

MRHA як цифровий двійник учня: адаптивна модель оцінювання на основі когнітивної резильєнтності

Василь Залізко^a

^a Міжнародний інноваційний центр штучного інтелекту, Цюрих, Швейцарія

Abstract

У цьому дослідженні ми концептуалізуємо українську онлайн платформу MRHA як цифровий двійник учня та пропонуємо модель неперервного оцінювання когнітивної резильєнтності. Запропонована модель навчання більше не оцінює учнів за бінарним критерієм «правильно/неправильно», натомість аналізується, наскільки стабільно учень утримує коректне міркування в умовах misleadingness. Реалізовано адаптивний цикл уточнювальних запитань і shape-preserving spline-калібрування, що дозволяє надавати частковий бал за відновлення reasoning після помилки. В емпіричній демонстрації ($N = 280$) середні значення субіндексів резильєнтності до втручання становили: Visual = 0.6113, Textual = 0.5469, Mixed = 0.5230, що відповідало інтегральному показнику до втручання $H_{pre} = 0.5528$. Після адаптивної підтримки середній інтегральний індекс зріс до $H_{post} = 0.7732$, із середнім приростом $\Delta H = +0.2204$. Це покращення є статистично значущим (ANOVA $F(2, 277) = 97.66, p < 10^{-30}$), медіанний приріст становив приблизно 0.22. Розподіл ΔH ($Q_1 = 0.0640$, Med = 0.2225, $Q_3 = 0.3524$) свідчить, що більшість учнів отримали позитивний ефект від адаптивної взаємодії. Отримані результати підтверджують, що інтегральний resilience-індекс цифрового двійника, який агрегує візуальну, текстову та змішану компоненти, відображає стан учня значно повніше, ніж одиничний показник правильності. Також продемонстровано нелінійне spline-відображення індексу резильєнтності в інтервалі $[0, 1]$ на звичну 12-бальну шкалу, що дозволяє зберегти деталізацію внутрішнього learner state при зовнішній дискретній інтерпретації MRHA з ядром адаптивної системи оцінювання нового покоління. На відміну від існуючих методів оцінювання знань ми використали OCR аналіз рукописних текстів для інтелектуальної трансформації у динамічну модель когнітивного стану учня.

Keywords: MRHA, цифровий двійник учня, когнітивна резильєнтність, неперервне оцінювання, resilience to misleadingness, knowledge tracing, adaptive learning, spline-апроксимація, 12-бальна шкала, рукописний когнітивний слід, штучний інтелект в освіті

1. Вступ

У цій статті ми розглядаємо навчальну екосистему MRHA не як черговий генеративний освітній чат-інструмент, а як реальний прототип цифрового двійника учня. Така постановка питання є принципово важливою, оскільки вона зміщує фокус з окремих AI-функцій – генерації пояснень, рекомендації вправ, автоматичної перевірки чи відповіді на запитання – на побудову цілісної динамічної моделі учня, яка оновлюється в часі, симулює можливі сценарії навчання і підтримує адаптивне оцінювання.

Сучасні цифрові навчальні середовища містять дедалі більше мультимодальних матеріалів, у яких текстові, візуальні, формальні та інтерактивні сигнали взаємодіють між собою. Водночас оцінювання в багатьох освітніх системах усе ще спирається на дискретну схему: відповідь або правильна, або неправильна. Такий підхід виявляється недостатнім

у тих ситуаціях, коли учень помиляється не через повну відсутність знання, а через misleadingness: хибну візуальну інтерпретацію, надмірний текстовий шум, невдале формулювання задачі, когнітивну ілюзію, помилкову математичну евристику або логічну пастку у доведенні. Саме тому в цій роботі пропонується перейти від бінарної моделі “знає / не знає” до моделі *когнітивної резильєнтності*. Під когнітивною резильєнтністю ми розуміємо стійкість учня до misleadingness у різних модальностях, здатність утримувати або відновлювати коректне reasoning після помилки, а також збереження мотиваційно-когнітивної активності після негативного початкового сигналу.

У межах цього підходу цифровий двійник оцінює не лише фінальну правильність відповіді, а й здатність учня:

- витримувати вплив misleading stimulus;
- виявляти помилку у власному міркуванні;
- коригувати відповідь після уточнювального запитання;

4. зберігати когнітивну стабільність у мультимодальному середовищі;
5. не втрачати мотивацію після першої невдачі.

Метою статті є теоретичне, математичне та інформатичне обґрунтування розвитку MRPIA як цифрового двійника учня, а також побудова моделі неперервного оцінювання когнітивної резильєнтності, що базується на трьох субіндексах, adaptive follow-up questioning, spline-апроксимації та інтегральному resilience-індексі.

2. Огляд літератури

Ідея моделювання учня має довгу історію в дослідженнях intelligent tutoring systems, де student model виступала ядром адаптивності. Класичні ITS оцінювали, що саме учень уже знає, де припускається помилок і яку дидактичну дію слід запропонувати далі. У міру розвитку adaptive e-learning systems learner model стала багатшою: почали враховуватися не лише результати тестів, а й темп роботи, патерни взаємодії, мотивація, історія помилок і контекст виконання завдань.

Новіший етап цього розвитку – концепція цифрового двійника студента (digital twin of student, DTS), яку можна розглядати як інтегровану, динамічну і семантично інтерпретовану learner model. У такому підході цифровий двійник акумулює академічні, поведінкові, компетентнісні та цільові сигнали, а також виступає прескриптивним механізмом для побудови персоналізованої траєкторії навчання [1]. На відміну від традиційної learner model, яка часто залишається внутрішнім модулем системи, цифровий двійник розглядається як постійно оновлюване цифрове представлення учня, здатне поєднувати спостереження, прогноз і рекомендацію.

Табл. 1 показує, що релевантна література для MRPIA формується на перетині кількох дослідницьких ліній. Перша група охоплює власне цифрові двійники учня та персоналізоване навчання. Тут цифровий двійник подається як динамічна модель студента, що інтегрує результати, компетентності, поведінкові сигнали, цілі та персоналізовані траєкторії навчання. Водночас більшість таких робіт або мають концептуально-архітектурний характер, або спираються на симуляції, а не на прямі шкільні експерименти.

Друга група літератури стосується ширших освітніх digital twin ecosystems, включно з IoT-enabled smart learning, immersive training та education-and-training frameworks. Ці праці важливі для архітектурного обґрунтування, але вони здебільшого описують інфраструктуру, взаємодію пристроїв, моделювання середовищ або галузеві training applications, а не власне тонке когнітивне оцінювання учня в шкільному класі.

Третя група походить з healthcare та mental-health digital twins. Саме тут цифровий двійник найчастіше

тракується як неперервна, прогностична та адаптивна модель людини, що оновлюється за мультимодальними сигналами. Для нашої роботи ця лінія важлива насамперед концептуально: вона показує, як можна перейти від разового вимірювання до динамічного оцінювання латентного стану. Проте ці праці не дають готової шкільної моделі оцінювання когнітивної резильєнтності.

Окрема група охоплює student knowledge assessment та machine-learning-based evaluation of student performance. Ці дослідження корисні тим, що показують перехід від грубого підсумкового бала до data-driven оцінювання знань. Однак і тут у більшості випадків ідеться або про prediction of academic performance, або про класифікацію рівня знань, а не про цифровий двійник як систему адаптивного дооцінювання після помилки.

Для нашої теми особливо важливою є група *когнітивного оцінювання*. Тут уже існують добре відомі інструменти для вимірювання working memory, executive functions, processing speed, inhibition та attention у дітей і підлітків. До найуживаніших належать NIH Toolbox Cognition Battery, CANTAB, Digit Span, Corsi Block-Tapping Test, Stroop, CPT та інші батареї. Проте ці інструменти зазвичай оцінюють окремі когнітивні функції як тести або батареї, а не як компонент цифрового двійника учня, який оновлюється в реальному часі в навчальному середовищі.

Саме тому MRPIA позиціонується не як ще один cognitive test battery і не як ще одна система prediction of grades. Його новизна полягає в поєднанні цифрового двійника учня, adaptive assessment, мультимодального misleadingness analysis, когнітивного оцінювання та нелінійного spline-based grading в одній explainable framework. У цьому сенсі таблиця показує не лише наявні джерела, а й наукову прогалину: існує багато суміжних напрямів, але відсутня цілісна school-oriented система, що оцінює когнітивну резильєнтність учня як динамічний latent state.

Важливою математичною лінією розвитку learner modeling є knowledge tracing, тобто оцінювання того, як змінюється стан знань учня в часі на основі послідовності навчальних подій. Традиційні підходи до КТ спиралися на ймовірнісні моделі типу Bayesian Knowledge Tracing, а пізніше з'явилися глибокі нейронні моделі, зокрема Deep Knowledge Tracing [2]. Сучасні огляди knowledge tracing підкреслюють, що ці моделі добре працюють для відстеження correctness trajectories, але значно слабше відображають пояснюваність, структуру помилки, латентну невизначеність і якість reasoning [8, 9, 10]. Для цифрового двійника учня однієї correctness-based логіки недостатньо. Якщо learner model має справді відображати стан учня, вона повинна працювати не лише з правильністю, а й з частковим розумінням, стабільністю відповіді, здатністю до самокорекції та стійкістю до misleadingness. У цьому сенсі explainable knowledge tracing є важливим містком між традиційною психометрією, аналітикою

Table 1: Основні напрями літератури, релевантні для MRІІА як цифрового двійника учня

Тематична група	Хронологія	Ключові публікації	+ / -
Цифрові двійники учня персоналізоване навчання	та 2024–2026	Omrany et al. (2024); Kabashkin (2025); Mishra et al. (2026, Chapter 8)	-
Цифрові двійники в освіті, тренуванні та smart learning ecosystems	2024–2026	Omrany et al. (2024); Lu and Hu (2025); Mishra et al. (2026, Chapter 8)	-
Цифрові двійники в healthcare та cog-nitive / mental-health twins	та cog- 2023–2025	Spitzer et al. (2023); Venkatesh et al. (2024); Doraiswamy et al. (2025); Nadeem et al. (2025); Ringeval et al. (2025)	-
Моделі оцінювання знань учнів та student knowledge assessment	та stu- 2015–2024	Piech et al. (2015); Zafari et al. (2021); Alruwais and Zakariah (2023); Wang et al. (2024); Shen et al. (2024); Bai et al. (2024)	-
Adaptive assessment, partial credit, formative assessment	1987–2026	Adams (1987); Weiss (2015); Khosla et al. (2026)	-
Когнітивне оцінювання	1920–2026	класичні та сучасні тести: NIH Toolbox Cognition Battery; CANTAB; Digit Span; Corsi Block-Tapping Test; Stroop Test; Continuous Performance Test (CPT); CABG-WM; Souissi et al. (2022)	+
Multimodal misleadingness, visual literacy, misleading charts	2015–2026	Pandey et al. (2015); Lee et al. (2016); Correll et al. (2020); Lo et al. (2022); Driessen et al. (2022); Rho et al. (2026)	-
Мультимодальне reasoning та multimodal AI	2022–2025	Dai and Brell-Cokcan (2022); Bharti et al. (2024); Zhou et al. (2025)	-
Semantic learner modeling та adaptive e-learning systems	2024–2025	Benhamdi et al. (2024); Shen et al. (2024); Kabashkin (2025)	-
Сучасні методи оцінювання в школі	2025–2026	Thianwan and Srikoon (2026); Schult et al. (2026); research on classroom assessment in EFL contexts (2025); game-based and cognitively informed assessment environments (2025); AI-awareness module evaluation in school classrooms (2025)	+

Примітка. Позначення “+” означає наявність задокументованого практичного використання у шкільному середовищі; “-” означає, що публікація або група публікацій не містить явно описаного school-based експерименту чи апробації у реальному шкільному класі.

