УДК 37.07

А.О. Іскандарова-Мала

АНАЛІЗ ЯКОСТІ СЕРЕДНЬОЇ ОСВІТИ

Анотація. В рамках освітнього контексту ключовою метою є оцінка придбаних навичок учнів і кластеризація студентів відповідно до їх рівня здібностей. У зв'язку з цим необхідно враховувати відповідний елемент, з якого випливає можливий вплив на школярів. З цією метою надано методологічний інструмент, який враховує багаторівневу структуру даних (тобто учнів в школах) відповідним чином. Такий підхід дозволяє згрупувати як учнів, так і школи в однорідні класи здібностей і ефективності, а також оцінити вплив певних характеристик учнів і шкіл на ймовірність приналежності до таких класів.

Ключові слова: Em-алгоритм, irt, багаторівневі моделі відповідей, двопараметрична логістична модель, ЗНО.

Вступ. Класичною моделлю для профілів запитань (ймовірність респондента з рівнем знань θі правильної відповіді на запитання зі складністю не вище за βj) вважають двопараметричну модель Бірнбаума:

$$P(\theta_i, \beta_j) = \frac{e^{D \cdot a_j(\theta_i - \beta_j)}}{1 + e^{D \cdot a_j(\theta_i - \beta_j)}},$$

де $D = 1.7$ константа, a_j - коефіцієнт дискримінації (item discrimination parameter), що визначає нахил характеристичної кривої.

Будемо вважати, що θ і β в процесі експерименту (тестування) залишаються незмінними, тоді можливо знайти оцінки ймовірностей, які пов'язані з θ і β. Припустимо, що результат відповіді і-го респондента на j-е завдання дорівнює $r_{i,j}$, де $r_{i,j} = 1$, якщо відповідь вірна (можна використовувати зважену оцінку $r_{i,j} > 0$), у зворотному випадку, дорівнює 0. Тоді оцінка рівня знань респондента дорівнює:

$$\widehat{\theta_i} = \frac{1}{M}\sum_{j=1}^{M} r_{i,j},$$

а оцінка рівня складності завдання тесту дорівнює:

$$\widehat{\beta_i} = 1 - \frac{1}{N}\sum_{i=1}^{N} r_{i,j},$$

де M — кількість завдань тесту, N — кількість респондентів.

Респондент з більш високим рівнем знань відповідає вірно на завдання з ймовірністю не менше ніж респондент з нижчим рівнем знань, отже, маємо:

$$P(\theta_i) \geq P(\theta_k), \text{якщо } \theta_i > \theta_k$$

Ми бачимо, що залежність $P(\theta_i)$ має неспадаючий характер для фіксованого рівня складності запитання β. Тобто, $P(\theta_i, \beta)$ є характеристичним профілем завдання. Виходячи з цього, $P(\theta_i, \beta)$ є кумулятивною кривою ймовірностей, кожна точка якої відповідає ймовірності того, що респондент із рівнем знань не більшим за θ дає правильну відповідь на запитання з рівнем складності β. Таким чином, оцінки характеристичної кривої запитання визначаються як:

$$P_j(\theta) = \frac{1}{N}\sum\left\{r_{i,j} \middle| \theta_i \leq \theta\right\}.$$

Оскільки ми маємо лише множину з $\tilde{N}$ оцінок θ, то отримуємо $\tilde{N}$ емпіричних точок

$$\tilde{P}_j(\theta_k) = \frac{1}{\tilde{N}}\sum\left\{r_{i,j} \middle| \theta_i \leq \theta_k\right\}, k = 1, \ldots, \tilde{N}.$$

Дані зовнішнього незалежного оцінювання (ЗНО).

Для аналізу було взято результати основної сесії ЗНО 2017 року з урахуванням різних показників. База містить деперсоніфіковані дані всіх учасників тестування.

Результати за шкалою 0-200 балів за найпопулярнішими предметами: Українська мова та література, Історія України, Математика, Фізика, Англійська мова.

Завдання на ЗНО різного типу – множинний вибір, відкриті питання, питання на відповідність тощо.

Попередні аналізи показують, що результати дуже різняться в залежності від статі учнів, географічного положення та виду навчального закладу. Наприклад, в цілому, частина учнів з міст виконала тестування успішніше, ніж із сел.

