



ПЕДАГОГІЧНА ОСВІТА

УДК 371.263:519.23

DOI <https://doi.org/10.5281/zenodo.20029707>

Оптимальний вибір моделі MIRT в аналізі тестів з вищої математики

Диховичний Олександр Олександрович

доцент, кандидат фізико-математичних наук,

доцент кафедри математичного аналізу та теорії ймовірностей

Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, a.dyx@ukr.net,

ORCID ID [0000-0001-8505-2878](https://orcid.org/0000-0001-8505-2878)

Круглова Наталія Володимирівна

кандидат фізико-математичних наук,

доцент кафедри математичного аналізу та теорії ймовірностей

Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, natahak@ukr.net,

ORCID ID [0000-0002-9143-9398](https://orcid.org/0000-0002-9143-9398)

Москвичова Катерина Костянтинівна

кандидат фізико-математичних наук,

старший викладач кафедри математичного аналізу та теорії ймовірностей

Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна,

moskvychovakateryna@gmail.com, ORCID ID [0000-0002-1870-9873](https://orcid.org/0000-0002-1870-9873)

Пелехата Ольга Богданівна

кандидат фізико-математичних наук,

доцент кафедри математичного аналізу та теорії ймовірностей

Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, pelekhata.olha@lil.kpi.ua,

ORCID ID [0000-0003-1318-856X](https://orcid.org/0000-0003-1318-856X)



Прийнято: 05.04.2026 | Опубліковано: 30.04.2026

Анотація

Мета. Метою даного дослідження є визначення найбільш оптимальної MIRT-моделі для аналізу якості тестів з вищої математики. Акцент зроблено на порівнянні політомічних моделей без їх розкладання на дихотомічні компоненти.

Методи. Емпірична база дослідження - результати контрольної роботи з математичного аналізу для 120 студентів РТФ «КПІ ім. І. Сікорського», яка включала 20 дихотомічних та політомічних завдань з максимальною кількістю рівнів 12. Дослідження проводилося у два етапи. На етапі експлораторного факторного аналізу (EFA) було застосовано низку алгоритмів (CD, ЕМПКС, HULL, MAP, NEVALSGT1, RAWPAR, SESCREE, SMT) для попереднього визначення розмірності. На етапі конфірматорного аналізу (CFA) оцінювалися параметри трьох типів MIRT-моделей: GPCM, GRM та NRM у одно-, дво- та тривимірних реалізаціях. Адекватність моделей порівнювалася за критеріями AIC, BIC, RMSEA, CFI та TLI.

Результати. Аналіз методів EFA виявив розбіжності у рекомендованій розмірності: шість методів (CD, HULL, MAP, RAWPAR, SESCREE, SMT) вказали на двовимірність, два методи (ЕМПКС, NEVALSGT1,) - на п'ятивимірність, що здається надлишковим. Конфірматорний аналіз дев'яти моделей показав, що найкращою відповідно до інформаційних критеріїв є двовимірна модель GPCM2. Проведено аналіз матриці факторних навантажень, перевірку завдань на монотонність і локальну незалежність. Виділено «невдалі» завдання, які потребують заміни. Дослідження показало, що одновимірні моделі є недостатніми для опису структури математичних знань, тоді як моделі вищої вимірності демонструють надмірність параметрів без суттєвого покращення якості підгонки.



Висновки. Дослідження встановило коректність багатоступінчатого підходу до вибору MIRT-моделі для математичних тестів, оскільки результати EFA послідовно узгоджуються з оптимальною двовимірною структурою, підтвердженою CFA. Модель GPCM2 виявилася оптимальною для аналізу політомічних математичних завдань. По результатах дослідження виявлено дві відносно незалежні компоненти математичної компетентності, які можна інтерпретувати як дві латентні характеристики, що відповідають природній структурі математичного аналізу: компетентність у диференціальному численні та компетентність в інтегральному численні.

Ключові слова: IRT, MIRT, EFA, CFA, GPCM, GRM, NRM, вища математика, якість тесту, латентні змінні.