Table 2: Розширене трактування misleadingness

Тип	Типові прояви
Візуальна misleadingness	misleading charts, truncated axes, dual axes, 3D distortion, salience bias, перцептивні ілюзії
Текстова misleadingness	надлишково навантажені інструкції, неоднозначні формулювання, мовні пастки, семантична двозначність
Математично-логічна misleadingness	хибні евристики, псевдокоректні доведення, приховані логічні зсуви, формально правдоподібні помилки
Змішана мультимодальна misleadingness	конфлікти між текстом і рисунком, суперечність між формулою, графіком і словесним описом

Примітка. У цій роботі misleadingness трактується ширше, ніж лише visual distortion: вона охоплює також текстові, математично-логічні та змішані мультимодальні джерела когнітивної помилки.

навчання та AI-based digital twin systems.

Окрема дослідницька традиція (Табл. 2 та 3) вивчає visualization literacy та misleading data visualizations. Вона показує, що навіть добре навчені користувачі можуть систематично помилятися, коли графіки містять truncated axes, inverted axes, dual axes, 3D distortions, irregular intervals або інші misleading visual features [3, 4, 5, 6, 7, 11]. Дослідження visualization literacy особливо важливі для освіти, оскільки вони демонструють: інтерпретація графічної інформації є не автоматичною, а когнітивно вразливою навичкою.

3. Мета дослідження, завдання та гіпотези

Метою статті є теоретичне, математичне та прикладне обґрунтування розвитку MRІІА як цифрового двійника учня для адаптивного оцінювання когнітивної резильєнтності, тобто стійкості до помилок, misleadingness, когнітивних ілюзій, логічних пасток і частково

некоректних стратегій міркування в мультимодальному освітньому середовищі.

На відміну від традиційного оцінювання, орієнтованого переважно на фінальну правильність відповіді, запропонований підхід виходить з того, що освітньо значущим є не лише результат, а й траєкторія досягнення цього результату: первинна реакція учня, стійкість до misleading stimuli, здатність до самокорекції, проходження уточнювальних кроків, а також стабільність reasoning після корекції. Саме тому центральним об’єктом аналізу в цій роботі виступає не окремий тестовий бал, а інтегральний resilience-індекс цифрового двійника, побудований на основі трьох субіндексів – візуального, текстового та математично-змішаного.

Для досягнення поставленої мети у статті сформульовано такі завдання:

1. концептуалізувати MRІІА як цифровий двійник учня в контексті adaptive assessment;
2. узагальнити літературу з learner modeling, knowledge tracing, digital twins, cognitive assessment, misleadingness studies та сучасних методів освітнього оцінювання;
3. запропонувати математичну модель інтегрального індексу когнітивної резильєнтності на основі трьох субіндексів;
4. обґрунтувати використання adaptive follow-up questioning як механізму відмови від жорсткого бінарного оцінювання;
5. описати spline-апроксимацію як основу неперервного внутрішнього оцінювання та

shape-preserving калібрування до зовнішньої 12-бальної шкали;

6. проаналізувати емпіричні візуалізації, побудовані на основі даних Book2026.csv, зокрема структуру субіндексів, pre-post зміну інтегрального показника та вікову динаміку ефекту;
7. окреслити педагогічні, етичні та методологічні наслідки впровадження такого підходу в освітню практику.

У межах дослідження розглядаються такі дослідницькі питання:

- RQ1: Чи може MRPIA бути інтерпретований як цифровий двійник учня, а не лише як генеративний або консультаційний AI-інструмент?
- RQ2: Чи є когнітивна резильєнтність продуктивнішою рамкою для оцінювання, ніж традиційна correctness-based логіка?
- RQ3: Чи дозволяє інтегральний індекс, побудований на трьох субіндексах, точніше описувати мультимодальний профіль учня?
- RQ4: Чи є adaptive questioning та spline-калібрування теоретично виправданою основою для переходу від бінарного оцінювання до неперервного?
- RQ5: Чи виявляється позитивна динаміка інтегрального індексу та чи залежить вона від віку й instructional condition?

Відповідно до мети та поставлених завдань формуються такі робочі гіпотези:

- H1.** Інтегральний індекс когнітивної резильєнтності, побудований як функція трьох субіндексів (V , T , M), є змістовно адекватнішою моделлю учнівського стану, ніж однофакторне оцінювання на основі фінальної відповіді.
- H2.** Adaptive follow-up questioning дозволяє розрізнити поверхневу помилку та глибоке нерозуміння, а отже зменшує ризик педагогічно грубого заниження оцінки.
- H3.** Неперервне spline-калібрування внутрішнього інтегрального індексу до 12-бальної шкали краще відображає часткове розуміння, самокорекцію та когнітивну стійкість, ніж лінійне або сходишкове порогове оцінювання.
- H4.** Після взаємодії з цифровим двійником інтегральний resilience-індекс демонструє додатну pre-post динаміку, що інтерпретується як ефект adaptive support.

H5. Вікові групи різняться за домінуючим типом вразливості до misleadingness: для молодших учнів більш характерна візуальна вразливість, тоді як для старших зростає роль текстово-аналітичної та математично-логічної misleadingness.

Слід підкреслити, що в цій роботі гіпотези мають статус теоретично та емпірично мотивованих робочих гіпотез. Вони спираються на подані візуалізації, описові статистики та загальну логіку побудови цифрового двійника, але потребують подальшої перевірки на ширших вибірках і в межах окремих inferential designs.

4. Methods

Методологічною основою запропонованого підходу є поєднання кількох дослідницьких традицій: adaptive testing, partial credit measurement, learner modeling, knowledge tracing та цифрових двійників учня. Для нашої аргументації особливо важливою є лінія, пов'язана з adaptive testing і partial credit measurement, у межах якої освітній результат не зводиться до жорсткої дихотомії “правильно / неправильно”, а інтерпретується як перехід між впорядкованими рівнями часткового досягнення. Adaptive testing підлаштовує складність і характер завдань до поточного estimated ability учня, тоді як partial credit models дозволяють оцінювати ordered response categories замість однієї бінарної ознаки [12, 13, 14].

У контексті цифрового двійника це має принципове значення: система повинна фіксувати не лише те, чи була відповідь правильною, а й те, наскільки далеко учень просунувся у відновленні правильного reasoning після початкової помилки, якою була його траєкторія самокорекції, і чи зберігалася стійкість міркування після додаткової перевірки. Саме тому в цій роботі пропонується оцінювати не просто *knowledge state*, а *resilience state*, тобто здатність учня протистояти misleadingness у різних модальностях, витримувати когнітивне навантаження, виявляти і виправляти власні помилки та утримувати коректне reasoning після уточнення.

На цій підставі інтегральний індекс цифрового двійника доцільно трактувати не як індекс “правильності” у вузькому сенсі, а як індекс *когнітивної резильєнтності*. Його зміст полягає не лише в тому, чи отримав учень правильну фінальну відповідь, а в тому, наскільки стійко він проходить через misleadingness, виявляє помилки, коригує reasoning і зберігає узгодженість мислення в різних модальностях. У цій роботі такий індекс подано через три базові субіндекси, узагальнені в табл. 3.

Таке розширення є принциповим, оскільки дає змогу охопити не лише misleading charts, а й значно ширший клас освітніх помилок: математично хибну логіку, псевдокоректні доведення теорем, задачі з прихованою пасткою, конфлікти між текстом і рисунком, а також частково правильні, але нестійкі міркування. Отже,

Table 3: Базові субіндекси когнітивної резильєнтності в моделі цифрового двійника

Субіндекс	Що саме відображає
Visual resilience	стійкість до misleading visual cues, misleading charts, перцептивних і просторових викривлень
Textual resilience	стійкість до misleading verbal/textual cues, перевантажених інструкцій, неоднозначних формулювань і мовних пасток
Mathematical / mixed resilience	стійкість до математично-хибної логіки, псевдокоректних доведень, прихованих логічних зсувів і змішаних мультимодальних конфліктів

Примітка. Інтегральний індекс формується як узагальнення трьох субіндексів і відображає не лише правильність відповіді, а й стійкість учня до когнітивно оманливих умов.

цифровий двійник оцінює не лише кінцевий результат, а структуру когнітивної стійкості, на якій цей результат ґрунтується. Це означає, що оцінювання в моделі цифрового двійника є процесуальним, а не лише результативним. Помилка на першому кроці не повинна автоматично трактуватися як “незнання”: вона може бути поверхневою, а не концептуальною; короткочасною, а не стабільною; misleading-induced, а не зумовленою відсутністю компетентності.

4.1. Математична модель інтегрального resilience-індексу

Нехай для учня i

$$V_i \in [0, 1], \quad T_i \in [0, 1], \quad M_i \in [0, 1]$$

– відповідно візуальний, текстовий та математично-змішаний субіндекси стійкості.

Тоді інтегральний індекс визначаємо як геометричне середнє:

$$H_i = (V_i T_i M_i)^{1/3}. \quad (1)$$

Перевага геометричного середнього полягає в тому, що низький рівень у будь-якій одній модальності не маскується високим рівнем у двох інших. Це особливо важливо для цифрового двійника, оскільки в реальному навчанні слабкість у текстовому reasoning або математично-логічному reasoning не повинна зникати за рахунок сильнішого візуального сприйняття.

У зваженому варіанті формулу можна подати як

$$H_i = V_i^\alpha T_i^\beta M_i^\gamma, \quad \alpha + \beta + \gamma = 1, \quad (2)$$

де ваги можуть бути адаптивно налаштовані відповідно до віку, предметної галузі або типу навчального завдання.

