PeerScholar виявилася підтримка рукописних робіт. Як зазначає К. Недялкова (Недялкова, 2025), у математиці неможливо повністю перейти на цифровий ввід без втрати

когнітивної глибини розв'язання задачі. Успіх нашої апробації зумовлений поєднанням традиційного письма з цифровою платформою модерації. Досвід використання платформ Gradescope та ALEKS вказує на те, що майбутнє за гібридними системами, де ІІІ розпізнає рукописний текст, кластеризує помилки, а студенти обговорюють методичні нюанси у формувальному діалозі.

Основним обмеженням дослідження є відсутність контрольної групи, що не дозволяє однозначно стверджувати, що прогрес у надійності оцінок зумовлений саме платформою, а не загальним навчанням на курсі. Крім того, англomовний інтерфейс платформи PeerScholar залишається бар'єром для частини студентів, що знижує швидкість їхньої роботи. Подальші дослідження мають зосередитися на: 1) лонгitudному вивченні впливу досвіду PeerScholar на реальну педагогічну діяльність у школі випускників педагогічного університету. 2) використанні генеративного ІІІ для автоматичного створення персоналізованих рубрик під конкретні математичні задачі. 3) аналізі впливу різних моделей розподілу рецензентів (наприклад, парування «сильний-слабкий» на основі ІІІ-аналітики) на якість навчання.

Проведене дослідження дозволяє констатувати, що інтеграція інтелектуальних платформ взаємооцінювання є необхідною умовою модернізації вищої педагогічної освіти в умовах цифровізації. Системний порівняльний аналіз виявив, що ринок EdTech пропонує диференційовані рішення: від адаптивних систем індивідуального навчання (платформа ALEKS) до комплексних рефлексивних екосистем (PeerScholar), вибір яких має базуватися на специфіці дисципліни та бюджетних можливостях ЗВО. Взаємооцінювання з підтримкою ІІІ не просто автоматизує перевірку – воно створює нове якісне середовище для професійного діалогу, де помилка стає не приводом для покарання, а ресурсом для глибинного розуміння предмета та вдосконалення педагогічної майстерності.

Апробація платформи PeerScholar у викладацькій діяльності підтвердила її педагогічну й технологічну доцільність у системі підготовки майбутніх учителів математики. Інтеграція цифрових технологій, елементів штучного інтелекту та методично продуманого дизайну взаємооцінювання створює умови для формування сучасної методичної компетентності, орієнтованої на рефлексію, об'єктивність та професійну відповідальність. Результати дослідження узгоджуються із висновками К. Недялкова (Недялкова, 2025) про ефективність структурованого багаторівневого процесу взаємооцінювання та підтверджують спостереження

K. Guo (Guo, 2024) щодо важливості етапу рефлексії для розвитку метакогнітивних навичок студентів.

На основі результатів апробації сформульовано низку рекомендацій для ефективного впровадження PeerScholar у освітній процес. По-перше, доцільно проводити вступне навчальне заняття з принципів конструктивного зворотного зв'язку та роботи з платформою перед початком взаємооцінювання, що дозволить знизити початкову невпевненість студентів та підвищити якість оберненого зв'язку. По-друге, необхідно закладати достатні часові рамки для кожного етапу (мінімум 60 хвилин на оцінювання однієї роботи), враховуючи реальні часові витрати студентів. По-третє, рекомендується поступове збільшення складності завдань протягом семестру для формування навичок оцінювання від простіших до більш складних методичних ситуацій. По-четверте, викладачу варто здійснювати вибірковий контроль якості наданого студентами оберненого зв'язку, особливо на початкових етапах, надаючи метакоментарі щодо якості рецензій.

Емпірична апробація PeerScholar у підготовці вчителів математики довела, що цифрова трансформація оцінювання дозволяє успішно вирішити проблему «дефіциту оберненого зв'язку». Статистично підтверджено, що студенти високо цінують методичну цінність взаємоперевірки, розглядаючи її як засіб виявлення власних когнітивних лакун через аналіз альтернативних підходів колег. Зростання міжрейтерської надійності до $r=0,71$ свідчить про реальне формування оцінювальної компетентності майбутніх педагогів протягом одного семестру.