Choice of MIRT Model for Analysis of Higher Mathematics Tests

Oleksandr O. Dykhovychnyi

PhD (physical and mathematical sciences),

Associate professor of Mathematical analysis and Probability Theory Department
National Technical University of Ukraine “Igor Sikorsky Kiev Polytechnic Institute”,

Kyiv, Ukraine, a.dyx@ukr.net,

<https://orcid.org/0000-0001-8505-2878>

Nataliia V. Kruglova

PhD (physical and mathematical sciences),

Associate professor of Mathematical analysis and Probability Theory Department
National Technical University of Ukraine “Igor Sikorsky Kiev Polytechnic Institute”,

Kyiv, Ukraine, natahak@ukr.net,

<https://orcid.org/0000-0002-9143-9398>

Kateryna K. Moskvychova

PhD (physical and mathematical sciences),

Senior teacher of Mathematical analysis and Probability Theory Department



National Technical University of Ukraine “Igor Sikorsky Kiev Polytechnic
Institute”, Kyiv, Ukraine,

moskvychovakateryna@gmail.com, <https://orcid.org/0000-0002-1870-9873>

Olha B. Pelekhata

PhD (physical and mathematical sciences),

Associate professor of Mathematical analysis and Probability Theory Department

National Technical University of Ukraine “Igor Sikorsky Kiev Polytechnic

Institute”, Kyiv, Ukraine, pelekhata.olha@lil.kpi.ua,

<https://orcid.org/0000-0003-1318-856X>

Abstract

Purpose. *This study aims to identify the most appropriate MIRT model for evaluating higher mathematics tests, with a focus on polytomous models without decomposition into dichotomous components.*

Methods. *The empirical data were obtained from a calculus test administered to 120 students at Igor Sikorsky Kyiv Polytechnic Institute. The test included 20 dichotomous and polytomous items with up to 12 response categories. The study was conducted in two stages. First, Exploratory Factor Analysis (EFA) was applied using multiple criteria (CD, EMPKC, HULL, MAP, NEVALSGT1, RAWPAR, SESCREE, SMT) to determine the latent dimensionality. Second, Confirmatory Factor Analysis (CFA) was used to estimate and compare nine MIRT models (GPCM, GRM, NRM) in one-, two-, and three-dimensional specifications. Model fit was evaluated using AIC, BIC, RMSEA, CFI, and TLI indices.*

Results. *The EFA results showed partial inconsistency across methods: most criteria supported a two-dimensional structure, whereas some suggested higher dimensionality. The CFA results indicated that the two-dimensional GPCM model (GPCM2) provided the best fit according to the information criteria. Analysis of factor loadings, as well as checks for monotonicity and local independence, confirmed the*



overall adequacy of the model while identifying several misfitting items. The findings demonstrate that unidimensional models are insufficient to capture the structure of mathematical knowledge, whereas higher-dimensional models introduce parameter redundancy without substantial improvement in model fit.

Conclusions. *The results support the effectiveness of a multi-stage approach to MIRT model selection. The two-dimensional GPCM model is the most appropriate for analyzing polytomous mathematics test items. Two relatively independent latent dimensions were identified, corresponding to key components of calculus competence: differential and integral calculus.*

Keywords: *IRT, MIRT, EFA, CFA, GPCM, GRM, NRM, higher mathematics, test quality, latent variables.*

Постановка проблеми. В умовах дистанційного та змішаного навчання комп'ютерне тестування залишається необхідним інструментом оцінювання знань іспитників. Від якості тестових завдань залежить правильність оцінювання, діагностика прогалин та планування коригувальних заходів. В Україні традиційно аналіз якості тестів проводять методами Класичної теорії тестів (КТТ) [1] та Item Response Theory (IRT)[2].

Однак класичні одновимірні моделі IRT передбачають, що успішність виконання всіх завдань тесту характеризується єдиною компетентністю. Це припущення часто виявляється нереалістичним для математичних дисциплін, де поєднуються розуміння теоретичних концепцій, технічні обчислювальні навички, вміння читати формальну символіку, здатність застосовувати знання до розв'язування задач. Використання одновимірних моделей призводить до втрати важливої інформації та зниження точності оцінювання індивідуальних здібностей студентів.

Багатовимірна Теорія Тестів (MIRT) [3] є більш точним математичним інструментом для аналізу складних латентних структур. Проте практичне

застосування MIRT вимагає визначення оптимальної кількості вимірів латентних параметрів і підбору моделі з найменшою похибкою.

Аналіз останніх досліджень і публікацій. Основи MIRT закладено в роботі [3]. В останні роки спостерігається інтенсивний розвиток методів та моделей MIRT.

Li, Gibbons та Ročková [4] розробили новий байєсівський підхід до MIRT, який автоматично ідентифікує розріджену структуру латентних рис без попереднього визначення кількості вимірів. Їхній метод особливо цінний для великих наборів даних і мінімізує необхідність у суб'єктивному налаштуванні параметрів. Cho, Wang, Zhang та Xu [5] запропонували гаусівську варіаційну оцінку для параметрів MIRT як швидку альтернативу традиційним алгоритмам максимальної правдоподібності, яку згодом було удосконалено Ma, Ouyang, Wang та Xu [6] через зважену вибірку для корекції систематичних похибок.