На відміну від бінарного тестового оцінювання, де неправильна відповідь автоматично знижує результат до нуля в конкретному вузлі, у цифровому двійнику використовується цикл уточнювального діалогу. Нехай після початкової відповіді система ставить k

уточнювальних запитань, де

$$k \in \{3, 4, 5\}.$$

Нехай далі $r_{ij} \in [0, 1]$ – оцінка відповіді учня i на j -те уточнення, а w_j – відповідна вага цього кроку. Тоді адаптивно уточнений показник можна записати як

$$A_i = \sum_{j=1}^k w_j r_{ij}, \quad \sum_{j=1}^k w_j = 1. \quad (3)$$

Принципово важливо, що на внутрішньому рівні цифровий двійник не зводиться до кількох дискретних значень на кшталт 0, 0.5 або 1. Він працює з неперервною latent score, а зазначені рівні є лише грубою педагогічною інтерпретацією окремих моментів оцінювання. Реально відповідь після кількох уточнень може набувати проміжних значень на кшталт 0.18, 0.36, 0.61 або 0.82, що дає змогу тонше розрізняти поверхневу нестійкість, часткове розуміння та відновлене коректне reasoning.

Саме це уможливорює spline-апроксимацію відповідей як основу неперервності оцінювання. Нехай послідовність уточнювальних кроків або когнітивних реакцій учня задається точками

$$\{(t_j, y_j)\}_{j=1}^m,$$

де t_j – номер кроку, а y_j – спостережуваний рівень локальної стабільності reasoning. Тоді неперервна оцінка моделюється функцією

$$\hat{y}_i(t) = \sum_{r=1}^R c_r B_r(t), \quad (4)$$

де $B_r(t)$ – spline-базисні функції, а c_r – їхні коефіцієнти.

Для освітнього оцінювання критично важливо, щоб апроксимація була *shape-preserving*, тобто не породжувала паразитних штучних коливань. Саме тому доцільно використовувати монотонні або *shape-preserving cubic splines*. У такій рамці цифровий двійник не стрибає між різкими дискретними оцінками, а формує гладку когнітивну траєкторію учня, яка краще відображає процес становлення, уточнення і стабілізації відповіді.

4.2. Архітектура цифрового двійника

Першим кроком у побудові інноваційної архітектури цифрового двійника є процес нелінійного калібрування 12-бальної шкали. Оскільки внутрішня оцінка цифрового двійника є неперервною, то

$$H_i \in [0, 1],$$

для педагогічної інтерпретації доцільно розділити шкалу наступним чином:

$$G_i = \Phi(H_i), \quad G_i \in [1, 12], \quad (5)$$

де Φ – монотонне shape-preserving spline-відображення.

У найпростішому лінійному варіанті можна було б використати

$$G_i = 12H_i.$$

Однак для реального оцінювання це надто грубо, оскільки розподіл учнівських індексів не є лінійним. Крім того часткове відновлення міркування після помилки не повинно оцінюватися так само, як первинно повна правильність, а, різні ділянки шкали мають різну педагогічну чутливість. Саме тому spline-калібрування є більш адекватним рішенням, оскільки воно зберігає неперервну внутрішню структуру learner state і водночас переводить її у зовнішньо зрозумілу шкільну шкалу.

Таким чином, ми можемо виконати наступний крок – представити цифровий двійник учня i в момент часу t у векторному полі у вигляді вектора

$$\mathbf{x}_i(t) = [\mathbf{a}_i(t), \mathbf{c}_i(t), \mathbf{b}_i(t), \mathbf{m}_i(t), \mathbf{g}_i(t), \mathbf{h}_i(t), \mathbf{s}_i(t)]^\top,$$

де:

- $\mathbf{a}_i(t)$ – академічні індикатори;
- $\mathbf{c}_i(t)$ – компетентнісний профіль;
- $\mathbf{b}_i(t)$ – поведінковий профіль;
- $\mathbf{m}_i(t)$ – мотиваційно-залученісний профіль;
- $\mathbf{g}_i(t)$ – освітні цілі та орієнтації;
- $\mathbf{h}_i(t)$ – рукописний когнітивний слід;
- $\mathbf{s}_i(t)$ – spline-ознаки динаміки відповіді.

Векторний підхід дозволяє враховувати не тільки абсолютні значення, але й траєкторію отримання правильної відповіді. Тоді еволюцію цифрового двійника запишемо у такому вигляді

$$\mathbf{x}_i(t+1) = F(\mathbf{x}_i(t), e_t, \mathbf{u}_t, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t, \quad (6)$$

де e_t – нова освітня подія, \mathbf{u}_t – інтервенція системи, а F – оператор переходу стану.

Система працює з непрямыми спостереженнями

$$\mathbf{y}_i(t) = H(\mathbf{x}_i(t)) + \boldsymbol{\eta}_t, \quad (7)$$

які включають правильність відповіді, час, кількість уточнень, типи помилок, рукописні артефакти та інші сигнали. Завдання цифрового двійника – оцінювати прихований когнітивний стан за цією послідовністю спостережень.

4.3. Spline-ознаки поведінкової та когнітивної динаміки

Поведінка учня в цифровому середовищі рідко є лінійною. Після швидкого успіху може настати plateau, за ним – спад, а потім новий ріст після вдалої підказки або уточнення. Тому у векторному просторі саме spline-апроксимація відкриває унікальну можливість виділити ознаки вищого порядку:

- локальний нахил $\widehat{y}'_i(t)$ – швидкість прогресу або деградації;
- друга похідна $\widehat{y}''_i(t)$ – прискорення або уповільнення;
- точки перегину – моменти когнітивної нестабільності;
- залишки апроксимації – можливі аномальні реакції.

Для цифрового двійника це означає, що оцінювання може спиратися не лише на один бал, а й на геометрію траєкторії міркування. Саме в цьому полягає принципова відмінність від звичайного тесту.

З інформатичної точки зору MRPIA доцільно описувати як багаторівневу систему:

1. **Рівень збору даних:** LMS, тести, журнали активності, текстові відповіді, рукописні артефакти.
2. **Рівень нормалізації:** очищення, синхронізація, learner events.
3. **Семантичний рівень:** онтології тем, компетентностей, типів помилок, цілей.
4. **Рівень моделі стану:** learner model, knowledge tracing, spline dynamics.
5. **Рівень рекомендації:** adaptive feedback, sequencing, пояснення.
6. **Human-in-the-loop:** педагогічна валідація та корекція.

Особливо важливим є контур human-in-the-loop, оскільки навіть найкраща AI-модель не повинна остаточно витіснити педагогічне судження. Цифровий двійник має бути не заміною вчителя, а інструментом більш тонкої інтерпретації стану учня.

4.4. База даних

У результаті проведення педагогічного експерименту було сформовано та систематизовано базу даних, яка використовується для аналітичної інтерпретації моделі MRPIA як цифрового двійника учня. Опис цифрового twin-підходу, загальної логіки експерименту та відповідної інформаційної інфраструктури подано на ресурсі Міжнародної інноваційної школи [39].

Розширені матеріали та деталізовані дані можуть бути надані за запитом.

Для побудови емпіричного шару цієї роботи було використано дані педагогічного експерименту [39], які містять, зокрема, такі змінні:

- вік (age);
- навчальну умову (condition);
- візуальний субіндекс (V_index_pre);
- текстовий субіндекс (T_index_pre);
- математично-змішаний субіндекс (M_index_pre3);
- інтегральний індекс до втручання (НММІ_index_pre);
- інтегральний індекс після втручання (НММІ_index_post).

У межах цієї статті зазначений масив даних використовується як емпірична основа для демонстрації того, яким чином модель цифрового двійника може бути операціоналізована через систему субіндексів та інтегрального показника. Ці дані дають змогу простежити структуру трьох субіндексів, їхній внесок у побудову інтегрального показника, pre-post зміну НММІ, а також вікову динаміку позитивного ефекту цифрового двійника.

Таким чином, база даних виконує подвійну функцію: з одного боку, вона фіксує результати педагогічного експерименту, а з іншого – забезпечує аналітичний міст між теоретичною моделлю цифрового двійника та її кількісною реалізацією. Саме на цій основі в наступному підрозділі подано результати у вигляді системи візуалізацій, які відображають структуру субіндексів, розподіл інтегрального показника, pre-post динаміку та вікові відмінності позитивного ефекту.

5. Результати

Спираючись на детально окреслену вище методику дослідження, представимо у вигляді рис. 1 середні значення трьох базових субіндексів: візуального, текстового та математично-змішаного. Окремо розрахуємо інтегральний resilience-індексу на основі бази даних, отриманих під час педагогічного експерименту [39] ($n = 280$) ми отримали середні pretest-значення окремих компонентів:

$$\bar{V} = 0.6113, \quad \bar{T} = 0.5469, \quad \bar{M} = 0.5230,$$

тоді як середнє значення інтегрального індексу дорівнювало

$$\bar{H}_{pre} = 0.5528.$$

Ці значення є змістовно важливими, оскільки вони показують, що в досліджуваній вибірці найвищим

виявився саме візуальний субіндекс, тоді як найнижчим – математично-змішаний. Отже, вже на базовому рівні можна говорити про неоднорідність когнітивної стійкості за модальностями: учні в середньому краще витримують візуальні стимули, ніж складні математично-логічні або змішані misleading conditions. При цьому інтегральний індекс $\bar{H}_{pre} = 0.5528$ лежить нижче за \bar{V} , але близько до \bar{T} та \bar{M} , що узгоджується з геометричною природою агрегації в (1).

На відміну від арифметичного середнього, інтегральний показник у нашій моделі визначається мультиплікативній формі

$$H_i = (V_i T_i M_i)^{1/3}. \quad (8)$$

Така конструкція робить інтегральний індекс чутливим до “вузького місця”. Якщо хоча б один субіндекс є низьким, загальний показник зменшується сильніше, ніж у разі простого адитивного усереднення. Саме тому інтегральний resilience-profile учня не маскує слабкість в окремій модальності, а навпаки – робить її аналітично видимою. У контексті цифрового двійника це є методологічною перевагою, оскільки learner model має фіксувати не лише сумарний результат, а структуру когнітивної стійкості.

З огляду на отримані середні значення, можна сформулювати перше емпіричне спостереження, яке надалі підтримує дослідницьку гіпотезу про мультимодальну природу когнітивної резильєнтності: навіть до втручання система виявляє не один однорідний рівень “успішності”, а внутрішньо диференційований профіль, де візуальна, текстова та математично-змішана стійкість мають різні середні рівні. Отже, цифровий двійник працює не з одним “непрозорим” числом, а зі структурованою конфігурацією когнітивних каналів.