 ISSN 1562-9945

Таблиця 1

Середній показник по типу території і географічної області для п'яти вимірів тестів ЗНО

	Українська мова та література		Історія України		Математика		Фізика		Англійська мова	
	Місто	Село	Місто	Село	Місто	Село	Місто	Село	Місто	Село
Вінницька область	142.3	120.3	126.1	108.9	123.3	95.7	110.7	83.8	125.8	99.8
Волинська область	148.6	120.2	129.5	104.6	135.9	103.7	120.3	95.1	137.1	118.2
Дніпропетровська область	139.7	114.0	122.4	100.6	121.9	86.3	108.7	79.4	119.6	89.3
Донецька область	138.0	118.0	121.5	99.9	116.5	83.3	100.4	76.8	115.1	93.1
Житомирська область	143.9	113.5	126.5	103.5	120.1	83.4	110.0	71.7	129.1	91.3
Закарпатська область	125.9	80.6	104.3	70.1	127.8	76.7	115.7	70.4	125.0	81.9
Запорізька область	137.6	113.2	121.4	102.8	120.9	84.1	101.1	71.3	123.5	76.8
Івано-Франківська область	145.0	114.4	125.8	98.0	136.1	104.6	121.3	95.7	135.6	102.8
Київська область	143.7	125.1	125.6	109.9	126.4	102.3	114.3	97.0	123.7	99.2
Кіровоградська область	139.6	114.7	123.4	98.1	119.4	86.9	107.7	66.5	124.1	93.5
Луганська область	137.8	118.6	123.8	103.5	116.3	95.2	98.3	71.1	119.6	96.2
Львівська область	152.7	125.4	134.3	110.6	137.6	108.5	119.7	92.1	135.8	110.8
Миколаївська область	136.9	109.9	118.3	98.6	116.7	81.3	98.4	65.0	110.5	70.0
Одеська область	135.7	110.9	118.8	98.1	119.2	85.6	98.4	70.8	117.6	82.3
Полтавська область	142.8	122.9	125.1	107.7	121.6	97.6	106.6	83.2	126.4	105.6
Рівненська область	144.7	104.6	127.6	90.5	128.6	85.3	114.1	70.7	131.7	92.9
Сумська область	144.5	121.0	127.5	109.5	126.0	97.1	107.8	91.6	129.6	99.6
Тернопільська область	149.0	122.2	134.5	111.6	131.2	98.7	110.5	81.4	135.3	113.8
Харківська область	141.5	119.3	125.2	102.9	124.7	92.2	109.7	73.6	128.7	106.1
Херсонська область	134.3	111.3	117.9	99.1	115.6	82.9	104.5	69.9	115.2	88.8
Хмельницька область	143.6	114.1	129.0	101.8	125.0	90.6	111.8	76.3	124.7	88.4
Черкаська область	146.1	122.2	127.0	107.6	131.0	103.9	114.1	87.3	130.5	103.9
Чернівецька область	136.9	88.1	119.0	73.3	119.5	80.3	111.6	65.3	134.3	115.4
Чернігівська область	142.7	115.9	125.4	104.6	125.2	93.6	110.3	80.4	127.2	98.1
м. Київ	149.7	127.2	134.0	125.4	143.2	101.3	130.6	88.9	140.0	100.3

Висновок на основі вірогідності.

Нехай існує вектор $\xi_v^{(V)}$ з елементами $\widehat{\theta_v}$, що відповідають рівню знань всіх без виключення респондентів, де v – клас. Отже, k_V – це кількість класів, на які можна поділити здатності учнів. За принципом IRT 2PL, k_V прийнято вважати рівним 3, тобто класи з низьким, середнім та високим рівнями здатностей.

$\pi_v^{(V)}$ – вага класу v , де $\sum_v \pi_v^{(V)} = 1$ та $\pi_v^{(V)} > 0$ для $v = 1, \ldots, k_V$.

Також задаємо вектор $\xi_u^{(U)}$ з елементами $\overline{\theta_u}$, що відповідають рівню знань згрупованих учнів (наприклад, в нашому випадку, за регіонами, або, як варіант, за статтю, за типом навчального закладу тощо), де u – тип групи.

Кількість типів груп обираємо використовуючи Байєсівський інформаційний критерій (BIC). Виходячи з цього критерію, кількість типів груп відповідає мінімальному значенню

$$BIC = -2\ell(\hat{\theta}) + \log(n)\#par$$

На практиці ми пропонуємо встановити модель для збільшення значень k_U при обмеженні $k_U \geq k_V$. Коли BIC починає збільшуватися, попереднє значення k_U приймається як оптимальне. Зверніть увагу, що всі інші елементи, які характеризують модель, залишаються фіксованими.