На основі отриманих результатів ми пропонуємо такі практичні рекомендації:

1. Для викладачів: впроваджувати взаємооцінювання як необхідний елемент методичної підготовки, обов'язково включаючи етап рефлексії над отриманим зворотним зв'язком. Слід закладати достатні часові рамки (мінімум 45–60 хв. на оцінювання однієї роботи) та проводити попередній інструктаж щодо принципів конструктивної критики.

2. Для адміністрації ЗВО: надавати перевагу платформам, що підтримують LTI-інтеграцію та мають безкоштовні рівні для пілотних проектів. Пріоритетом має бути створення інституційної політики ШІ-грамотності, яка б регламентувала етичне використання алгоритмів в оцінюванні.

3. Для розробників EdTech: зосередитися на локалізації інтерфейсів та вдосконаленні OCR-модулів для розпізнавання математичних формул у рукописних роботах, що

значно знизить технічний поріг входу для природничих спеціальностей.

Перспективи подальших досліджень включають лонгitudне вивчення довгострокового впливу досвіду взаємооцінювання в PeerScholar на професійну діяльність майбутніх фахівців під час педагогічної практики та першого року роботи в школі, порівняльне дослідження ефективності PeerScholar та інших платформ взаємооцінювання в різних дисциплінарних контекстах, а також дослідження впливу різних моделей розподілу рецензентів (випадковий розподіл проти розподілу за рівнем успішності і проти розподілу за методичним підходом) на якість взаємооцінювання та навчальні результати студентів.