На рис. 2 показано розподіл трьох субіндексів і інтегрального показника. Цей рисунок є методологічно важливим, оскільки дозволяє побачити не лише середні значення, а й форму емпіричних розподілів, тобто варіативність, концентрацію та можливу асиметрію результатів. Для інтегрального pretest-індексу були отримані такі описові характеристики:

$$\begin{aligned} \min(H_{pre}) &= 0.1923, \\ Q_1 &= 0.4421, \\ \text{Med}(H_{pre}) &= 0.5414, \\ Q_3 &= 0.6514, \\ \max(H_{pre}) &= 1.0000. \end{aligned} \quad (9)$$

Ці параметри показують, що вибірка не є ні тривіальною однорідною, ні поляризованою. Медіана 0.5414 близька до середнього $\bar{H}_{pre} = 0.5528$, що свідчить про відсутність різко перекошеного розподілу. Водночас

Three resilience subindices and the integrated index

Assessment is based on resilience to misleadingness rather than binary correctness

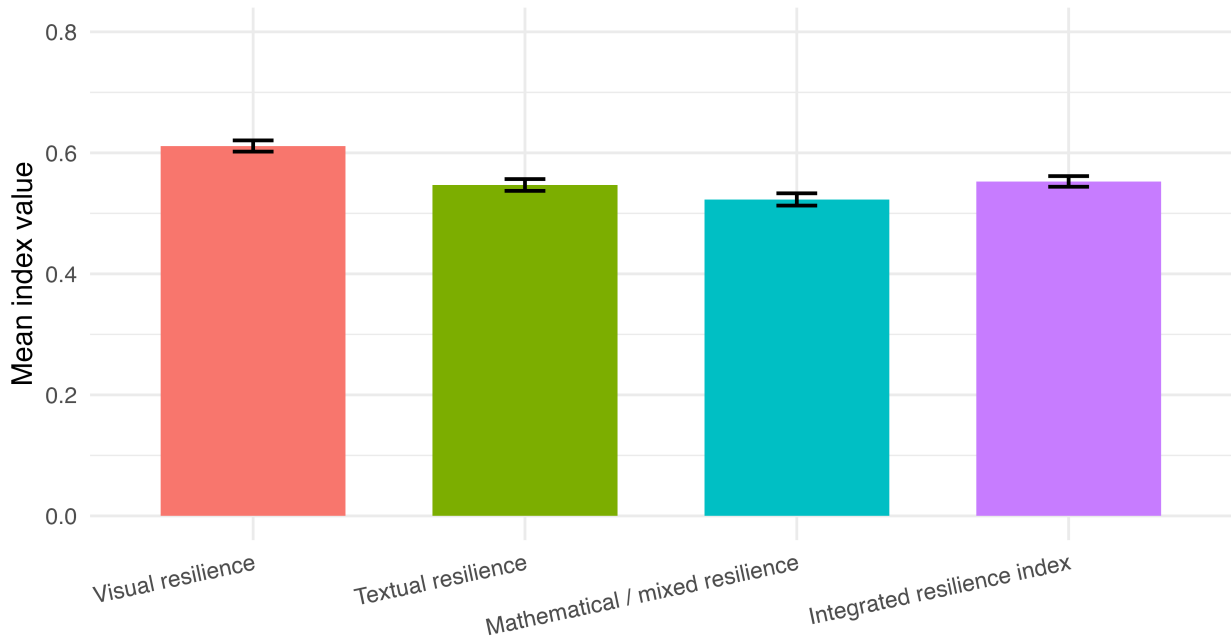


Figure 1: Три субіндекси стійкості (visual, textual, mathematical/mixed) та інтегральний resilience-індекс.

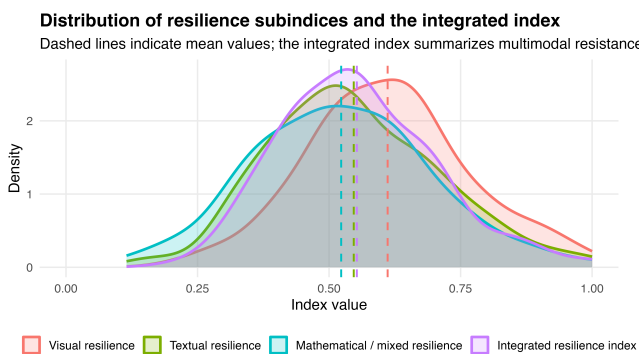


Figure 2: Розподіл трьох resilience-субіндексів та інтегрального індексу.

міжквартильний інтервал

$$IQR = Q_3 - Q_1 = 0.6514 - 0.4421 = 0.2093$$

вказує на достатньо виразну внутрішню варіативність. Це означає, що в межах вибірки присутні як учні з порівняно низькою стійкістю до misleadingness, так і учні з відносно сформованим resilience-profile.

У термінах learner modeling така картина особливо важлива. Вона означає, що цифровий двійник не накладає на всіх учнів єдину “норму”, а працює з реальною неоднорідністю когнітивних станів. Отже, розподіл субіндексів та інтегрального показника

підсилює головну тезу статті: система має оцінювати не тільки факт правильності, а й архітектуру когнітивної стабільності, тобто варіативний профіль вразливості та стійкості в різних модальностях.

5.1. Pre-post зміна інтегрального індексу

На рис. 3 показано зміну інтегрального показника до і після роботи цифрового двійника. Для всієї вибірки середнє значення інтегрального індексу зросло з

$$\bar{H}_{pre} = 0.5528$$

до

$$\bar{H}_{post} = 0.7732.$$

Відповідно середній приріст становив

$$\Delta \bar{H} = \bar{H}_{post} - \bar{H}_{pre} = 0.2204.$$

Цей результат є кількісно суттєвим. Якщо інтерпретувати ΔH_i як індикатор зміни latent resilience state, то спостережуваний приріст означає не просто механічне підвищення бала, а зміщення внутрішнього когнітивного профілю в бік більшої стійкості до misleadingness. Це узгоджується з логікою цифрового двійника, де система не зупиняється на фіксації помилки, а використовує adaptive follow-up questioning, уточнення і корекцію reasoning.

Додатково розподіл приросту також є

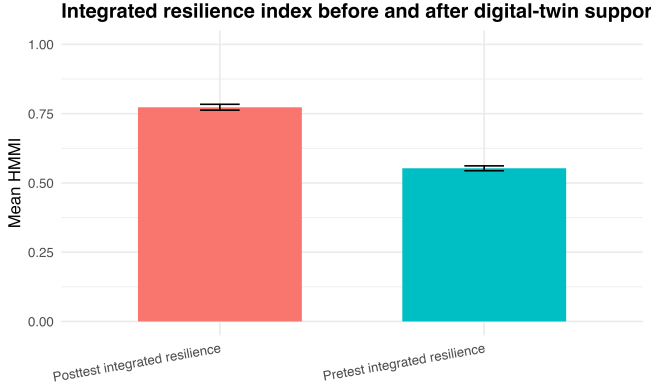


Figure 3: Pre–post зміна інтегрального resilience-індексу після втручання цифрового двійника.

інформативним:

$$\begin{aligned}
 \min(\Delta H) &= -0.3000, \\
 Q_1 &= 0.0640, \\
 \text{Med}(\Delta H) &= 0.2225, \\
 Q_3 &= 0.3524, \\
 \max(\Delta H) &= 0.7077.
 \end{aligned} \tag{10}$$

Отже, хоча в окремих випадках зафіксовано і від’ємні зміни, медіана 0.2225 і додатний міжквартильний інтервал свідчать, що для більшості учнів ефект був позитивним. Це дозволяє трактувати observed pre–post dynamics як систематичну, а не випадкову.

Формально приріст визначався як

$$\Delta H_i = H_{i,\text{post}} - H_{i,\text{pre}}. \tag{11}$$

Саме ця величина надалі використовується як центральний індикатор ефективності адаптивної підтримки цифрового двійника.

5.2. Міжумовні відмінності приросту інтегрального індексу

Для перевірки того, чи пов’язаний приріст інтегрального показника з instructional condition, було проведено однофакторний дисперсійний аналіз. Для pretest-рівня статистично значущих відмінностей між умовами виявлено не було:

$$F(2, 277) = 1.031, \quad p = 0.358,$$

що є важливим методологічним результатом. Він означає, що групи не відрізнялися істотно на старті за інтегральним resilience-індексом, а отже подальші відмінності приросту не можна просто пояснити початковою нерівністю груп.

Натомість для приросту ΔH було отримано виразний умовний ефект:

$$F(2, 277) = 97.66, \quad p < 2 \times 10^{-16}.$$

Це один із найсильніших статистичних результатів у дослідженні. Він показує, що instructional condition суттєво пов’язана зі зміною інтегрального індексу. Таким чином, дані підтримують гіпотезу про те, що різні режими підтримки не є педагогічно еквівалентними, а повноцінний digital-twin-guided support асоційований із вищим приростом resilience-індексу, ніж стандартний або частковий режим.

У термінах дослідницької логіки цей результат є ключовим для подальших висновків: якщо pretest-групи статистично не відрізнялися, але posttest-динаміка і приріст відрізняються дуже суттєво, то це підтримує центральну гіпотезу статті про ефективність adaptive digital-twin framework як механізму розвитку когнітивної резильєнтності.

На початку експерименту вікова динаміка показувала, що молодші класи допускають значно більше помилок із завданнями, які містять візуальні ілюзії, але ми отримали для усіх вікових груп позитивний ефект використання цифрового двійника (Рис. 4). Саме ці показники дозволяють перейти від загальної тези про позитивний ефект цифрового двійника до більш тонкого аналізу: у яких вікових групах ефект є виразнішим, як він змінюється залежно від типу підтримки, та які когнітивні механізми можуть стояти за цими змінами.

Важливо зауважити, що нелінійний аналіз вікового ефекту було виконано за допомогою generalized additive model:

$$H_i \sim s(\text{Age}_i) + \text{Gender}_i + \text{Condition}_i + \text{GradeGroup}_i. \tag{12}$$

Отримана модель для pretest-індексу показала статистично значущий нелінійний ефект віку:

$$\chi^2 = 41.82, \quad \text{edf} = 3.181, \quad p = 8.43 \times 10^{-6},$$

при поясненій частці девіації

$$\text{Deviance explained} = 30.7\%.$$

Це означає, що вік не можна трактувати як лінійну другорядну ковариату. Навпаки, він відіграє роль нелінійного структурного фактора learner model.

Крім того, у цій моделі ефект instructional condition також виявився значущим для однієї з умов:

$$\beta_{\text{generic}} = 0.2569, \quad p = 0.0345,$$

тоді як для control-умови було зафіксовано гранично значущий ефект:

$$\beta_{\text{control}} = 0.2112, \quad p = 0.0761.$$

Хоча ці коефіцієнти слід інтерпретувати обережно з огляду на параметризацію моделі, вони все ж вказують, що instructional mode пов’язаний із систематичними відмінностями когнітивної стійкості.