При заданих k_U та k_V параметри можуть бути оцінені шляхом максимізації log-правдоподібності (ρ_h):

$$\ell(\theta) = \sum_{h=1}^{H} \log \sum_{u=1}^{k_U} \pi_{hu}^{(U)} \rho_h\left(\xi_u^{(U)}\right)$$

де θ - вектор, що містить всі вільні параметри, і

$$\rho_h\left(\xi_u^{(U)}\right) = \prod_{i=1}^{n_h} \sum_{v=1}^{k_V} \pi_{(hi,v|u)}^{(V)} \prod_{j=1}^{r} p\left(\psi_{hij} \middle| V_{hi} = v\right)$$

де h – група, i – учень (респондент), j – завдання тесту, $\psi_{\downarrow}hij \mid V_{\downarrow}hi = v$ – означає реалізацію випадкової величини, яка відповідає відповіді на пункт j, наданий i -м респондентом.

Щоб зробити модель ідентифікуючою, ми приймаємо обмеження

$$\beta_{j_d} = 0, \gamma_{j_d} = 1, d = 1, \ldots, s,$$

де β_j (складність завдання) та γ_j (дискримінаційний індекс) для d -го.

Фактично існують коефіцієнти регресії $(k_V - 1)(m_V + k_U) + (k_U - 1)(m_U + 1)$ для прихованих класів, параметри здатності $k_V s$, параметри дискретизації $r - s$ і параметри труднощів $r - s$, вектор коваріацій індивідуального рівня (m_U) та групового рівня (m_V). При параметризації їх кількість зменшується на $r - s$, так як дискримінаційні індекси не оцінюються.

 ISSN 1562-9945

Для вибору кількості типів груп ми складаємо модель з коваріатами (тип території і географічна область) в п'ятивимірній версії (предмети: V1 - Українська мова та література, V2 - Історія України, V3 - Математика, V4 - Фізика, V5 - Англійська мова). Нехай значення k_U приймемо 6. Результати цієї підгонки представлені в таблиці 2. Отже, виберемо $k_U = 5$, так як відповідно до цієї кількісті типів спостерігається найменше значення BIC .

Таблиця 2

Визначення k_U; жирним шрифтом виділено найменше значення BIC

k	ℓ	#par	BIC
1	-551442.8	301	1069688
2	-550185.9	309	1067501
3	-559744.5	315	1067035
4	-549930.3	319	1066838
5	-549789.9	335	1066531
6	-549571.8	339	1066665

Щоб максимізувати log-правдоподібність $\ell(\theta)$, ми використовуємо алгоритм максимізації очікування (ЕМ).

Повна логарифмічна ймовірність, на якій заснований алгоритм ЕМ, може бути виражена як

$$\ell^*(\theta) = \sum_{h=1}^{H} [\ell_{1h}^*(\theta) + \ell_{2h}^*(\theta) + \ell_{3h}^*(\theta)],$$

$$\ell_{1h}^*(\theta) = \sum_{i=1}^{n_h} \sum_{j=1}^{J} \sum_{v=1}^{k_V} z_{hiv} \log p\left(\psi_{hij} \middle| V_{hi} = v\right),$$

$$\ell_{2h}^*(\theta) = \sum_{i=1}^{n_h} \sum_{u=1}^{k_u} \sum_{v=1}^{k_V} z_{hu} z_{hiv} \log \pi_{(hi,v|u)}^{(V)},$$

$$\ell_{3h}^*(\theta) = \sum_{u=1}^{k_1} z_{hu} \log \pi_{hu}^{(U)},$$

де z_{hiv} - індикаторна функція для суб'єкта i , яка перебуває в латентному класі v $(V_{hi} = v)$, а z_{hu} – індикаторна функція для кластера h , що має типологію u $(U_h = \xi_u^{(U)})$. Отже, $z_{hu} z_{hiv}$ дорівнює 1, якщо обидві умови виконані і 0 в іншому випадку.

