

Л. С. СМІДОВИЧ, О. В. АРТЮХ, Н. В. КОСЕНКО

АВТОМАТИЗАЦІЯ АНАЛІЗУ ЯКОСТІ ТЕСТОВИХ ЗАВДАНЬ ТА РЕЗУЛЬТАТІВ ТЕСТУВАННЯ СТУДЕНТІВ ІТ-СПЕЦІАЛЬНОСТЕЙ

Предметом дослідження статті є статистичні моделі і методи оцінювання якості тестових завдань, а також процеси та технології інформаційної підтримки оцінювання результатів тестування та якості тестів. **Метою** є підвищення якості тестових завдань та точності оцінювання результатів тестування в умовах постійних змін змісту навчальних програм і вмісту тестових завдань, та скорочення часу на обробку результатів тестування за рахунок автоматизації, застосування розроблених алгоритмів та інформаційної системи тощо. В роботі вирішуються наступні **задачі**: аналіз процесу підготовки та проведення тестування; проведення аналітичного огляду статистичних методів оцінювання якості тестових завдань та показників якості тестових завдань; розробка узагальненого процесу підготовки та проведення тестування; автоматизація аналізу результатів тестування студентів ІТ-спеціальностей. Для вирішення зазначених завдань були використані **методи** статистичного аналізу, модель Rasch IRT (Item Response Theory), технології розробки програмних додатків. Отримані наступні **результати**. Проаналізовано процес підготовки та проведення тестування, запропоновано його модифікацію з урахуванням особливостей тестування студентів ІТ-спеціальностей. Вибрано показники якості тестових завдань та моделі і методи їх розрахунку, зокрема модель Rasch IRT. Наведено інтерпретацію значень показників якості тестових завдань, та розрахунок коефіцієнтів аномалії якості завдання, які є основою для формування поради щодо поліпшення якості тесту. Розроблено прототип інформаційної системи підтримки процесу аналізу якості тестових завдань, оцінювання результатів тестування та прийняття рішень про модифікацію тесту. Проведено тестування системи при оцінюванні результатів модульного контролю студентів третього курсу. **Висновки**: автоматизація та інформаційна підтримка розрахунку та аналізу показників якості тестових завдань та результатів тестування дозволить підвищити точність оцінювання та скоротити час на обробку результатів тестування.

Ключові слова: оцінка якості тестових завдань; тестове завдання; процес тестування; статистичні методи; моделі IRT; валідність; надійність; диференціююча здатність.

Вступ

Питанню якості освіти та її управління надається важливе значення в законах і інших нормативних актах України, зокрема в Законі про вищу освіту. Впровадження стандартизованих освітніх вимірювань у вищій школі є одним із інструментів управління якістю вищої освіти. Одним із методів вимірювання і оцінювання рівня навчальних компетенцій та контролю якості знань студентів є застосування педагогічних тестів [1 - 3]. Тестування є одним із базових інструментів модульно-рейтингової системи (МРС) оцінювання знань студентів. При проектуванні тестів та тестових завдань, та їх застосуванні, актуальним є завдання оцінювання та контролю їх якості.

Особливістю навчання студентів ІТ-спеціальностей є швидка зміна технологій і відповідно вимог до компетенцій майбутніх фахівців, що потребує постійної зміни учбового матеріалу. Це в свою чергу вимагає постійної модернізації тестів та

розробки нових тестових завдань, і відповідно оцінювання їх якості з метою коригування.

Аналіз останніх досліджень і публікацій

При проектуванні та застосуванні тестів вирішуються наступні завдання: конструювання тесту, апробація – проведення пробного тестування, визначення та розрахунок показників якості тесту, оцінювання якості тесту, корегування тестових завдань та побудова кінцевого варіанту тесту, проведення тестування, оцінювання результатів тестування [3, 4].

Особливістю класичного процесу підготовки та застосування тесту (рис. 1) є наявність етапу апробації тесту перед його застосуванням, а також його лінійний характер в межах однієї сесії тестування. Після тестування може бути проведено розрахунок і аналіз показників якості тесту та відповідно корегування тестових завдань.