Додатково лінійна модель взаємодії віку й статі для

pretest-індексу показала позитивний основний ефект віку:

$$\beta_{\text{age}} = 0.0339, \quad p = 3.16 \times 10^{-5},$$

а також значущу взаємодію

$$\beta_{\text{age} \times F} = -0.0212, \quad p = 0.00462.$$

Зазначену модель апріорі ми трактуємо у контексті особливості кодування факторів, тому ми припускаємо, що вікова динаміка когнітивної резильєнтності може бути різною для різних підгруп учнів. Змістовно це добре узгоджується з гіпотезою про зміну домінуючого типу misleadingness з віком. Молодші учні, ймовірно, більш чутливі до візуальної misleadingness, тоді як зі зростанням віку посилюється роль текстово-аналітичної та математично-логічної вразливості. Саме тому позитивний ефект цифрового двійника не є монотонним і не може бути зведений до єдиного усередненого “бонусу”: він є віково-чутливим і, ймовірно, модально-специфічним.

Логіка цього аналізу полягає в тому, що вік тут не розглядається як другорядна демографічна характеристика. Навпаки, він виступає як змінна, пов’язана з відмінностями у структурі когнітивної вразливості. Молодші учні можуть бути більш чутливими до візуальної misleadingness, тоді як старші – до текстово-аналітичної або математично-логічної misleadingness. Отже, вікова динаміка на рисунку може відображати не просто різницю у “силі ефекту”, а зміну типу когнітивних труднощів, на які цифровий двійник реагує адаптивно.

Таким чином, ми маємо аналітичну рамку для подальшого тестування віково-умовних взаємодій. У цьому сенсі рис. 4 виконує подвійну функцію. З одного боку, він ілюструє, що позитивний ефект цифрового двійника не є рівномірним по всьому віковому спектру. З іншого – він обґрунтовує необхідність подальшого моделювання, у якому вік має розглядатися як важливий параметр калібрування learner model, adaptive questioning та spline-based grading. Також впадає в око те, що крива, яка відповідає повноцінному digital-twin support, систематично лежить вище за інші умови майже в усьому досліджуваному віковому діапазоні. Це означає, що повна модель цифрового двійника – на відміну від часткової або стандартної підтримки – пов’язана з найбільшим приростом інтегрального resilience-індексу. Причому цей ефект не є локальним або випадковим для одного віку: він спостерігається як у молодших групах, так і в середньому підлітковому інтервалі.

Другий важливий момент полягає в тому, що вікова крива ефекту не є монотонною. На графіку видно локальні підйоми, що вказує: цифровий двійник не просто “додає однаковий бонус” усім учням, а взаємодіє з віковим профілем когнітивної вразливості. Іншими словами, ефект цифрового двійника, ймовірно,

є найбільшим там, де природна вікова конфігурація помилок найбільш чутлива до adaptive clarification, поетапної корекції та мультимодальної підтримки.

Особливо цікавою є гіпотеза про зміну домінуючого типу misleadingness з віком. Молодші учні, ймовірно, більш вразливі до візуальної misleadingness: яскравих, просторово-салентних, перцептивно домінуючих стимулів. У старших учнів слабке місце може зміщуватися до аналітичної misleadingness: перевантажених текстових формулювань, прихованих логічних зсувів, псевдокоректних доведень, багатокрокових символічних переходів та надлишку формул. У термінах субіндексів це означає, що для молодших груп більшу роль може відігравати V_i , тоді як для старших важливішими стають T_i та особливо M_i .

Статистично цей ефект доцільно моделювати як

$$\Delta H_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Cond}_i + \beta_3 (\text{Age}_i \times \text{Cond}_i) + \varepsilon_i, \quad (13)$$

або, у нелінійному варіанті, через GAM:

$$\Delta H_i = \beta_0 + s(\text{Age}_i) + \text{Cond}_i + s(\text{Age}_i, \text{by} = \text{Cond}_i) + \varepsilon_i, \quad (14)$$

де $s(\cdot)$ – гладка spline-функція.

У психофізіологічному сенсі це також узгоджується з тим, що у молодшому віці сприйняття частіше керується безпосередньою наочністю, тоді як у старшому зростає когнітивне навантаження абстрактними структурами, формулами, доведеннями та текстово-символічними переходами. Відповідно цифровий двійник може виступати віково-чутливим механізмом компенсації різних типів когнітивної вразливості.

6. Нелінійне spline-калібрування та adaptive grading

Ми здійснюємо нелінійне калібрування інтегрального індексу до 12-бальної шкали, яка є офіційною в Україні. Як показано в табл. 4, у різних країнах використовуються формально відмінні системи оцінювання, однак у більшості випадків вони мають спільну рису: зовнішній результат подається через дискретні категорії, порогові інтервали або фіксовані рівні досягнення, що і зумовлює потребу в spline-калібруванні, яке дозволяє узгодити внутрішню неперервну модель цифрового двійника із зовнішньою дискретною системою шкільного оцінювання. У цьому контексті принципово важливим для всієї логіки роботи цифрового двійника є рис. 5, оскільки він наочно демонструє перехід від внутрішньої неперервної learner model до зовнішньої педагогічно зрозумілої 12-бальної шкали. У традиційній шкільній практиці оцінювання зазвичай має дискретний характер: відповідь або зараховується, або ні; часткова правильність фіксується грубо; а помилка на ранньому етапі нерідко непропорційно знижує підсумковий бал. Натомість у

Positive effect of the digital twin across age

Gain in HMMI (posttest minus pretest) by age and instructional condition

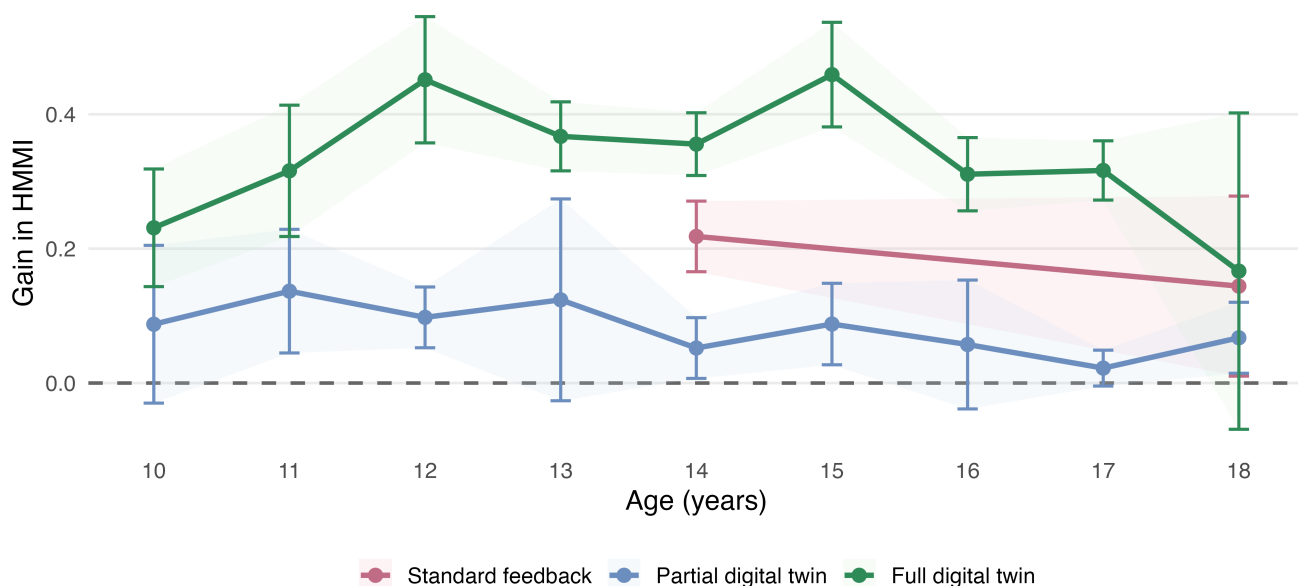


Figure 4: Позитивний ефект цифрового двійника за віком: приріст інтегрального індексу (posttest minus pretest) як функція віку та instructional condition.

Table 4: Систематизація національних шкал оцінювання в Європі

Група країн	Типова шкала
Німецькомовні країни	Austria: 1–5; Germany: 1.0–6.0; Switzerland: 6–1
Франкомовні та романські системи	France: 0–20; Belgium: 0–20; Luxembourg: 0–20; Portugal: 0–20; Spain: 0–10; Italy: 18–30 (<i>e lode</i>)
Центральна та Східна Європа	Poland: 2.0–5.0; Czech Republic: 1–5; Slovakia: A–FX/F; Hungary: 1–5; Romania: 1–10; Bulgaria: 2–6; Croatia: 1–5
Північна Європа	Denmark: –3–12; Sweden: U/G/VG; Norway: A–F; Finland: 0–5; Iceland: 0–10
Англомовні системи	Ireland: 0–100%; Malta: A–F
ECTS reference	A, B, C, D/E, F/FX

Примітка. Таблицю укладено на основі [40].

МРІІА внутрішній стан учня подано як неперервний інтегральний індекс, що змінюється в інтервалі $[0, 1]$ і відображає не лише correctness у вузькому сенсі, а й стійкість до misleadingness, здатність до самокорекції, стабільність reasoning та узгодженість між різними модальностями. У міжнародному академічному контексті інтерпретація результатів навчання потребує зіставлення локальних національних шкал із узагальненими європейськими орієнтирами. На платформі Study in Europe подано порівняльний огляд університетських систем оцінювання в країнах Європи, де як референтну рамку використано ECTS-подібне зіставлення категорій “excellent”, “very good”, “good”, “pass” і “fail”. Зокрема, джерело показує,

Standard step grading versus adaptive digital-twin grading

The digital twin stays above the standard rule almost everywhere, while smoothing rigid threshold e

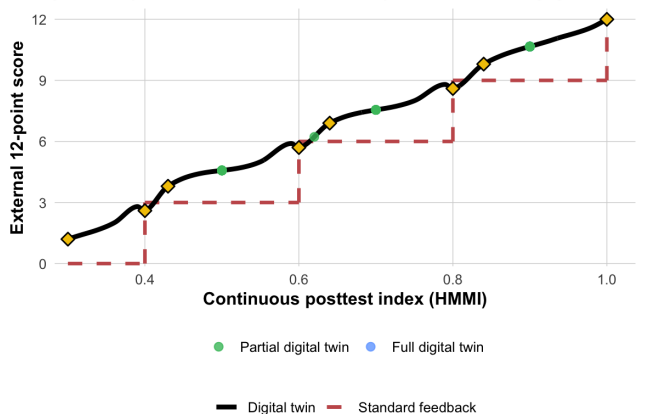


Figure 5: Порівняння жорсткого порогового оцінювання та adaptive digital-twin scoring.