Зазвичай $\ell^*(\theta)$ набагато простіше максимізувати по відношенню до $\ell(\theta)$. Алгоритм ЕМ чергує наступні два кроки:

– Е-крок. Він складається в обчисленні очікуваного значення повної лог-правдоподібності $\ell^*(\theta)$. На практиці це еквівалентно обчисленню задніх очікуваних значень змінних індикатора. Зокрема, ми маємо:

$$\hat{z}_{hiv} = p(V_{hi} = v|\mathcal{D}) = \sum_{u=1}^{k_U} [(z)]_{hu} \widehat{z_{hiv}})$$

де $\mathcal{D}$ – скорочений запис спостережуваних даних. Більш того, ми маємо:

$$|V_{hi} - v^{\square}\uparrow$$

та

$$\hat{z}_{hu} = p\left(U_h = \xi_u^{(U)} \middle| \mathcal{D}\right) = \frac{\rho_h\left(\xi_u^{(U)}\right)}{\sum_{h'=1}^{H} \rho_{h'}\left(\xi_u^{(U)}\right)}.$$

– М-крок. Він складається з поновлення параметрів моделі, максимізуючи очікуване значення $\ell^*(\theta)$, обчислене на Е-кроці. Точніше, в той час як для окремих ваг класу існує явне рішення, а для інших параметрів явне рішення не існує. Тому використовується ітераційний алгоритм оптимізації типу Ньютона-Рафсона. Отримані оцінки θ використовуються для поновлення $\ell^*(\theta)$ на наступному Е-кроці.

Коли алгоритм сходиться, останнє значення θ, що позначається $\hat{\theta}$, відповідає максимуму $\ell(\theta)$, а потім береться як оцінка максимальної правдоподібності цього параметра. Важливо підкреслити, що кількість ітерацій і, зокрема, виявлення глобальної, а не локальної точки максимуму у більшій мірі залежить від ініціалізації ЕМ-алгоритму. Тому рекомендовано спробувати кілька ініціалізацій цього алгоритму.

Після оцінки параметрів кожен суб'єкт i може бути призначений одному з класів k_V на основі шаблону відповіді ψ_i, яку було надано учнем, а також його коваріацій x_i і типології групи, до якої він належить. Аналогічно, кожна група h може бути віднесена одному з k_U- класів. В обох випадках найбільш поширеним підходом є призначення суб'єкта і групи класу з найвищою постеріорною ймовірністю, розрахованої, як в рівняннях на Е-кроці.

Застосування. Ми маємо справу з оцінкою параметрів IRT з $k_V = 3$ на рівні учнів, $k_U = 5$ на рівні груп і п’яти вимірів (V1, V2, V3, V4, V5), що позначають 5 предметів ЗНО.

 ISSN 1562-9945

Надалі здатності виражаються через стандартизований масштаб, а вагу класів отримуються як середні значення оцінених питомих ваг класу, які охоплюють $\hat{\pi}_v^{(V)}$, і виходять як

$$\hat{\pi}_v^{(V)} = \frac{1}{n}\sum_{h=1}^{H}\sum_{i=1}^{n_h}\hat{\pi}_{(hi,v|u)}^{(V)}.$$

У таблиці 4 показані оціночні здібності і середні ваги для трьох класів учнів і п'яти пов'язаних вимірювань разом з рангом здібностей в кожному стовпці для полегшення інтерпретації. Перевірка цих оцінок показує, що учні, які належать до класу 3 в тестах V1-V5, мають найменший рівень здатності. В цілому, частка учнів з низьким рівнем підготовленості, згрупованих за класом 3, вельми велика з точки зору класових пропорцій, оскільки вона становить трохи більше 32 відсотків учнів в цілому. Учні з найвищими рівнями здібностей відносяться до класу 1, який становить трохи більше 18 відсотків, а клас 2 – це клас випробовуваних з середніми рівнями здатності (більше 51%).

Таблиця 3

Розподіл здібностей на рівні учнів ($\hat{\xi}_v^{(V)}$)

	V1	V2	V3	V4	V5	$\hat{\pi}_v^{(V)}$
Class 1	0.878	0.815	0.732	0.614	0.642	0.183
Class 2	0.521	0.509	0.448	0.399	0.474	0.512
Class 3	0.298	0.191	0.188	0.140	0.194	0.321