Критичним аспектом застосування MIRT є правильне визначення розмірності латентного простору. Markos та Tsigilis [7] провели комплексне симуляційне порівняння паралельного аналізу та експлораторного () графічного аналізу для порядкових даних, виявивши, що ефективність методів суттєво залежить від характеристик даних. Xia та Zhou [8] удосконалили традиційний паралельний аналіз, врахувавши варіабельність вибірки кореляційної матриці, що особливо важливо для малих вибірок.

Питання вибору між різними специфікаціями MIRT-моделей досліджено в роботі Cole та Paek [9], які порівняли програмні реалізації в SAS PROC IRT та R-пакеті mirt. Hosseinzadeh та Cole [10] виявили, що неправильна специфікація структури перехресних навантажень призводить до систематичних похибок у параметрах. Jewsbury та van Rijn [11] математично довели, що використання окремих одновимірних моделей для багатоетапних тестів є некоректним, тоді як MIRT-моделі забезпечують валідні оцінки.

Для освітнього контексту особливо важливою є робота Askerman [12], який проаналізував причини обмеженого практичного впровадження MIRT, попри його суттєві теоретичні переваги. Автор підкреслює необхідність узгодження між форматом тесту, передбаченими компетентностями та статистичною моделлю. Kim, Lee [13] продемонстрували важливість діагностики розмірності латентного вимірів перед застосуванням MIRT аналізу.

Щодо інтерпретації латентних вимірів у контексті математичного оцінювання, класичні роботи Hartig та Höhler [14] виокремлюють три сценарії використання MIRT для оцінювання компетентностей. Jiao та співавтори [15] застосували MIRT до великомасштабного наукового оцінювання з множинними контентними областями, обґрунтовуючи, що різні предметні області можуть відповідати окремим латентним вимірам.

В українському контексті дослідження MIRT представлені обмежено. Круглова, Диховичний та Лисенко [16] застосували MIRT до аналізу тесту з аналітичної геометрії, порівнявши одновимірні моделі Муракі та Бірнбаума з багатовимірними 2-PL і GPCM. Круглова та Диховичний [17] досліджували вибір MIRT-моделі для психологічних тестів.

Шимчик, Круглова та Диховичний [18] обмежено розглядали питання вибору розмірності MIRT-моделей для тестів з вищої математики.

Аналіз літератури показує, що переважна більшість досліджень зосереджена або на розробці алгоритмів оцінювання параметрів, або на порівнянні окремих моделей у межах конкретних предметних областей. Однак бракує комплексних досліджень, які б систематично порівнювали широкий спектр методів визначення розмірності (більше ніж 2-3 методи) та різні типи політомічних моделей (GPCM, GRM, NRM) на одних і тих самих даних з чіткою педагогічною інтерпретацією результатів.

Виділення невирішених раніше частин загальної проблеми.



Незважаючи на значний обсяг досліджень з MIRT, залишаються невирішеними наступні методологічні питання аналізу математичних тестів:

1. В більшості досліджень обмежуються застосуванням 1-3 методів. Відсутні роботи, які б комплексно порівнювали результати застосування CD, EMPKS, HULL, MAP, NEVALSGT1, RAWPAR, SESCREE, SMT на реальних освітніх даних з вищої математики та дослідили їх узгодженість.
2. Політомічні завдання часто розкладаються на набір дихотомічних для спрощення аналізу, або порівнюються лише два типи моделей (наприклад, GPCM і GRM).
3. Більшість психометричних досліджень зосереджені на статистичних показниках якості моделі, приділяючи недостатню увагу педагогічній інтерпретації виявлених латентних параметрів. Для математичного оцінювання необхідне розуміння, чи відповідають виміри дисциплінам (лінійна алгебра, аналітична геометрія, математичний аналіз) або когнітивним процесам (обчислення, моделювання, доведення).
4. Відсутні структуровані рекомендації щодо послідовності вибору оптимальної MIRT-моделі, які могли б використовувати розробники тестів без глибокої психометричної підготовки.

Заповнення цих прогалин має як теоретичне, так і практичне значення для удосконалення методів освітнього оцінювання з вищої математики та підвищення валідності педагогічних висновків.

Формулювання цілей статті (постановка завдання).

Метою даного дослідження є розробка та апробація комплексної методології вибору оптимальної багатовимірної моделі MIRT для аналізу якості математичних тестів.