що в Європі співіснують різноманітні шкали: числові шестибальні та п’ятибальні системи, 10-, 20- і 30-бальні шкали, а також літерні формати типу A–F. Це ще раз підкреслює, що проста лінійна інтерпретація балів є недостатньою, а перехід від внутрішнього неперервного індексу до зовнішньої шкали повинен бути каліброваним і контекстно чутливим [40].

Таким чином, ми задіяли не просто іншу техніку візуалізації, а іншу філософію оцінювання: від порогового “або перейшов, або ні” до діагностичного

“наскільки близько підійшов і чи здатен виправити помилку”. Саме тому зовнішня 12-бальна оцінка не повинна будуватися простим лінійним правилом типу $G = 12H$, а має бути результатом монотонного shape-preserving spline-калібрування:

$$G_i = \Phi(H_i), \quad H_i \in [0, 1], \quad G_i \in [1, 12], \quad (15)$$

де H_i – інтегральний posttest-індекс цифрового двійника для учня i , а Φ – монотонне нелінійне відображення, яке зберігає порядок учнів, але не нав’язує штучно лінійну структуру шкали.

Форма spline-кривої показує, що зв’язок між інтегральним індексом і підсумковою оцінкою є виразно нелінійним. На нижчих рівнях шкали навіть невеликий приріст індексу ще не означає різкого зростання зовнішнього бала, оскільки система інтерпретує початкові успіхи як крихкі й нестійкі. У середній зоні спостерігається більш плавний перехід: тут цифровий двійник “бачить” часткову сформованість компетентностей і допускає м’яке зростання оцінки. На верхніх рівнях крива стає знову крутішою, оскільки високі значення інтегрального індексу означають не просто правильні відповіді, а стабільну, відтворювану й мультимодально узгоджену стійкість до помилок.

Ключова інновація запропонованого підходу полягає в тому, що неперервний інтегральний індекс формується не після одного-єдиного акту відповіді, а після *адаптивного циклу уточнення*. Якщо учень дає неповну, нестійку або misleading-induced відповідь, система не обнуляє результат автоматично, а ставить різну кількість follow-up questions залежно від характеру помилки, віку, профілю субіндексів і поточного стану цифрового двійника. Отже, цифровий двійник оцінює не просто outcome, а *траєкторію реконструкції правильного мислення*.

У практичному сенсі це означає, що нелінійне spline-калібрування не лише покращує точність grading, а й відкриває шлях до раннього профілювання здібностей. Якщо протягом відносно раннього періоду розвитку система стабільно фіксує високі значення математично-змішаного субіндексу за збереження аналітичної стійкості у follow-up questioning, це може вказувати на потенціал до напрямів, де потрібне послідовне абстрактне reasoning. Якщо ж виразно домінує візуально-просторовий компонент, а учень демонструє стійкість до графічних та конфігураційних ілюзій, це може бути раннім індикатором іншого типу освітньої траєкторії.

6.1. Порівняння adaptive digital-twin scoring та жорсткого порогового оцінювання

Рис. 5 візуалізує різницю між двома принципово різними моделями освітнього оцінювання. Червона пунктирна сходинова функція репрезентує стандартний підхід, у якому неперервний внутрішній індекс фактично зводиться до грубих порогових зон. У

цій логіці до моменту формального подолання порога учень не отримує нового рівня оцінки.

Чорна крива, навпаки, подає адаптивне spline-калібрування цифрового двійника. Вона розташована вище стандартної сходинової функції майже на всій області визначення, хоча в околах самих точок стрибка тимчасово наближається до порогової межі. Інтерпретаційно це означає, що цифровий двійник не карає учня за першу ж помилку автоматичним недоотриманням бала, а відкриває можливість компенсувати частину втрати через уточнювальні adaptive follow-up questions. Якщо учень після початкової помилки демонструє самокорекцію, частково правильний хід міркування, відновлення правильної логіки або стійкість до misleadingness після додаткового запиту, система може присвоїти вищу оцінку раніше, ніж це дозволила б жорстка дискретна шкала.

З точки зору психометрії така логіка добре узгоджується з partial credit measurement. Освітній результат природно описується не бінарною ознакою, а переходом між сусідніми рівнями майстерності. У цьому сенсі цифровий двійник ближчий до partial credit logic, ніж до класичного right/wrong scoring.

Математично стандартну порогову схему можна подати як сходинову функцію

$$G_{\text{std}}(H) = \begin{cases} 0, & H < 0.4, \\ 3, & 0.4 \leq H < 0.6, \\ 6, & 0.6 \leq H < 0.8, \\ 9, & 0.8 \leq H < 1.0, \\ 12, & H = 1.0. \end{cases} \quad (16)$$

Натомість оцінювання цифрового двійника описується монотонною spline-каліброваною функцією

$$G_{\text{DT}}(H) = \Phi(H), \quad (17)$$

де Φ є shape-preserving monotone spline.

Якщо після первинної відповіді система ставить k_i додаткових діагностичних питань, де

$$k_i \in \{0, 1, 2, 3, 4, 5\},$$

то адаптивно скоригований бал для учня i можна подати як

$$A_i = \sum_{j=0}^{k_i} w_{ij} r_{ij}, \quad \sum_{j=0}^{k_i} w_{ij} = 1, \quad (18)$$

де r_{i0} – початкова відповідь, r_{ij} для $j \geq 1$ – відповіді на уточнення, а w_{ij} – ваги, що залежать від діагностичної цінності кожного кроку.

7. Ймовірнісне моделювання мотивації до навчання

У межах цифрового двійника мотивація до навчання розглядається не як фіксована психологічна риса, а як латентний стан, який може змінюватися в процесі взаємодії учня з навчальною системою. На відміну від грубих анкетних оцінок, MRPIA інтерпретує мотивацію через поведінкові індикатори: чи продовжує учень відповідати після помилки, як реагує на уточнювальні питання, чи демонструє самокорекцію, як швидко втрачає залученість і чи зберігає когнітивну наполегливість у складних задачах.

Нехай латентний мотиваційний стан учня M_i належить до трьох категорій:

$$M_i \in \{\text{Low}, \text{Medium}, \text{High}\}.$$

Тоді цифровий двійник оцінює апостеріорний розподіл

$$P(M_i = m \mid \mathcal{D}_i),$$

де \mathcal{D}_i – сукупність доступних даних про учня: результати первинної відповіді, реакції на уточнення, час відповіді, кількість спроб, частка самовиправлень і показники стійкості до misleadingness.

7.1. Динаміка мотивації: від первинної невдачі до snowball effect

У системі цифрового двійника мотивація до навчання розглядається як динамічний латентний стан, що змінюється під впливом успіху, помилки, зворотного зв'язку та історії попередніх взаємодій. Такий підхід узгоджується з expectancy-value logic, self-efficacy perspective та моделями learned helplessness. Якщо учень отримує негативний результат одразу після першої відповіді без можливості уточнення, це зменшує його очікування успіху в наступних спробах, послаблює відчуття контролю над результатом і може знижувати willingness to persist. Натомість adaptive follow-up questions створюють умови для partial recovery of agency: учень одержує сигнал, що помилка ще не є фінальним вираженням, а отже зберігається простір для самокорекції, когнітивного відновлення і підтримки мотивації.

Особливо важливо, що мотивація в такій моделі залежить не лише від однієї відповіді, а від накопиченої історії освітніх подій. Якщо учень систематично отримує низькі результати на серії завдань, це породжує кумулятивний негативний ефект: кожна нова невдача знижує очікування успіху в наступному кроці, а зменшення очікування успіху, своєю чергою, знижує готовність інвестувати зусилля. Такий механізм можна описати як motivational snowball effect.

Нехай $M_i(t) \in [0, 1]$ – латентний рівень навчальної мотивації учня i у момент часу t . Тоді в найпростішому

Table 5: Ймовірнісна модель мотиваційних станів у цифровому двійнику

Освітня ситуація	$P(\text{Low})$	$P(\text{Medium})$	$P(\text{High})$
Перша відповідь хибна, система одразу фіксує невдачу	0.70	0.25	0.05
Перша відповідь хибна, але система ставить уточнювальне питання	0.40	0.45	0.15
Учень частково відновлює правильне reasoning після 1–2 уточнень	0.20	0.55	0.25
Учень проходить усю серію уточнювальних питань і не припиняє роботу	0.10	0.35	0.55
Учень сам знаходить і виправляє помилку без прямої підказки	0.05	0.20	0.75
Серія низьких оцінок на кількох завданнях поспіль	0.78	0.18	0.04
Серія низьких оцінок переривається частковим успіхом через adaptive support	0.25	0.50	0.25
Висока точність і стійкість у складних multimodal tasks	0.03	0.17	0.80

вигляді його динаміка може бути подана як

$$M_i(t+1) = \rho M_i(t) + \beta_1 S_i(t) + \beta_2 C_i(t) - \beta_3 F_i(t) - \beta_4 L_i(t), \quad (19)$$

де:

- $S_i(t)$ – сигнал успіху або часткового успіху;
- $C_i(t)$ – сигнал самокорекції або успішної відповіді на уточнювальне питання;
- $F_i(t)$ – сигнал первинної невдачі;
- $L_i(t)$ – накопичений ефект серії низьких результатів;
- $\rho \in (0, 1)$ – інерційний коефіцієнт мотивації.

У цій моделі негативна первинна відповідь зменшує мотивацію через $F_i(t)$, тоді як уточнювальне питання з можливістю часткового відновлення переводить систему в режим adaptive support і може підвищити $M_i(t)$ через компонент $C_i(t)$. Якщо ж серія завдань завершується низькими результатами, накопичений член $L_i(t)$ посилює спад мотивації, утворюючи snowball effect.

Табл. 5 показує, що мотивація в цифровому двійнику трактується як динамічний ймовірнісний стан. Якщо після першої хибної відповіді система одразу фіксує невдачу, різко зростає ймовірність стану Low: це означає втрату навчальної впевненості, зниження очікування успіху і ризик раннього виходу із задачі. Якщо ж після тієї самої помилки система ставить уточнювальне питання, відбувається зсув розподілу від Low до Medium: учень ще не досяг правильного результату, але вже не сприймає помилку як остаточний провал. Коли після уточнень з'являється часткове або повне відновлення правильного reasoning, розподіл зміщується далі – від Medium до High, тобто система фіксує не

лише правильність, а й наполегливість, самокорекцію та збереження когнітивної залученості.

У термінах переходів станів це можна читати так: Low → Low означає, що негативний досвід закріплюється і кожна нова невдача посилює наступну; це і є початок snowball effect. Перехід Low → Medium означає, що адаптивне уточнення зупиняє різке падіння мотивації і переводить учня в режим часткового відновлення. Перехід Medium → High означає, що після серії уточнювальних кроків учень не лише не втрачає інтерес, а починає демонструвати стійкість до помилки, внутрішню наполегливість і здатність самостійно вийти на правильний хід думки.