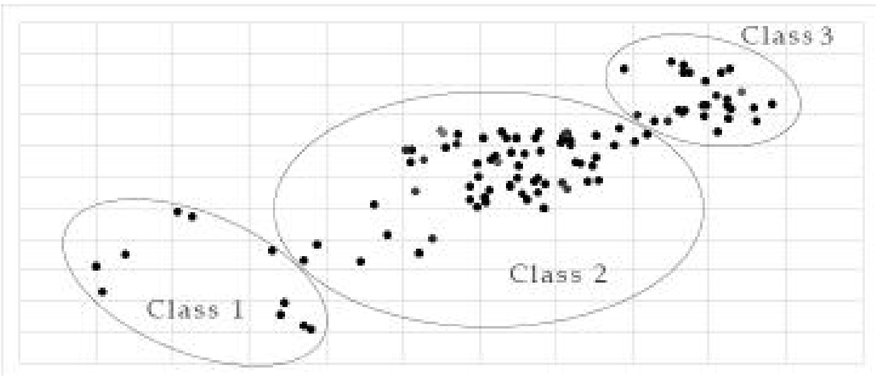

Рисунок 1 - Розподіл здібностей на рівні учнів на 3 класи

На рівні груп (міські та сільські школи за областями) розподіл оціночних середніх здібностей для п'яти обраних типів (k_U = 5) дозволяє нам класифікувати школи від найгірших (тип 1) до кращих (тип 5) шкіл. Ми бачимо, що 18,2% шкіл належать до кращих типів (тип 4 і 5), тоді як 31,1% ставляться до найгірших; інші 50,6% є проміжними, тобто середніми.

Таблиця 4

Розподіл здібностей на рівні шкіл

	Type 1	Type 2	Type 3	Type 4	Type 5
$\xi_u^{(U)}$	0.216	0.307	0.488	0.679	0.821
$\hat{\pi}_u^{(U)}$	0.098	0.213	0.506	0.123	0.059

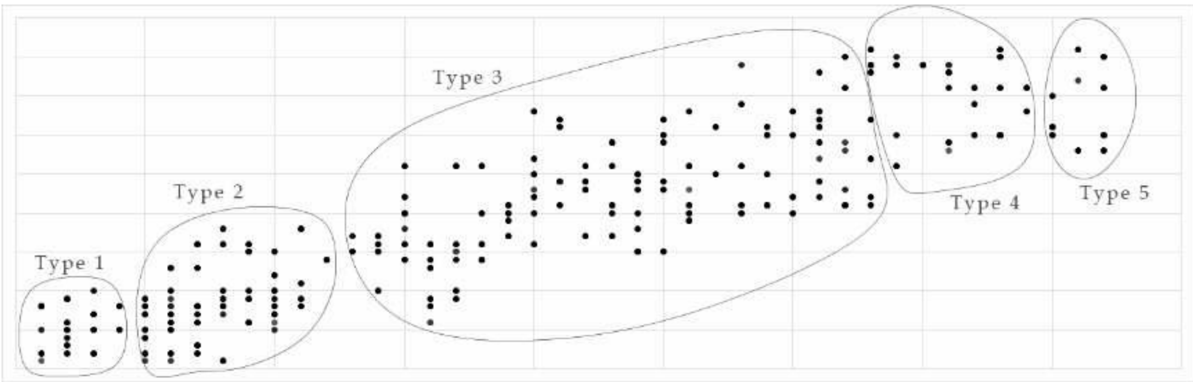

Рисунок 2 - Розподіл здібностей на рівні шкіл на 5 типів

У перших двох типах, тобто, школи з найслабшою підготовкою учнів, найбільшу частку займають сільські школи різних областей країни. «Міста», як правило, групуються у 3-5 типи, їх результати вищі.

Тепер згрупуємо результати міських і сільських шкіл за районами України и розглянемо розподіл здібностей.

Результати, наведені в таблиці 5, підтверджують, що попередній описовий аналіз (див. Таблицю 1) вже частково виявлено, тобто дуже різноманітні рівні досягнень при обліку шкільних географічних районів. Але тепер чітко видно, що найгірший рівень підготовки учнів мають школи Західної частини країни (частка таких шкіл становить більше 60%). Як показують попередні аналізи, ця ситуація склалась через те, що кількість сіл у цьому районі найбільша, а, як вже відомо, сільські школи – це «слабке місце» у системі середньої освіти України.

Проте, слід відмітити, що центральна частина України по кількості сільських шкіл знаходиться на другому місці і частка з низьким рівнем підготовки складає приблизно 23%.