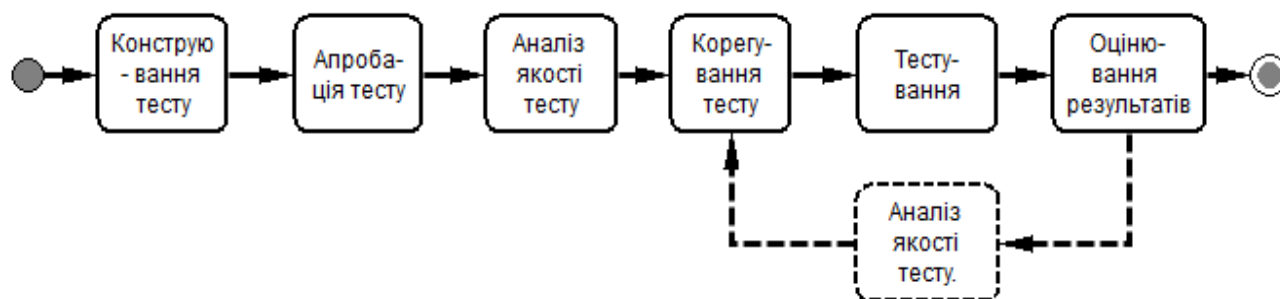


Рис. 1. Класичний процес підготовки тесту, тестування та оцінювання

В умовах швидкої зміни змісту тестових завдань використання традиційних методів конструювання тесту, які включають в себе експериментальну апробацію та наступне коригування тестових завдань з метою покращення їх якості, не завжди є ефективним. Актуальним є завдання розробки процесу, який би дозволяв проводити модифікацію тесту безпосередньо при оцінюванні результатів, без необхідності проведення іншої сесії тестування.

Для оцінювання показників якості тесту та результатів тестування використовують статистичні методи та моделі – як класичні статистичні методи [4, 5], так і моделі IRT (Item Response Theory) [6 - 8], які дозволяють оцінити якість тестових завдань з різних сторін. Початковими даними для статистичних методів оцінювання якості тестів є результати тестування.

В основі класичних методів оцінювання якості тестів лежить статистична теорія вимірювань та ідея еталонної оцінки [4, 5, 9, 10]. Оцінка якості тесту та тестових завдань полягає в розрахунку таких характеристик, як валідність, надійність, диференціююча здатність.

Валідність – одна з характеристик якості тесту, що визначає відповідність змісту тесту, тому, що він повинен оцінити. Варто звернути увагу на те, що кореляційний зв'язок тестового завдання з тестом в цілому може бути нижчим за рахунок того, що випробуваний володіє тільки фрагментарними знаннями в предметі, списував чи намагався вгадати відповідь при проведенні тестування.

Надійність тесту вимірює ступінь стійкості результатів тестування. Отриманий коефіцієнт надійності показує кореляційний зв'язок між результатами різних вимірювань, проведених в однакових умовах.

Диференціююча здатність – спроможність тестового завдання розрізнити успішних випробовуваних від слабких. У кожній групі є випробовувані з різним рівнем підготовки (знань). Якщо диференціююча здатність тестового завдання висока, то більш успішні випробовані краще впораються із запропонованим тестом, а менш успішні – гірше.

Основне поняття теорії IRT [10 – 12] – це латентний параметр, який визначає вимірюється властивість випробованого, недоступне для прямого спостереження. Латентний параметр можна виміряти за допомогою його індикатора. Індикатор – властивість (параметр) випробованого, який доступний для прямого спостереження і може бути вимірний в числовій шкалою. Для того щоб виміряти значення латентного параметра, необхідно виміряти значення пов'язаного з ним індикатора. Наприклад, якщо латентним параметром є рівень знань випробованого з певного предмета, то індикатором може бути бал, який випробуваний отримав в результаті проходження тестового завдання.

Основні припущення IRT:

1) існують латентні параметри особистості випробованого, які недоступні для безпосереднього спостереження;

2) існують індикаторні змінні, пов'язані з латентними параметрами, які доступні для спостереження і які можливо розрахувати. За цим значенням можна судити про значення латентних параметрів;

3) оцінюваний латентний параметр повинен бути гомогенним. Це означає, що тест має вимірювати знання тільки в одній, чітко заданій, предметній області. Якщо ця умова не виконується, то необхідно переробити тест, видаливши або змінивши завдання, що порушують його гомогенність.