Принциповою перевагою такого підходу є те, що мотивація більше не виступає зовнішнім поясненням результату, а стає внутрішнім параметром learner model. Це особливо важливо в ситуаціях, коли двоє учнів мають однакову початкову помилку, але один одразу припиняє спроби, тоді як інший проходить через 3–5 уточнювальних кроків і відновлює правильне reasoning. У традиційному тесті обидва випадки можуть дати однаково низький бал, тоді як цифровий двійник розрізняє їх як різні когнітивно-мотиваційні профілі. Саме це логічно вимагає введення нелінійного калібрування: якщо learner model враховує траєкторію відновлення після помилки, кількість уточнювальних кроків, самокорекцію та динаміку мотивації, то підсумкова 12-бальна шкала не може залишатися простою лінійною пропорцією.

8. Обговорення

Отримані результати дають підстави інтерпретувати запропоновану модель не просто як ще один варіант автоматизованого оцінювання, а як зміну самої логіки оцінювального процесу в цифровому освітньому середовищі. Подані розрахунки, описові статистики та побудовані на основі файлу Book2026.csv візуалізації показують, що learner state у межах MR11A доцільно описувати не лише через correctness of final response, а через інтегральний профіль когнітивної резильентності, який охоплює візуальний, текстовий та математично-змішаний компоненти.

Насамперед результати підтверджують продуктивність переходу від traditional correctness-based assessment до resilience-based assessment. У цій логіці освітньо значущим є не лише те, чи дав учень правильну відповідь, а й те, наскільки стабільним було його reasoning, чи зберігав він когнітивну стійкість під впливом misleading cues, чи був здатний до самокорекції та чи проходив через уточнювальні кроки без втрати смислової цілісності. Відповідно, помилка більше не інтерпретується автоматично як еквівалент “незнання”, а розглядається як потенційний прояв нестійкого, частково сформованого або misleading-induced міркування.

Дані також підтримують тезу про те, що інтегральний показник не повинен бути арифметично грубим агрегатом. Структура трьох субіндексів і спосіб побудови НММІ показують, що низька стійкість хоча б в одній модальності може суттєво знижувати загальний resilience-profile учня навіть за відносно високих двох інших компонентів. Це узгоджується з обраною геометрично чутливою моделлю агрегації та додатково підтверджує гіпотезу **H1**: інтегральний індекс когнітивної резильентності є змістовно сильнішою моделлю learner state, ніж оцінка, побудована лише на фінальному correctness outcome.

Окремо слід підкреслити значення adaptive questioning. Концептуально й аналітично ця частина моделі підтримує гіпотезу **H2**. Якщо учень після первинної помилки отримує 3–5 уточнювальних кроків і демонструє часткове або повне відновлення правильного reasoning, то система фіксує не лише кінцевий бал, а сам механізм когнітивного відновлення. Саме це відрізняє цифровий двійник від жорсткої тестової моделі, де початкова помилка часто непропорційно знижує результат. У такому сенсі MR11A працює не як автоматичний інспектор відповідей, а як система реконструкції латентного стану учня через послідовність відповідей, уточнень і корекцій.

Подані pre-post візуалізації дають підстави обережно підтримати і гіпотезу **H4**. Зсув інтегрального індексу між НММІ_index_pre та НММІ_index_post, а також позитивна динаміка на відповідних графіках можуть бути інтерпретовані як свідчення того, що adaptive digital-twin support асоціюється з покращенням resilience-profile. Водночас у цій статті ми принципово не абсолютизуємо причинність: подані результати мають статус концептуально сильних ілюстрацій і попередньої емпіричної підтримки, але для строгого підтвердження причинного ефекту потрібні окремі inferential designs з чітко визначеними контрольними умовами, статистичним тестуванням та перевіркою ефектів на незалежних вибірках.

Надзвичайно важливою є і вікова динаміка ефекту. Побудовані графіки дозволяють сформулювати змістовну інтерпретацію на користь гіпотези **H5**: різні вікові групи можуть мати різну домінуючу вразливість до misleadingness. Молодші учні, імовірно, більш чутливі до візуальної misleadingness, оскільки їхнє первинне судження частіше залежить від безпосередньої перцептивної сентентності. Старші учні, навпаки, можуть виявляти вищу вразливість до текстово-аналітичної та математично-логічної misleadingness: прихованих логічних зсувів, псевдокоректних доведень, переважаних формулювань, конфліктів між формулою та інструкцією. Отже, цифровий двійник є перспективним не лише як універсальний оцінювальний механізм, а і як віково-чутлива модель когнітивної підтримки.

Окремого значення набуває нелінійне spline-калібрування до 12-бальної шкали. Воно безпосередньо

пов'язане з гіпотезою **H3**. Якщо внутрішній стан учня є неперервним і включає часткове розуміння, самокорекцію, коливання стійкості та адаптивно відновлене reasoning, то зовнішня оцінка не може адекватно відображатися через жорстке лінійне або сходишкове rule-based grading. Shape-preserving spline-модель у цьому контексті є не просто технічним інструментом візуалізації, а математично обґрунтованим механізмом переходу від latent score до педагогічно інтерпретованого результату. Саме вона дозволяє зберегти порядок і неперервність learner state, водночас роблячи результат придатним для шкільної практики.

Особливо перспективним є поєднання цієї моделі з аналізом рукописного когнітивного сліду. У такому випадку цифровий двійник може оцінювати не лише те, що учень написав, але і як саме він дійшов до відповіді: чи був reasoning послідовним, чи з'являлися спроби самоперевірки, чи містило доведення приховані логічні розриви, чи виникала нестійкість у нотації, алгебраїчних переходах або формальному синтаксису доведення. Це виводить оцінювання за межі традиційної тестології і наближає його до моделі когнітивної динаміки.

Таким чином, результати статті в цілому підтримують сформульовану мету дослідження, узгоджуються з поставленими завданнями та на концептуально-емпіричному рівні дають попередню підтримку робочим гіпотезам. MRHA постає не як допоміжний чат-інструмент, а як основа для нового покоління adaptive assessment, у якому штучний інтелект працює не з окремою відповіддю, а з динамічною моделлю учня.

9. Етичні та методологічні виклики

Попри високий потенціал, запропонований підхід пов'язаний із низкою етичних і методологічних обмежень, які не можна ігнорувати.

По-перше, оцінювання на основі цифрового двійника працює з чутливими освітніми даними, а в перспективі – також із рукописними артефактами, поведінковими логами, часовими патернами та історією помилок. Тому така система має будуватися на принципах privacy by design, data minimization, контрольованого доступу до даних і прозорого пояснення того, які саме сигнали використовуються для побудови learner model.

По-друге, adaptive questioning і spline-based grading повинні бути пояснюваними для всіх учасників освітнього процесу. Учень, учитель і, за потреби, батьки повинні розуміти, чому система поставила додаткове уточнювальне питання, як було інтерпретовано часткове розуміння, яким чином змінювався інтегральний індекс і як саме відбулося його калібрування до 12-бальної шкали. Без explainability така система ризикує бути сприйнятою як непрозора “чорна скринька”.

По-третє, цифровий двійник не повинен створювати ілюзію абсолютної об'єктивності. Навіть якщо

модель працює з більшою кількістю сигналів, ніж традиційний тест, вона все одно спирається на певні припущення, архітектурні рішення, вагові коефіцієнти й інтерпретаційні правила. Отже, остаточне педагогічне судження має залишатися за людиною, а не повністю делегуватися алгоритму.

Методологічно важливо також наголосити, що наведені в статті емпіричні візуалізації мають характер концептуально значущої ілюстрації та попередньої аналітичної підтримки, а не завершеної статистично вичерпної валідації. Подальші дослідження мають перевірити запропоновану модель на ширших вибірках, у різних предметних галузях, на різних типах misleading tasks, а також із явним розкладом ефекту на окремі субіндекси V , T і M . Окремого тестування потребують також вікові взаємодії, стабільність spline-калібрування, валідність adaptive follow-up questioning та надійність інтерпретації рукописного когнітивного сліду.

10. Висновки

Проведене дослідження забезпечує як теоретичне, так і емпіричне підґрунтя для концептуалізації MRHA як цифрового двійника учня, побудованого не на бінарній логіці правильності відповіді, а на моделі resilience-based assessment. На вибірці з $n = 280$ учнів мультимодальна структура когнітивної резильєнтності виявилася аналітично змістовною: середні pretest-значення ($\bar{V} = 0.6113$, $\bar{T} = 0.5469$, $\bar{M} = 0.5230$) продемонстрували диференційовану модально-специфічну вразливість, тоді як інтегральний геометричний індекс становив $\bar{H}_{pre} = 0.5528$. Це підтверджує гіпотезу **H1**: стан учня адекватніше репрезентується як конфігурація субіндексів, а не як одиничний показник правильності.

Адаптивна підтримка цифрового двійника була пов'язана з істотним зростанням інтегрального показника до $\bar{H}_{post} = 0.7732$, що відповідає середньому приросту $\Delta\bar{H} = 0.2204$. Вплив instructional condition на приріст виявився статистично дуже сильним ($F(2, 277) = 97.66$, $p < 2 \times 10^{-16}$), тоді як початкові міжгрупові відмінності не були статистично значущими ($p = 0.358$). Ці результати підтримують гіпотезу **H4** і свідчать, що adaptive follow-up questioning та resilience-oriented modeling систематично пов'язані з позитивними змінами learner state, а не з випадковими флуктуаціями.

Нелінійний віковий ефект, виявлений за допомогою GAM ($\chi^2 = 41.82$, $p = 8.43 \times 10^{-6}$; explained deviance = 30.7%), підтверджує гіпотезу **H5**: когнітивна резильєнтність і вразливість до misleadingness є віково-чутливими та структурно немонотонними характеристиками. Це означає, що цифрові двійники повинні враховувати віково-адаптивне калібрування ваг субіндексів і стратегій уточнювального опитування.

Монотонне spline-калібрування латентного індексу $H \in [0, 1]$ до 12-бальної шкали підтверджує гіпотезу **H3**. На відміну від жорсткого порогового оцінювання,

spline-відображення зберігає порядкову структуру, але водночас залишається чутливим до часткового відновлення, самокорекції та динаміки reasoning. У поєднанні з adaptive questioning (гіпотеза H2) такий підхід запобігає непропорційно жорсткому заниженню оцінки після початкової помилки та узгоджує grading не з одиничним outcome, а з когнітивною траєкторією учня.