Виклад основного матеріалу дослідження.

Учасники та контекст дослідження.



Дослідження проводилися за результатами контрольної роботи з математичного аналізу, проведеної у Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» у 2025 році. Загальна вибірка включала 120 студентів першого курсу РТФ. Контрольна робота проводилась в тестовому форматі на платформі MOODLE після завершення відповідних розділів курсу математичного аналізу.

Результати контрольної роботи наведені у додатку А.

Інструмент дослідження. Тест містив 20 завдань на теми: похідні функцій, дослідження функцій за допомогою похідних, невизначений та визначений інтеграл, застосування інтегралів. Тест містив як дихотомічні, так і політомічні питання з кількістю рівнів від 3 до 12. Політомічні завдання у формі «вбудованих відповідей» дозволяли оцінити покрокове виконання задачі та запобігають застосуванню різноманітних «помічників».

Приклад приклад політомічного завдання з дослідження функції з 12-ма рівнями наведено на рисунку 1.

Етап 1: Експлораторний факторний аналіз (EFA). На першому етапі було застосовано вісім різних алгоритмів для попереднього оцінювання розмірності моделі:

1. Comparison Data (CD) – метод, що порівнює власні значення реальних даних з власними значеннями випадкових даних подібної структури;
2. Empirical Kaiser Criterion (EMPKC) – емпірична версія критерію Кайзера;
3. Hull Method (HULL) – метод, що балансує якість підгонки та складність моделі через оптимізацію показника критерію;
4. Minimum Average Partial (MAP) – метод мінімуму середнього часткового кореляційного залишку;
5. NEVALSGT1 – класичний критерій Кайзера;
6. Raw Parallel Analysis (RAWPAR) – паралельний аналіз на основі необроблених даних;

Рисунок 1

Дослідити функцію: $y = \frac{x^2 - x - 1}{x - 2}$.

1) Область визначення:

☐ $x \in \mathbb{R}$ ☐ $x \neq 2$ ☐ $x \neq 0$ ☐ $x > 0$

2) Функція

☐ загального вигляду
☐ непарна
☐ парна

3) Якщо функція не має певного виду асимптот, то у відповідні комірки внесіть символ "-" (вносити слід без лапок)!

Має вертикальну асимптоту: $x =$.

Має горизонтальну асимптоту: $y =$.

Має похилу асимптоту: $y =$ $x +$.

4) Функція має максимум в точці $x =$ і $y_{max} =$.

Функція має мінімум в точці $x =$ і $y_{min} =$.

5) Функція опукла вгору при $x <$.

Функція опукла вниз при $x >$.

Джерело: власна розробка авторів у середовищі MOODLE

7. Scree Test (SESCREE) – метод другого власного значення для scree-графіку;

8. Sequential Mean Test (SMT) – послідовний тест середніх значень.

Усі методи було реалізовано в середовищі R з використанням пакетів **psych**, **EFAtools** та **EFAfactors** [19]. Метою цього етапу було отримати попередні оцінки розмірності для подальшого конфірматорного аналізу.

Етап 2: Конфірматорний факторний аналіз (CFA).

На етапі конфірматорного факторного аналізу (CFA) для кожної з імовірних трьох розмірностей ($s=1, 2, 3$; високі розмірності було відкинуто) було

оцінено параметри наступних політомічних MIRT-моделей: **GPCM** – модель Muraki [20], **GRM** – модель Samejima [21], **NRM** – модель Bock [22].

Оцінка латентних параметрів усіх моделей проводилася за допомогою алгоритму максимізації математичного сподівання (EM), реалізованому у пакеті **mirt** середовища R.

Для порівняння адекватності вибраних моделей використано п'ять критеріїв [23]: інформаційний критерій Акаїке (AIC), Байєсівський інформаційний критерій (BIC), RMSEA, порівняльний індекс придатності (CFI), індекс Такера-Льюїса (TLI).

Результати та їх обговорення.

Результати застосування восьми алгоритмів EFA представлені в Таблиці 1. Спостерігається суттєва варіативність у рекомендованій розмірності латентного простору.

Таблиця 1.

Результати експлораторного факторного аналізу

Метод	Попередня оцінка розмірності
CD	2
EMPKC	5
HULL	2
MAP	2
NEVALSGT1	5
RAWPAR	2
SESCREE	2
SMT	2

Джерело: власні обчислення авторів

Методи CD, MAP, SMT, HULL, RAWPAR, SESCREE узгоджено вказали на двовимірну структуру. Класичний критерій Кайзера (NEVALSGT1) і EMPKC

дали явно завищену оцінку (5 вимірів), що можна пояснити їх тенденцією до переоцінки розмірності, особливо при наявності слабких факторів або шуму в даних.