Список бібліографічних посилань

- Недялкова, К.В. (2025). Використання технологій штучного інтелекту для оцінювання та взаємооцінювання студентів. *Проблеми сучасних трансформацій. Серія: педагогіка та психологія*, (9). Doi: <https://doi.org/10.54929/2786-9199-2025-9-08-04>.
- Babik et al., 2024 – Babik, D., Gehringer, E., Kidd, J., Sunday, K., Tinapple, D., & Gilbert, S. (2024). A systematic review of educational online peer-review and assessment systems: Charting the landscape. *Educational Technology Research and Development*, 72(3): 1653–1689. Doi: <https://doi.org/10.1007/s11423-024-10349-x>.
- British Columbia University, 2026 – University of British Columbia. (n.d.). peerScholar instructor guide. Learning Technology Hub. URL: <https://lthub.ubc.ca/guides/peerscholar-instructor-guide/>.
- Guo, 2024 – Guo, K. (2024). EvalMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing*, 61: 100864. Doi: <https://doi.org/10.1016/j.asw.2024.100864>.
- Joordens et al., 2009 – Joordens, S., Pare, D.E., & Pruesse, K. (2009). PeerScholar: An Evidence-based online peer assessment tool supporting critical thinking and clear communication. In *Proceedings of the 2009 International Conference on e-Learning* (pp.236–240). Academic Conferences Limited.
- Li, 2023 – Li, A.W. (2023). Using Peerceptiv to support AI-based online writing assessment across the disciplines. *Assessing Writing*, 57: 100746. Doi: <https://doi.org/10.1016/j.asw.2023.100746>.
- Liu et al., 2024 – Liu, C., Cui, J., Shang, R., Jia, Q., Rashud, P., & Gehringer, E. (2024). Generative AI for peer assessment helpfulness evaluation. In B. Paaßen & C.D. Epp (eds.), *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 412–419). URL: <https://educationaldatamining.org/edm2024/proceedings/2024.EDM-short-papers.39/2024.EDM-short-papers.39.pdf>.
- Pare, 2020 – Pare, D.E. (2020). The peerScholar process: Using formative peer and self-assessment to enhance and quantify critical thinking skills, metacognitive awareness, and sense of community: Theses of Ph.D dissertation. University of Toronto. URL: <https://utoronto.scholaris.ca/server/api/core/bitstreams/1ae4ab91-bef6-40e2-8e6c-e68cfe4435af/content>.
- Pare, Joordens, 2008 – Pare, D.E., Joordens, S. (2008). Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning*, 24(6): 526–540. Doi: <https://doi.org/10.1111/j.1365-2729.2008.00290.x>.
- Topping et al., 2025 – 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(1): 3. Doi: <https://doi.org/10.1186/s41239-024-00501-1>.
- Toronto University, 2026 – Toronto Metropolitan University. (n.d.). peerScholar instructor guide. URL: <https://www.torontomu.ca/courses/toolbox/peerscholar/instructor-guide/>.
- Wei, Yu, Gao, 2022 – Wei, W., Yu, Y., & Gao, G. (2022). Investigating learners' changing expectations on learning experience in a MOOC of professional translation and interpreter training. *SAGE Open*, 12(4). Doi: <https://doi.org/10.1177/21582440221134577>.
- Xiong et al., 2012 – Xiong, W., Litman, D., Schunn, C.D. (2012). Natural language processing techniques for researching and improving peer feedback. *Journal of Writing Research*, 4(2): 155–176.
- References**
- Nedyalkova, K.V. (2025). Using artificial intelligence technologies for assessing and peer-assessing students. *Problems of modern transformations. Series: pedagogy and psychology*, (9). Doi: <https://doi.org/10.54929/2786-9199-2025-9-08-04> [in Ukr.].
- Babik et al., 2024 – Babik, D., Gehringer, E., Kidd, J., Sunday, K., Tinapple, D., & Gilbert, S. (2024). A systematic review of educational online peer-review and assessment systems: Charting the landscape. *Educational Technology Research and Development*, 72(3): 1653–1689. Doi: <https://doi.org/10.1007/s11423-024-10349-x>.
- British Columbia University, 2026 – University of British Columbia. (n.d.). peerScholar instructor guide. Learning Technology Hub. URL: <https://lthub.ubc.ca/guides/peerscholar-instructor-guide/>.
- Guo, 2024 – Guo, K. (2024). EvaluMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing*, 61: 100864. Doi: <https://doi.org/10.1016/j.asw.2024.100864>.
- Joordens et al., 2009 – Joordens, S., Pare, D.E., & Pruesse, K. (2009). PeerScholar: An Evidence-based online peer assessment tool supporting critical thinking and clear communication. In *Proceedings of the 2009 International Conference on e-Learning* (pp.236–240). Academic Conferences Limited.
- Li, 2023 – Li, A.W. (2023). Using Peerceptiv to support AI-based online writing assessment across the disciplines. *Assessing Writing*, 57: 100746. Doi: <https://doi.org/10.1016/j.asw.2023.100746>.
- Liu et al., 2024 – Liu, C., Cui, J., Shang, R., Jia, Q., Rashud, P., & Gehringer, E. (2024). Generative AI for peer assessment helpfulness evaluation. In B. Paaßen & C.D. Epp (eds.), *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 412–419). URL: <https://educationaldatamining.org/edm2024/proceedings/2024.EDM-short-papers.39/2024.EDM-short-papers.39.pdf>.
- Pare, 2020 – Pare, D.E. (2020). The peerScholar process: Using formative peer and self-assessment to enhance and quantify critical thinking skills, metacognitive awareness, and sense of community: Theses of Ph.D dissertation. University of Toronto. URL: <https://utoronto.scholaris.ca/server/api/core/bitstreams/1ae4ab91-bef6-40e2-8e6c-e68cfe4435af/content>.
- Pare, Joordens, 2008 – Pare, D.E., Joordens, S. (2008). Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning*, 24(6): 526–540. Doi: <https://doi.org/10.1111/j.1365-2729.2008.00290.x>.
- Topping et al., 2025 – 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(1): 3. Doi: <https://doi.org/10.1186/s41239-024-00501-1>.
- Toronto University, 2026 – Toronto Metropolitan University. (n.d.). peerScholar instructor guide. URL: <https://www.torontomu.ca/courses/toolbox/peerscholar/instructor-guide/>.
- Wei, Yu, Gao, 2022 – Wei, W., Yu, Y., & Gao, G. (2022). Investigating learners' changing expectations on learning experience in a MOOC of professional translation and interpreter training. *SAGE Open*, 12(4). Doi: <https://doi.org/10.1177/21582440221134577>.
- Xiong et al., 2012 – Xiong, W., Litman, D., Schunn, C.D. (2012). Natural language processing techniques for researching and improving peer feedback. *Journal of Writing Research*, 4(2): 155–176.