Загалом MRIIA демонструє можливість інтегрувати learner modeling, adaptive assessment, аналіз мультимодальної misleadingness та explainable spline grading в єдину архітектуру цифрового двійника. Внесок цієї роботи в AI in Education полягає у переході від статичного вимірювання результату до динамічного моделювання когнітивної резильєнтності як латентного стану, що може оновлюватися в процесі взаємодії. Подальші дослідження мають бути спрямовані на валідацію причинних механізмів на ширших і міжкультурних вибірках, перевірку стабільності spline-калібрування та подальшу операціоналізацію мотиваційних переходів у межах ймовірнісних learner models.

Declaration of competing interest

Автор заявляє про відсутність конфлікту інтересів.

Data availability

Емпірична ілюстрація в роботі ґрунтується на файлі Book2026.csv та похідних візуалізаціях, сформованих автором.

References

- [1] I. Kabashkin, AI-Based Digital Twins of Students: A New Paradigm for Competency-Oriented Learning Transformation, *Information* 16 (10) (2025) 846. <https://doi.org/10.3390/info16100846>
- [2] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, J. Sohl-Dickstein, Deep Knowledge Tracing, in: *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [3] S. Lee, S.-H. Kim, B. C. Kwon, VLAT: Development of a Visualization Literacy Assessment Test, *IEEE Transactions on Visualization and Computer Graphics* 23 (1) (2017) 551–560. <https://doi.org/10.1109/TVCG.2016.2598920>
- [4] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, E. Bertini, How Deceptive Are Deceptive Visualizations? An Empirical Analysis of Common Distortion Techniques, in: *Proceedings of IEEE VIS 2015*, 2015, pp. 1469–1478.
- [5] M. Correll, E. Bertini, S. Franconeri, Truncating the Y-Axis: Threat or Menace?, in: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.
- [6] L. Y.-H. Lo, A. Gupta, K. Shigyo, A. Wu, E. Bertini, H. Qu, Misinformed by Visualization: What Do We Learn from Misinformative Visualizations?, *Computer Graphics Forum* 41 (3) (2022) 515–525. <https://doi.org/10.1111/cgf.14510>
- [7] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, J. Hullman, The Science of Visual Data Communication: What Works, *Psychological Science in the Public Interest* 22 (3) (2021) 110–161. <https://doi.org/10.1177/15291006211051956>
- [8] S. Shen et al., A Survey of Knowledge Tracing: Models, Variants, and Applications, *IEEE Transactions on Learning Technologies* (2024).
- [9] Y. Bai, J. Zhao, T. Wei, Q. Cai, L. He, A Survey of Explainable Knowledge Tracing, *arXiv preprint arXiv:2403.07279* (2024).
- [10] Y. Zhang et al., Knowledge ontology enhanced model for explainable knowledge tracing, *Journal of King Saud University – Computer and Information Sciences* 36 (10) (2024) 102065. <https://doi.org/10.1016/j.jksuci.2024.102065>
- [11] J. Rho, M. A. Rau, X. Zhu, Comparing the deceptive impact of misleading data visualizations: Implications for adaptive data literacy support in computer-based learning environments, *Computers & Education* 242 (2026) 105508. <https://doi.org/10.1016/j.compedu.2025.105508>
- [12] R. J. Adams, *Adaptive Testing, Information and the Partial Credit Model: A Simulation Study*, Centre for the Study of Higher Education, The University of Melbourne, 1987.
- [13] G. N. Masters, A Rasch Model for Partial Credit Scoring, *Psychometrika* 47 (2) (1982) 149–174. <https://doi.org/10.1007/BF02296272>
- [14] D. J. Weiss, Adaptive Testing, in: *The SAGE Encyclopedia of Educational Technology*, SAGE, 2015. <https://doi.org/10.4135/9781483346397.n13>
- [15] A. Khosla, H. Singh, A. Aggarwal, J. Bedi, A knowledge tracing based intelligent framework for formative assessment of students skills, *Complex & Intelligent Systems* 12 (2026) 22. <https://doi.org/10.1007/s40747-025-02149-4>
- [16] A. K. Mishra, C. R. Dudhatra, R. K. Mishra, R. P. Mishra, IoT and digital twins enabled smart learning in education, in: *Digital Twins for Sustainable Development*, Elsevier, 2026, pp. 183–210. <https://doi.org/10.1016/B978-0-443-27388-9.00006-4>

- [17] H. Omrany, K. M. Al-Obaidi, A. Ghaffarianhoseini, R.-D. Chang, C. Park, F. Rahimian, Digital twin technology for education, training and learning in construction industry: implications for research and practice, *Engineering, Construction and Architectural Management* 33 (3) (2024) 1836–1870. <https://doi.org/10.1108/ECAM-10-2024-1376>
- [18] N. Alruwais, M. Zakariah, Evaluating Student Knowledge Assessment Using Machine Learning Techniques, *Sustainability* 15 (7) (2023) 6229. <https://doi.org/10.3390/su15076229>
- [19] S. Wang, H. Wang, Y. Lu, J. Huang, Toward a Comprehensive Evaluation of Student Knowledge Assessment for Art Education: A Hybrid Approach by Data Mining and Machine Learning, *Applied Sciences* 14 (12) (2024) 5020. <https://doi.org/10.3390/app14125020>
- [20] A. K. Mishra, C. R. Dudhatra, R. K. Mishra, R. P. Mishra, IoT and digital twins enabled smart learning in education, in: *Digital Twins for Sustainable Development*, Elsevier, 2026, Chapter 8, pp. 183–210. <https://doi.org/10.1016/B978-0-443-27388-9.00006-4>
- [21] H. Omrany, K. M. Al-Obaidi, A. Ghaffarianhoseini, R.-D. Chang, C. Park, F. Rahimian, Digital twin technology for education, training and learning in construction industry: implications for research and practice, *Engineering, Construction and Architectural Management* (2024). <https://doi.org/10.1108/ECAM-10-2024-1376>
- [22] M. Lu, Z. Hu, Digital Twin-Enhanced Programming Education: An Empirical Study on Learning Engagement and Skill Acquisition, *Computers* 14 (8) (2025) 322. <https://doi.org/10.3390/computers14080322>
- [23] P. M. Doraiswamy, J. A. Duñabeitia, C. Rodriguez, D. P. Devanand, Digital cognitive twins in mental health, *Nature Mental Health* 3 (10) (2025) 1106–1108. <https://doi.org/10.1038/s44220-025-00482-8>
- [24] M. Spitzer, I. Dattner, S. Zilcha-Mano, Digital twins and the future of precision mental health, *Frontiers in Psychiatry* 14 (2023) 1082598. <https://doi.org/10.3389/fpsy.2023.1082598>
- [25] M. Nadeem, S. Kostic, M. Dornhöfer, C. Weber, M. Fathi, A comprehensive review of digital twin in healthcare in the scope of simulative health-monitoring, *Digital Health* 11 (2025). <https://doi.org/10.1177/20552076241304078>
- [26] K. P. Venkatesh, G. Brito, M. N. Kamel Boulos, Health digital twins in life science and health care innovation, *Annual Review of Pharmacology and Toxicology* 64 (2024) 159–170. <https://doi.org/10.1146/annurev-pharmtox-022123-022046>
- [27] M. Ringeval et al., Advancing health care with digital twins: Meta-review of applications and implementation challenges, *Journal of Medical Internet Research* 27 (2025) e69544. <https://doi.org/10.2196/69544>
- [28] N. Alruwais, M. Zakariah, Evaluating Student Knowledge Assessment Using Machine Learning Techniques, *Sustainability* 15 (7) (2023) 6229. <https://doi.org/10.3390/su15076229>
- [29] S. Wang, H. Wang, Y. Lu, J. Huang, Toward a Comprehensive Evaluation of Student Knowledge Assessment for Art Education: A Hybrid Approach by Data Mining and Machine Learning, *Applied Sciences* 14 (12) (2024) 5020. <https://doi.org/10.3390/app14125020>
- [30] M. Zafari, A. Sadeghi-Niaraki, S.-M. Choi, A. Esmaeily, A Practical Model for the Evaluation of High School Student Performance Based on Machine Learning, *Applied Sciences* 11 (23) (2021) 11534. <https://doi.org/10.3390/app112311534>
- [31] P. J. Weintraub et al., NIH Toolbox Cognition Battery (CB): Introduction and pediatric data, *Monographs of the Society for Research in Child Development* 78 (4) (2013/2014). <https://pmc.ncbi.nlm.nih.gov/articles/PMC3954750/>
- [32] B. K. Taylor et al., Reliability of the NIH Toolbox cognitive battery in children and adolescents: a 3-year longitudinal examination, *Psychological Medicine* 52 (9) (2022) 1718–1727. <https://doi.org/10.1017/S0033291720003487>
- [33] S. Souissi, K. Chamari, T. Bellaj, Assessment of executive functions in school-aged children: A narrative review, *Frontiers in Psychology* 13 (2022) 991699. <https://doi.org/10.3389/fpsyg.2022.991699>
- [34] S. B. Dallaire-Demers et al., Computerized Cognitive Test Batteries for Children and Adolescents—A Scoping Review of Tools for Lab- and Web-Based Settings From 2000 to 2021, *Archives of Clinical Neuropsychology* / review source (2023). <https://pmc.ncbi.nlm.nih.gov/articles/PMC10681451/>
- [35] K. Cabbage, S. Brinkley, S. Gray, M. Alt, N. Cowan, S. Green, T. Kuo, T. P. Hogan, Assessing Working Memory in Children: The Comprehensive Assessment Battery for Children – Working Memory (CABC-WM), *Journal of Visualized Experiments* 124 (2017) e55121. <https://doi.org/10.3791/55121>
- [36] K. Thianwan, S. Srikoon, Development of an AI literacy self-assessment questionnaire in upper primary school students, *Social Sciences & Humanities Open* 13 (2026) 102424. <https://doi.org/10.1016/j.ssaoh.2025.102424>

- [37] J. Schult, B. Fauth, R. Schneider, M. A. Lindner, How schools rebound from pandemic learning loss: Longitudinal findings from mandatory large-scale assessments, *Learning and Instruction* 101 (2026) 102263. <https://doi.org/10.1016/j.learninstruc.2025.102263>
- [38] M. Iqbal, K. Singh, S. Khan, O. Osho, E. Sidnam-Mauch, N. Bannister, K. Caine, B. Knijnenburg, Teaching AI Awareness in Middle School Classrooms: Design, Implementation and Evaluation of Two Education Modules on Algorithmic Bias and Filter Bubbles, *Computers and Education: Artificial Intelligence* 8 (2025) 100425. <https://doi.org/10.1016/j.caeai.2025.100425>
- [39] Mriia International School, Artificial Intelligence and Digital Twins in Education: Project Description and Experimental Data Infrastructure, institutional web resource, 2026. <https://mriia-school.com/shtuchnyj-intelekt-digital-twins/>
- [40] Study in Europe, Compare European Grades, comparative overview of university grading systems in Europe, 2026. <https://www.studyineurope.eu/grades/>