Згода більшості методів щодо двовимірності вказує на доцільність розгляду багатовимірних моделей і пріоритетність такої розмірності для конфірматорного аналізу.

Результати конфірматорного аналізу та порівняння MIRT-моделей.

Було оцінено дев'ять моделей: три типи політомічних моделей (GPCM, GRM, NRM) для трьох розмірностей ($s=1, 2, 3$). Результати оцінки за критеріями якості наведено в Таблиці 2.

Таблиця 2.

Порівняння якості MIRT-моделей за критеріями адекватності

	Вимірність	AIC	BIC	M2	RMSEA	CFI	TLI
GPCM	1	2286.332	2419.022	70.19218	0.05825498	0.986349	0.9808888
GRM	1	2285.039	2420.729	46.8472	0.05263132	0.9678345	0.9802984
NRM	1	2283.802	2473.358	23.63699	0.0532811	0.9584391	0.9845721
GPCM	2	2251.812	2443.043	21.88048	0.033101	0.9986891	0.9913922
GRM	2	2287.175	2455.406	22.84791	0.03861495	0.9867854	0.9712445
NRM	2	2261.388	2486.486	20,71856	0.05512665	0.9591264	0.9688748
GPCM	3	2291.253	2492.656	9.524399	0.0602111	0.9768548	0.9912652
GRM	3	2290.479	2491.882	17.06431	0.05125645	0.9972523	0.9852048
NRM	3	2261.851	2510.121	2.3454	0.0362228	0.9345222	0.90444655

Джерело: власні обчислення авторів

Для ранжування моделей за кожним показником застосовано такі правила: **AIC**, **BIC**, **M2**, **RMSEA** – мінімальне значення отримує ранг 1; **CFI**, **TLI** – максимальне – ранг 1. Далі підсумовано ранги і відсортувано за сумою (менший кращий).

Двовимірна модель **GPCM** має найменший сумарний ранг (7) – тобто вона лідирує у більшості критеріїв. Серед одновимірних, найкраща за сумою – одновимірна GPCM (24), однак її AIC/RMSEA значно гірші. Як бачимо, хоча дві одновимірні моделі (GPCM 1D та GRM 1D) мають менші BIC (парсимонія), вони програють за AIC та критеріями підгонки. Водночас перехід до тривимірних моделей не приводить до покращення

Таблиця 3

Ранжування MIRT-моделей за критеріями адекватності

Модель	Вимір	Ранг AIC	Ранг BIC	Ранг RMSEA	Ранг CFI	Ранг TLI	Сумарний ранг
GPCM	2	1	3	1	1	1	7
GPCM	1	6	1	8	4	5	24
GRM	1	5	2	5	6	6	24
GRM	2	7	4	3	3	7	24
GRM	3	8	7	4	2	3	24
NRM	1	4	5	6	8	4	27
NRM	2	2	6	7	7	8	30
NRM	3	3	9	2	9	9	32
GPCM	3	9	8	9	5	2	33

Джерело: власні обчислення авторів

, що свідчить про надлишковість третього латентного виміру.

Згідно з класичними джерелами (Hu & Bentler 1999 та ін.)[24] прийнято вважати хорошою моделлю таку, у якої $RMSEA \leq 0.06$, $CFI \geq 0.95$, $TLI \geq 0.95$. Значення RMSEA для двовимірної GPCM перебуває в діапазоні, що відповідає відмінній якості моделі, тоді як альтернативні специфікації демонструють гірші або лише прийнятні показники. CFI та TLI у двовимірній моделі GPCM наближаються до ідеальних, що свідчить про високу відповідність моделі емпіричним даним. Моделі GRM, також показують прийнятну підгонку, але поступаються за інформаційними критеріями. Номінальна модель NRM виявляється найменш адекватною, що проявляється як у значеннях інформаційних критеріїв, так і в показниках підгонки.