SKVORTSOVA Svitlana

Doctor Sciences of Pedagogy, Professor, Professor at the Department of Mathematics and Methods of its Teaching, South Ukrainian National Pedagogical University named after K.D. Ushynsky

SYMONENKO Tetyana

Doctor Sciences of Pedagogy, Professor,
Professor at the Department of Teaching Methods, Stylistics and Culture of the Ukrainian Language,
Bohdan Khmelnytsky National University of Cherkasy;
Director of the Center for Ukrainian Researchers in Austria,
Affiliated Professor of the Department of Strategic Communications WU Universität

HNEZDILOVA Kira

Doctor Sciences of Pedagogy, Professor, Professor at the Department of Primary and Special Education,
Bohdan Khmelnytsky National University of Cherkasy

NIEDIALKOVA Kateryna

Ph.D in Pedagogy, Associate Professor at the Department of Mathematics and Methods of its Teaching,
South Ukrainian National Pedagogical University named after K.D. Ushynsky

AI-SUPPORTED PEER ASSESSMENT PLATFORMS IN THE METHODOLOGICAL TRAINING OF PRE-SERVICE MATHEMATICS TEACHERS: A COMPARATIVE ANALYSIS AND THE IMPLEMENTATION OF PEERSCHOLAR

Summary. Introduction. The modern paradigm of higher education is characterized by a transition from traditional hierarchical methods of control to participatory models, among which peer assessment occupies a key place. This mechanism involves the involvement of education seekers in assessing the learning outcomes of their colleagues, which contributes to the development of critical thinking and self-regulation skills. However, the wide implementation of peer assessment in the university environment faces a number of challenges: the subjectivity of the assigned scores, a significant cognitive load on the teacher, and the problem of the "depth deficit" of feedback. The integration of artificial intelligence technologies, in particular

natural language processing (NLP) methods, opens up new opportunities for addressing these problems.

Purpose. The purpose of the study is to conduct a systematic comparative analysis of AI-supported peer assessment platforms and assess the effectiveness of PeerScholar in developing assessment competencies among future mathematics teachers in Ukrainian higher education.

Methods. To comprehensively assess the effectiveness of the PeerScholar platform, a mixed-method study design was used, combining quantitative and qualitative data collection methods. The quantitative component included a structured survey of students after the completion of the pilot using a specially developed questionnaire containing

32 questions with Likert scales (1-5 points) and 7 open-ended questions for detailed comments. The qualitative component of the study included in-depth semi-structured interviews with six students selected according to the principle of maximum variability (two students with a high level of success, two with an average and two with a lower than average level), as well as an analysis of peer assessment artifacts (samples of feedback provided by students and revisions of work after receiving comments from other students). Additionally, objective indicators of the platform were analyzed: the average completion time of each stage, the number of revisions made after receiving feedback, the degree of consistency of assessments between different reviewers (inter-rater reliability).

Statistical processing of quantitative data was carried out using the IBM SPSS Statistics 27 package and included descriptive statistics (means, standard deviations, frequency distributions), correlation analysis (Spearman's ρ correlation coefficient to identify relationships between different aspects of platform perception) and assessment of internal consistency of questionnaire blocks (Cronbach's alpha coefficient). Qualitative data from open-ended questions and interviews were processed using thematic analysis to identify recurring categories, count the frequency of mentions of key themes, and select illustrative quotes.