Основне завдання моделей IRT полягає в переході від індикаторних змінних до латентних параметрів. Після того, як будуть розраховані індикаторні змінні, пов'язані з цими параметрами, можна привести латентні параметри до однієї шкалою та порівняти їх. Також в IRT системи дозволяють встановлювати зв'язок між кількома латентними параметрами.

Моделі IRT і зокрема модель Раша має переваги перед класичною теорією тестів [13]: перетворює вимірювання, виконані в дихотомічних і порядкових шкалах, в лінійні вимірювання, в результаті якісні дані аналізуються за допомогою кількісних методів; міра вимірювання параметрів моделі Раша є лінійною, що дозволяє використовувати широкий спектр статистичних процедур для аналізу результатів вимірювань; оцінка труднощі тестових завдань не залежить від вибірки досліджуваних, на яких вона була отримана; оцінка рівня підготовленості випробовуваних не залежить від використовуваного набору тестових завдань.

Моделі IRT, зокрема модель Раша (Rasch model), застосовуються, в тому числі, для аналізу результатів тестування і якості тестових завдань [13 - 15].

В даному випадку в якості латентних параметрів виступають рівень підготовленості випробованого і рівень складності завдання.

Завдання дослідження. Таким чином, при практичному застосуванні тестування для контролю якості знань, зокрема студентів ІТ-спеціальностей, актуальним є завдання розробки процесу проведення тестування та аналізу його результатів, а також автоматизованої системи інформаційної підтримки тестування, які б забезпечили:

- мінімізацію впливу на результат тестування суб'єктивних факторів, як то списування, підказки, вгадування відповіді тощо;

- оперативний розрахунок показників якості всього тесту та окремих тестових завдань, та оцінювання результатів тестування;

- формування поради щодо поліпшення якості тесту;

- можливість оперативного корегування тесту за рахунок виключення неякісних завдань, та відповідного перерахунку результатів тестування;

- можливість автоматизації як самого процесу тестування, так і обчислення та аналізу результатів тестування та якості тесту.

Для забезпечення зазначених вище вимог треба виконати наступні завдання:

- провести дослідження і вибір методів аналізу якості тесту і тестових завдань;

- провести дослідження і вибір моделей IRT оцінки якості тестових завдань;

- надати інтерпретацію результатів аналізу тестових завдань, яка може бути використана при побудові системи підтримки прийняття рішень;

- виконати програмну реалізацію обраних методів, розробити прототип інформаційної системи підтримки та провести аналіз її застосування.

Метою даної статті є підвищення якості тестових завдань та точності оцінювання результатів тестування в умовах постійних змін змісту навчальних програм і змісту тестових завдань, та скорочення часу на обробку результатів тестування за рахунок автоматизації, застосування розроблених алгоритмів та інформаційної системи тощо.

Матеріали та методи

Методи розрахунку критеріїв якості тестових завдань. Після проведеного аналізу були обрані такі статистичні методи розрахунку перелічених вище критеріїв оцінки якості тестових завдань.

Оцінка валідності тестових завдань:

$$r_{x_j} = \frac{1}{s_x} (x_{j\theta} - x_{jH}) \sqrt{p_j q_j},$$

де $x_{j\theta}$ – середній арифметичний бал студентів, які успішно виконали j -е завдання тесту; x_{jH} – середній арифметичний бал студентів, які не виконали j -е завдання; $\sqrt{p_j q_j}$ – стандартне відхилення по j -му завданню; s_x – стандартне відхилення по всьому тесту.

Коефіцієнт кореляції показує, наскільки завдання відповідає тематиці та змісту всього тесту.

Оцінка надійності тесту:

$$r_H = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k p_i q_i}{s_x^2} \right),$$

де k – число завдань в тесті; $\sum_{i=1}^k p_i q_i$ – сума дисперсій завдань тесту; s_x^2 – дисперсія.

Оцінка диференціюючої здатності тестового завдання:

$$r_d = \frac{x_n - x}{\delta_x} \cdot \sqrt{\frac{n}{N_d - n}},$$

де x – середнє арифметичне значення всіх індивідуальних оцінок по тесту; x_n – середнє арифметичне значення оцінок по тесту у тих випробовуваних, які правильно розв'язали завдання; δ_x – середньоквадратичне відхилення індивідуальних оцінок по тесту для вибірки; n – число випробовуваних, які правильно розв'язали задачу; N_d – загальне число випробовуваних.