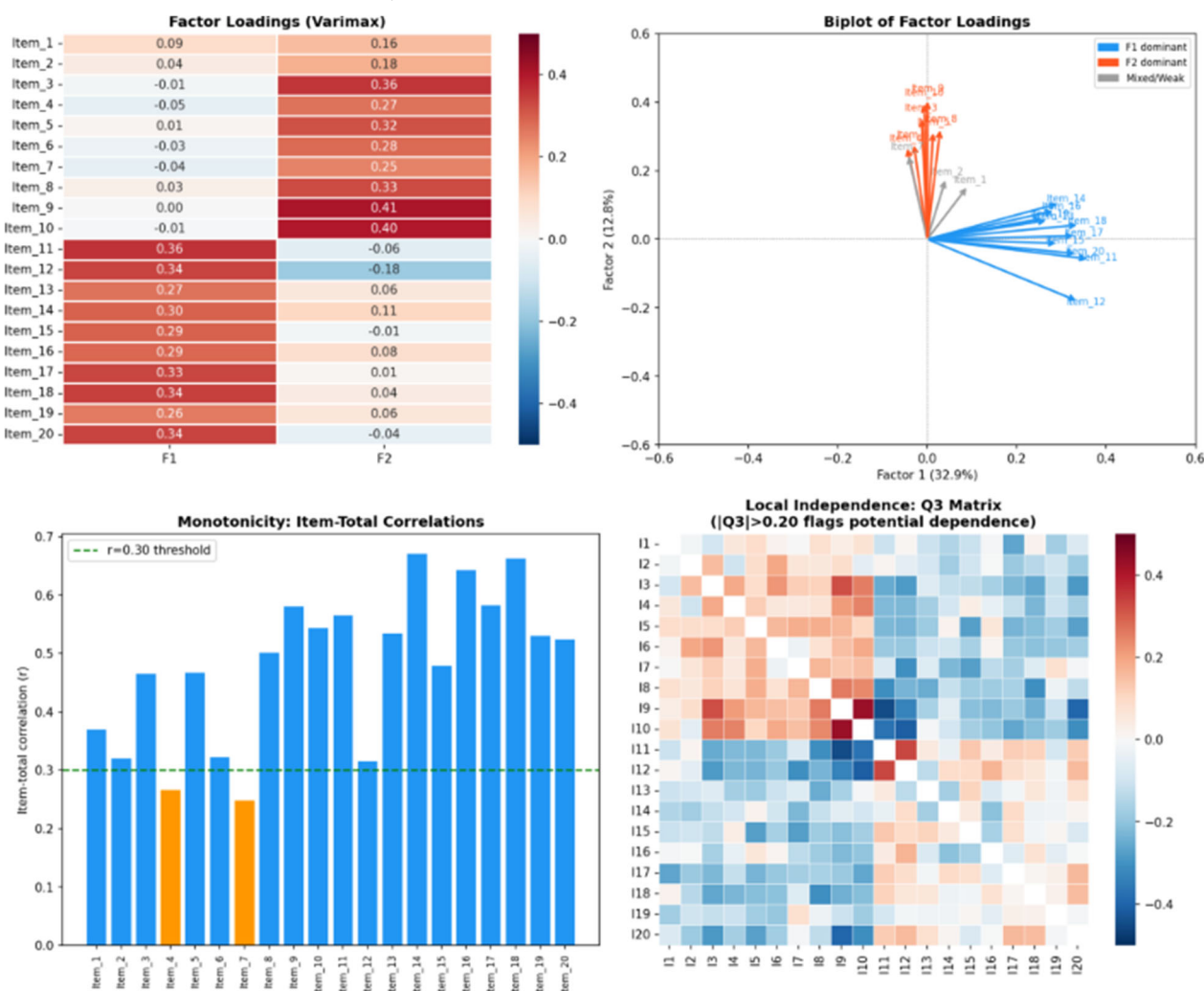
Отже, найкращою для даного тесту можна вважати двовимірну модель GPCM.

Аналіз факторного моделювання свідчить про двовимірну структуру тесту. Матриця факторних навантажень наведена у додатку В. Графічно факторні навантаження та biplot візуалізація відображені на рисунку 2. Очевидно, виокремлюються два виміри: F1 (32.9% дисперсії) – домінують завдання 11-20,

F2 (12.8% дисперсії) – домінують завдання 3-6, 8-10. У завданнях 1, 2, 7 – слабкі навантаження на обидва фактори (такі завдання доцільно виключити).

Рисунок 2

Аналіз результатів тестування на факторне навантаження, монотонність та локальну незалежність.