Results. Comparative analysis of peer assessment platforms revealed differentiated EdTech solutions from adaptive systems (ALEKS) to comprehensive reflective ecosystems (PeerScholar, RiPPLE), with the critical role of affordability for Ukrainian HEIs. The PeerScholar pilot demonstrated statistically significant positive student perceptions of the methodological ($M=4.36$, $SD=0.61$) and professional-reflective ($M=4.35$, $SD=0.65$) aspects of the platform. Strong correlations were found between the identification of alternative approaches and the value of formative feedback ($r_s=0.913$, $p<0.01$), as well as between the development of methodological and reflective skills ($r_s=0.713$, $p<0.05$). Inter-rater reliability increased from $r=0.51$ to $r=0.71$ over the semester, confirming the gradual development of assessment competence. Students showed a higher interest in


constructive feedback ($M=4.81$) compared to receiving points ($M=4.42$), which indicates a request for professional communication.


Originality. Based on the results of the testing, a number of recommendations were formulated for the effective implementation of PeerScholar in the educational process. First, it is advisable to conduct an introductory training session on the principles of constructive feedback and working with the platform before the start of peer assessment, which will reduce the initial uncertainty of students and improve the quality of feedback. Second, it is necessary to set sufficient time frames for each stage (at least 60 minutes for assessing one work), taking into account the real time costs of students. Third, it is recommended to gradually increase the complexity of tasks throughout the semester to develop assessment skills from simpler to more complex methodological situations. Fourth, the teacher should conduct selective quality control of the feedback provided by students, especially in the initial stages, providing meta-comments on the quality of the reviews.

Conclusion. PeerScholar integration creates a quality environment for professional dialogue, where peer assessment stimulates the transition from routine problem solving to deep methodological reflection. The sequence is revealed: identification of alternative approaches → critical analysis → request for qualitative feedback. Time costs ($M=3.40$) require planning adjustments, and the technological barrier of the English-language interface reduces efficiency for some students. The study offers practical recommendations for choosing platforms for institutions with limited resources and implementing structured peer assessment in teacher training.

Keywords: student peer assessment; artificial intelligence in education; PeerScholar; assessment competence; mathematics teacher training; digital transformation of assessment; reflective learning; peer assessment; methodological competence.


Одержано редакцією 09.02.2026
Прийнято до публікації 23.02.2026


 <https://doi.org/10.31651/2524-2660-2026-1-26-39>

 <https://orcid.org/0000-0002-4979-7669>

БОДНАРУК Світлана

кандидатка фізико-математичних наук, доцентка кафедри алгебри та інформатики,
Чернівецький національний університет імені Юрія Федьковича,


 s.bodnaruk@chnu.edu.ua

 <https://orcid.org/0000-0001-7796-7760>

КОЛІСНИК Руслана


кандидатка фізико-математичних наук, доцентка, завідувачка кафедри алгебри та інформатики,
Чернівецький національний університет імені Юрія Федьковича,

 r.kolisnyk@chnu.edu.ua

 <https://orcid.org/0000-0002-5971-8557>

ШЕВЧУК Наталія

кандидатка фізико-математичних наук, асистентка кафедри алгебри та інформатики,
Чернівецький національний університет імені Юрія Федьковича,

 n.shevchuk@chnu.edu.ua

УДК 378.091.33-027.22:004.8(045)

ДІАЛОГОВА ВЗАЄМОДІЯ З ШІ ЯК ІНСТРУМЕНТ РОЗВИТКУ НАУКОВОЇ СУБ'ЄКТНОСТІ СТУДЕНТА

У статті досліджено трансформацію методичних підходів до організації науково-дослідної роботи студентів в умовах тотальної доступності генеративних мовних моделей (LLM).

Запропоновано та апробовано інноваційну модель «цифрової маєвтики» (Rodo, 2025), яка зміщує фокус викладацького контролю з фінального тексту на рефлексивний аудит процесу дослідження.

Обґрунтовано концепцію «Ризику Гамільтона» – право дослідника на висування сміливих гіпотез у зонах фактологічного вакууму ШІ («галюцинацій»), що за умови належної верифікації стають джерелом наукової новизни.

Наукова новизна підходу полягає в інтеграції інтерактивного методу «акваріуму» у структуру цифрової маєвтики, що перетворює процес критичної верифікації відповідей ШІ на відкриту колективну дискусію.