Коефіцієнт дискримінації тестового завдання може приймати значення від -1 до +1.

При застосуванні моделі IRT Раша (Rasch model) для аналізу результатів тестування та показників якості тестових завдань в якості латентних параметрів виступають рівень підготовленість випробуваного і рівень складності завдання.

Перша множина містить значення рівнів підготовленості випробовуваних θ_i , де i – номер випробовуваного, який приймає значення в інтервалі від 1 до N (N – кількість випробовуваних). Друга множина містить значення труднощі j -го завдання β_j . Індекс j приймає значення в межах від 1 до M , де M – кількість завдань у тесті.

Рівень підготовленості випробуваного θ_i та рівень складності завдання β_j розміщені на одній шкалі і вимірюються в одних і тих же одиницях – логітах. Аргументом функції успіху випробуваного є різниця $\theta_i - \beta_j$.

Якщо ця різниця позитивна і велика, то відповідно висока ймовірність досягнення успіху i -го випробуваного в j -м завданні. Якщо ж ця різниця негативна і велика по модулю, то ймовірність досягнення успіху i -го випробуваного в j -м завданні буде низькою.

Розглянемо метод [14 – 16] обчислення і порівняння рівня підготовленості випробуваних θ_i і рівня складності тестових завдань β_j (рис. 1).

Спочатку необхідно обчислити частки вірних p_i і невірних q_i відповідей випробовуваних по тесту:

$$p_i = \frac{X_i}{M}; \quad q_i = 1 - p_i,$$

де X_i – індивідуальний бал випробовуваного; M – кількість завдань у тесті.

Частка вірних p_i та не вірних q_i відповідей студентів по кожному завданню окремо:

$$p_j = \frac{X_j}{N}; \quad q_j = 1 - p_j,$$

де X_j – сумарний бал відповідей на тестове завдання; N – кількість випробовуваних.

Початкові значення рівня підготовленості випробуваних розраховуються по формулі:

$$\theta_i^0 = \ln \frac{p_i}{q_i},$$

Початкові значення складності тестових завдань β_j :

$$\beta_j^0 = \ln \frac{q_j}{p_j}.$$

Середні значення рівня підготовленості випробуваних та складності тестових завдань:

$$\bar{\theta} = \frac{\sum_{i=1}^N \theta_i^0}{N}; \quad \bar{\beta} = \frac{\sum_{j=1}^M \beta_j^0}{M}.$$

Значення дисперсії рівня підготовленості випробуваних S_θ та складності S_β [14]:

$$S_\theta = \frac{\sum_{i=1}^N (\theta_i^0)^2 - N(\bar{\theta})^2}{N-1}; \quad S_\beta = \frac{\sum_{j=1}^M (\beta_j^0)^2 - M(\bar{\beta})^2}{M-1}.$$

Обчислюємо кутові коефіцієнти:

$$a_\theta = \sqrt{\frac{1+S_\beta/2,89}{1-S_\theta S_\beta/8,35}}; \quad a_\beta = \sqrt{\frac{1+S_\theta/2,89}{1-S_\theta S_\beta/8,35}}.$$

Нарешті, ми можемо записати оцінки параметрів θ і β в єдиній інтервальной шкалі:

$$\theta_i = a_\theta \theta_i^0 + \bar{\beta}; \quad \beta_j = a_\beta \beta_j^0 + \bar{\theta}.$$

Після того, як будуть розраховані рівні підготовленості випробуваних і рівні складності тестових завдань, необхідно їх порівняти. Різниця цих значень вказує на те, наскільки тест відповідає рівню знань студентів.

Далі розраховуються ймовірності правильної відповіді i -го випробуваного на j -е завдання:

$$P_{ij} = \frac{1}{1 + \exp(-(\theta_i - \beta_j))}.$$

Після чого можна побудувати характеристичну криву для кожного із завдань, яка показує залежність загальної оцінки за тест від кількості балів, набраних за дане завдання:

$$P_j(\theta) = \frac{1}{1 + \exp(-(\theta - \beta_j))}; \quad j = \overline{1, N}.$$

Результати досліджень

Методика тестування. Для мінімізації впливу суб'єктивних факторів використовуються тести із випадковою генерацією варіантів тестових завдань. Варіант формується випадковим вибором завдань із банку завдань, відповідно до профілю тесту. Профіль визначає, скільки завдань із кожного блоку (який відповідає розділу учебної програми) та/або кожного рівня складності має бути включено в варіант тесту.