Джерело: власна розробка авторів

Додатково, аналіз кореляцій «завдання–тест» (показник r_{it}) показав, що більшість завдань мають задовільні або високі значення (переважно понад 0.3), що підтверджує їх монотонність і узгодженість із загальною структурою тесту. Усі 20 завдань продемонстрували статистично значущі позитивні кореляції в діапазоні від $r = 0.25$ (завдання 7) до $r = 0.67$ (завдання 14), що доводить відсутність порушень монотонності даних. Отже, припущення монотонності

виконується для всього банку завдань, що є необхідною умовою коректного застосування MIRT-моделей

Перевірка локальної незалежності здійснювалася на підставі статистики Q3 Єна[25]. Перевірка за допомогою Q3-матриці загалом підтверджує коректність моделі, хоча для групи завдань одного фактора спостерігається помірна залишкова залежність, що може бути зумовлено змістовою близькістю або спільним контекстом.

Інтерпретація вимірів оптимальної моделі. Логічною представляється інтерпретація двовимірної моделі у відповідності до двох вимірів: перший – компетентність у диференціальному численні, другий – у інтегральному. Це підтверджується факторним навантаженням на завдання, які відповідають цим двом вимірам.

Педагогічна цінність такої двовимірної структури полягає в можливості диференційованої діагностики. Викладач може виявити студентів, які успішні в диференціюванні, але мають труднощі з інтегруванням, або навпаки. Це дозволяє надавати цілеспрямовану підтримку саме в тій області, де студент демонструє дефіцит знань.

Висновки.

У роботі розроблено та апробовано комплексну методику вибору оптимальної MIRT-моделі для аналізу математичних тестів, яка базується на систематичному застосуванні методів визначення розмірності та порівнянні різних типів політомічних моделей за комплексним набором критеріїв адекватності.

Результати систематичного порівняння методів експлораторного факторного аналізу виявили відносну узгодженість навколо двовимірної структури латентного простору.



Конфірматорний аналіз дев'яти моделей різних типів та розмірностей підтвердив перевагу двовимірної моделі GPCM. Важливо, що рекомендації EFA (переважно 2 виміри) повністю узгоджуються з результатами CFA.

Порівняння трьох типів політомічних моделей виявило перевагу GPCM над GRM та NRM у двовимірному випадку. Це можна пояснити специфікою математичних завдань з політомічним оцінюванням: бали зазвичай відображають послідовні кроки розв'язання з чітко визначеною порядковою структурою.

Результати дослідження підтверджують недостатність класичних одновимірних моделей IRT для аналізу складних математичних тестів.

Рекомендована послідовність вибору MIRT-моделі включає: застосування розширеного набору методів EFA з фокусом на узгодженості рекомендацій; специфікацію моделей GPCM, GRM та NRM для діапазону розмірностей, визначеного на етапі EFA; порівняння за комплексним набором критеріїв. Для тестів з математичного аналізу рекомендується двовимірна GPCM-модель.

Перспективи подальших досліджень включають застосування обраної двовимірної GPCM-моделі для аналізу завдань контрольних заходів для студентів інших спеціальностей, дослідження валідності оцінок за двома вимірами відносно майбутньої академічної успішності, а також розширення методології на інші розділи математики для перевірки виявленої структури латентних параметрів.

Список використаних джерел

1. Crocker L., Algina J. *Introduction to Classical and Modern Test Theory*. Belmont, CA: Wadsworth, 2006. 527 p.
2. Linden W., Hambleton R. *Handbook of modern item response theory*. New York: Springer, 2018. 1500 p.



3. Reckase M. D. *Multidimensional item response theory*. New York: Springer, 2009. 354 p.

4. Li J., Gibbons R., Ročková V. Sparse Bayesian Multidimensional Item Response Theory. *Journal of the American Statistical Association*. 2025.

DOI: <https://doi.org/10.1080/01621459.2025.2476786>

5. Cho A. E., Wang C., Zhang X., Xu G. Gaussian variational estimation for multidimensional item response theory. *British Journal of Mathematical and Statistical Psychology*. 2021. Vol. 74. P. 52-85.

DOI: <https://doi.org/10.1111/bmsp.12219>

6. Ma C., Ouyang J., Wang C., Xu G. A Note on Improving Variational Estimation for Multidimensional Item Response Theory. *Psychometrika*. 2024. Vol. 89. P. 89-113.

DOI: <https://doi.org/10.1007/s11336-023-09939-0>

7. Markos A., Tsigilis N. Dimensionality assessment in ordinal data: a comparison between parallel analysis and exploratory graph analysis. *Frontiers in Psychology*. 2024. Vol. 15. Article 1359111.

DOI: <https://doi.org/10.3389/fpsyg.2024.1359111>

8. Xia Y., Zhou X. Improving the Use of Parallel Analysis by Accounting for Sampling Variability of the Observed Correlation Matrix. *Educational and Psychological Measurement*. 2024. Vol. 84. P. 1156-1182.