 ISSN 1562-9945

Таблиця 5

Рівень підготовки у школах за районами України

	Кількість міст	Кількість сіл	Type 1 $\hat{\pi}_u^{(U)}$	Type 2 $\hat{\pi}_u^{(U)}$	Type 3 $\hat{\pi}_u^{(U)}$	Type 4 $\hat{\pi}_u^{(U)}$	Type 5 $\hat{\pi}_u^{(U)}$
Захід	134	8245	0.189	0.413	0.206	0.163	0.029
Північ	69	5781	0.031	0.223	0.305	0.321	0.119
Центр	81	6650	0.032	0.207	0.310	0.352	0.099
Схід	106	3869	0.036	0.234	0.406	0.263	0.061
Південь	51	3684	0.041	0.246	0.458	0.199	0.056

Висновки. Дані для аналізу були взяті на сайті Українського центру оцінювання якості освіти.

В цілому, представлене розширення моделі IRT дозволяє робити різноманітні «зрізи» для аналізу результатів і надання подальших висновків та рекомендацій. Ґрунтуючись на цій моделі, ми з'ясовуємо наявність прихованих класів випробовуваних, які демонструють послідовні рівні здатності по залученим вимірам (предметам ЗНО), а також кілька типів шкіл - від найменш підготовлених шкіл до найвищих досягнень. Потім вивчаємо взаємозв'язок, що спостерігається між змінними клас та тип. На рівні учнів ми виявляємо, що тип території (місто або село) значно впливає на членство в класі, як правило учні міських шкіл, групуються в клас з найвищим рівнем підготовки. На рівні школи результати показують, як і в якій мірі чинники, пов'язані зі шкільною географічною областю, впливають на ймовірність того, що школа буде згрупована за певним типом школи (наприклад, як за нашим аналізом, школи із заходу мають найнижчий рівень підготовки).

У представленому прикладі, було з’ясовано, які райони України мають проблеми з рівнем якості середньої освіти та визначені ці причини.

ЛІТЕРАТУРА

1. Челышкова М.Б. Теория и практика конструирования педагогических тестов.-М.:Логос, 2002.– 432 с.
2. Алгина Дж. Введение в классическую и современную теорию тестов: учебник / Дж. Алгина, Л. Крокер .– М.: Логос, 2010 .– 668 с.
3. Введение в теорию моделирования и параметризации педагогических тестов/Ю.М.Нейман, В.А.Хлебников.– М.:Прометей, 2000.– 168 с.
4. Шумейко А.А., Сотник С.Л. Интелектуальный анализ данных. Введение в Data Mining, 2014. – 223 с.

5. Abrahamowicz M., Multicategorical spline model for item response theory / M. Abrahamowicz, J. O. Ramsay // Psychometrika .– 1992 .– vol. 57, no. 1.– P. 5-27.

6. Winsberg S. Fitting item characteristic curves with spline functions / S. Winsberg, D. Thissen, H. Wainer // Educational Testing Service. Technical Report .– Princeton: NJ, 1984 .– P. 84-52.

7. Шумейко О.О., Кнуренко В.М. Побудова профілів IRT за допомогою сплайнів, які зберігають середнє значення/Математичне моделювання, №1 (32), 2015, с.28-32.

8. Reckase, M. D. (2009). Multidimensional Item Response Theory. Springer, New York.

9. Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. Psychometrika, 72:141.

10.Bartolucci, F., Bacci, S., and Gnaldi, M. (2014). MultiLCIRT: An R package for multidimensional latent class item response models. Computational Statistics and Data Analysis,71:971.

11.Шумейко О.О., Іскандарова-Мала А.О. Побудова профілів IRT за допомогою кусково-лінійної регресії з вільними вузлами / А.О.Іскандарова-Мала, О.О.Шумейко // збірник КНУ .— 2016 .

12.Шумейко О.О., Іскандарова-Мала А.О. Вплив адаптивного тестування на профілі IRT / А.О.Іскандарова-Мала, О.О.Шумейко // збірник КНУ .— 2017 .

13.Іскандарова-Мала А.О. Про вибір параметрів ЕМ-алгоритму для поділу суміші розподілів / А.О.Іскандарова-Мала, О.О.Шумейко // Математичне моделювання .— 2018 .

14.Іскандарова-Мала А.О. Адаптивне тестування на основі профілів IRT / А.О.Іскандарова-Мала // Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості та освіті: Тези XI Міжнародної науково-практичної конференції.— 2017 .— №11 .— С.178.

 ISSN 1562-9945