Таким чином, кожний студент отримує підмножину завдань із банку ($T_i \in T$), і кожне завдання виконується частиною студентів ($S_j \in S$).

Використовується дихотомічна схема оцінювання тестових завдань, тобто за виконання кожного завдання студент може отримати або 0 або 1.

Використаються тестові завдання двох типів: із одиночним вибором та із множинним вибором. При використанні завдання із множинним вибором воно вважається виконаним тільки якщо правильно вибрані всі варіанти відповідей.

Для зменшення вірогідності вгадування завдання тесту містять не менш 4 (зазвичай 4-5) варіантів відповіді (дистракторів).

Процес підготовки та застосування тесту змінено наступним чином: після сесії тестування виконується обробка результатів, яка включає в себе оцінювання результатів та розрахунок і аналіз показників якості тестових завдань.

По результатам аналізу може бути проведено корегування тесту (без проведення повторного тестування) – з тесту видаляються тестові завдання, якість яких не відповідає критеріям. Корегування тесту проводиться експертом, з використанням поради щодо поліпшення якості тесту, які надає автоматизована інформаційна система.

Після модифікації тесту перераховуються результати тестування та показники якості. Ці етапи – оцінювання, аналіз, корегування тесту – можуть повторюватися ітераційно. На кожній ітерації є можливість повернутися до попереднього варіанту тесту.

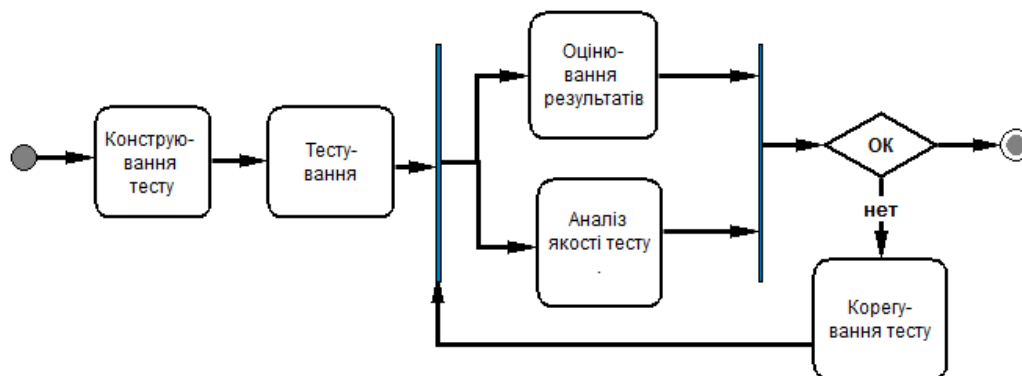


Рис. 2. Запропонована модифікація процесу підготовки та застосування тесту (при неможливості його попередньої апробації)

Інтерпретація результатів і формування поради щодо поліпшення тесту. Для аналізу якості тестових завдань рекомендують [17, 18] використовувати такі показники, як диференціююча здатність, надійність,

рівень складності, кутовий коефіцієнт по моделі IRT a_{θ} тощо.

Розраховані значення показників якості тестових завдань інтерпретуються наступним чином (табл. 1).

Таблиця 1. Значення показників якості тестових завдань та їх інтерпретація

Показник якості	Значення	Інтерпретація
Рівень складності	< 0,2	Надто легке, може бути виключено із тесту.
	0,2 – 0,8	Складність відповідає рівню підготовки студентів.
	> 0,8	Надто складне, може бути виключено із тесту.
Диференціююча здатність	< 0	Диференціююча здатність незадовільна, завдання має бути виключено із тесту.
	0 – 0,15	Диференціююча здатність погана, тестове завдання неправильно сформульовано, завдання необхідно переформулювати або замінити.
	> 0,15	Диференціююча здатність задовільна або добра.
Критерій надійності	0,8 – 1	Надійність тестового завдання хороша, відсутні похибки вимірювання.
	0,5 – 0,8	Надійність тестового завдання задовільна, наявне пропорційне співвідношення істинної оцінки і похибки.
	< 0,5	Надійність незадовільна, присутні значні похибки. Дане тестове завдання має бути виключено із тесту.
Кутовий коефіцієнт по моделі IRT a_{θ}	< 0,4	Диференціююча здатність незадовільний, завдання може бути виключено із тесту.
	0,4 – 0,7	Диференціююча здатність задовільна.
	> 0,7	Диференціююча здатність добра.
Критерій валідності	0,7 – 1	Сильний зв'язок тестового завдання з усім тестом, завдання повністю відповідає загальному змісту тесту.
	0,4 – 0,7	Задовільний зв'язок тестового завдання з усім тестом, завдання відповідає загальному змісту тесту.
	< 0,4	Незадовільний зв'язок тестового завдання з усім тестом, завдання не відповідає змісту тесту. Дане тестове завдання необхідно виключити із тесту.