DOI: <https://doi.org/10.1177/00131644241268073>

9. Cole K., Paek I. SAS PROC IRT and the R mirt Package: A Comparison of Model Parameter Estimation for Multidimensional IRT Models. *Psych*. 2023. Vol. 5(2). P. 416-426.

DOI: <https://doi.org/10.3390/psych5020028>

10. Hosseinzadeh M., Cole K. L. M. Effects of the Quantity and Magnitude of Cross-Loading and Model Specification on MIRT Item Parameter Recovery. *Educational and Psychological Measurement*. 2024. Vol. 84. P. 948-974.



DOI: <https://doi.org/10.1177/00131644231210509>.

11. Jewsbury P. A., van Rijn P. W. IRT and MIRT Models for Item Parameter Estimation With Multidimensional Multistage Tests. *Journal of Educational and Behavioral Statistics*. 2020. Vol. 45(4). P. 383-402.

DOI: <https://doi.org/10.3102/1076998619881790>.

12. Ackerman T. A. Still Interested in Multidimensional Item Response Theory Modeling? Here Are Some Thoughts on How to Make It Work in Practice. *Educational Measurement: Issues and Practice*. 2024. Vol. 43(4). P. 22-31.

DOI: <https://doi.org/10.1111/emip.12645>.

13. Kim S., Lee W. Several Variations of Simple-Structure MIRT Equating. *Journal of Educational Measurement*. 2022. Vol. 60(2), P/ 265-290.

DOI: <https://doi.org/10.1111/jedm.12341>.

14. Hartig J., Höhler J. Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*. 2009. Vol. 35(2-3). P. 57-63.

DOI: <https://doi.org/10.1016/j.stueduc.2009.10.002>.

15. Li Y., Jiao H., Lissitz R. W. Applying multidimensional item response theory models in validating test dimensionality: An example of K-12 large-scale science assessment. *Journal of Applied Testing Technology*. 2012. Vol. 13(2). P. 1-27. URL: <https://www.testpublishers.org/assets/applying%20multidimensional%20item.pdf>.

16. Круглова Н., Диховичний О., Лисенко Д. Застосування моделей IRT та MIRT до аналізу тестів з аналітичної геометрії. *Адаптивні системи автоматичного управління*. 2021. № 38. С. 15-28.

DOI: <https://doi.org/10.20535/1560-8956.38.2021.233179>.

17. Круглова Н., Диховичний О. Choosing MIRT Model for Analysis of Quality of Pedagogical and Psychological Tests. 2022 *IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC)*. Kyiv, Ukraine, 2022. P. 1-4.

DOI: <https://doi.org/10.1109/SAIC57818.2022.9922918>.



18.Шимчик С.Д., Круглова Н. В., Диховичний О. О. Вибір моделі MIRT для аналізу тестів з вищої математики. *Тези XIII Всеукраїнської наукової конференції молодих математиків*. Київ: Український державний університет імені Михайла Драгоманова, 2025. С. 196-197.

URL: <https://matan.kpi.ua/media/2025/young-math-2025/young-math-2025-abstracts.pdf>

19.Ayanwale M. A., Oke A. O., Ayanwale B. A. Multidimensional item Response Theory Calibration of Dichotomous Response Structure Using R Language for Statistical Computing. *Interchange*. 2024. Vol. 55. P. 89-108.

DOI: <https://doi.org/10.1007/s10780-024-09517-y>.

20.Muraki E. A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*. 1992. Vol. 16. P. 159-176.

URL: <https://scispace.com/pdf/a-generalized-partial-credit-model-application-of-an-em-135pwqlqbw.pdf>.

21. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*. 1969. No. 34. P. 100-114.

URL: <https://www.psychometricsociety.org/sites/main/files/file-attachments/mn17.pdf>.

22. Bock R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*. 1972. Vol. 37. P. 29-51.

URL: <https://link.springer.com/article/10.1007/BF02291411>.

23.Chalmers F. *Multidimensional Item Response Theory Workshop in R*. York University. 2015. 88 P.

URL:https://philchalmers.github.io/mirt/extra/mirt-Workshop-2015_Day-1.pdf.

24.Hu L. T., Bentler P. M. 1999. *Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives*. Structural Equation Modeling. Vol. 6(1): 1–55.

URL:<https://doi.org/10.1080/10705519909540118>



25. Yen W. M. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*. 1984. Vol. 8(2). P. 125-145.

DOI: <https://doi.org/10.1177/014662168400800201>.