Критерій відповідності рівня складності тестового завдання рівню підготовленості випробовуваних:

$$\sum_{j=1}^M \beta_j > 0 \quad - \text{складність тесту надто висока,}$$

необхідно спростити тест відповідно до рівня підготовки студентів;

$$\sum_{j=1}^M \beta_j = 0 \quad - \text{складність тесту відповідає рівню}$$

підготовки студентів;

$$\sum_{j=1}^M \beta_j < 0 \quad - \text{складність тесту надто низька,}$$

необхідно збільшити складність тесту відповідно до рівня підготовки студентів.

Графік характеристичної кривої (рис. 3), в першу чергу кут її нахилу, наочно відображає диференціюючу здатність тестового завдання. Зворотний нахил свідчить про те, що слабкі випробовувані відповідають на завдання краще, ніж сильні – таке завдання є некоректним і має бути виключено із тесту.

Слабкий нахил свідчить про поганий диференціює здатності завдання. На рис. 3 можна бачити, що завдання з групи (1) мають задовільну диференціюючу здатність, крива завдання (2) має менший нахил, що свідчить про гіршу диференціює здатності, а крива (3) має зворотний нахил, і це завдання має бути виключено з тесту.

Таким чином, кожне тестове завдання характеризується вектором значень показників якості.

$$Q_i = \{q_{i1}, q_{i2}, \dots, q_{iM}\}.$$

Якщо значення показника якості тестового завдання класифікується як незадовільне згідно правил описаних у табл. 1, відповідний коефіцієнт аномалії якості a_{ij} для i -го завдання встановлюється рівним 1.

Класифікація тестового завдання як неякісного виконується по значенню модуля вектору аномалії якості завдання

$$|A_i| = \sum_{j=1}^M a_{ij}.$$

Остаточне рішення про виключення неякісних завдань із тесту і відповідно перерахунок результатів приймає експерт. Рекомендується виключати не більш 10–15% завдань. Такі завдання можуть бути перероблені та включені в наступний варіант тесту, але не враховуються при розрахунку результатів поточної сесії тестування.

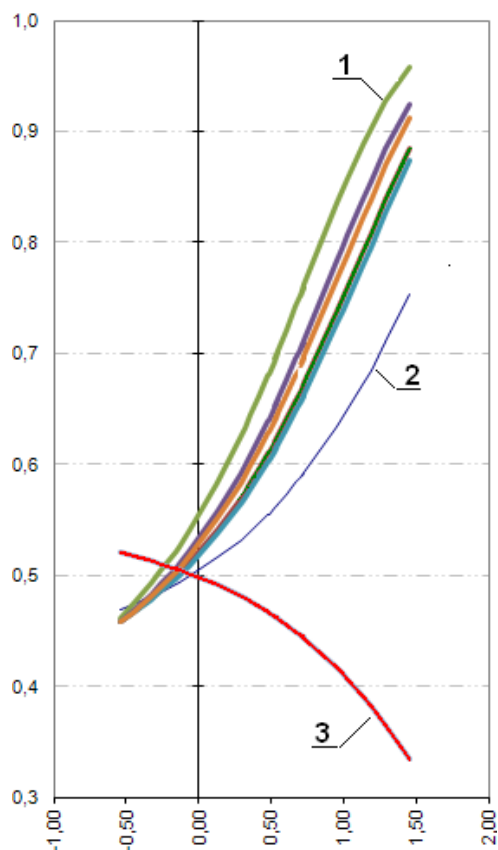


Рис. 3. Характеристичні криві тестових завдань, розраховані на основі моделі IRT Rasch

Аналіз результатів тестування і розрахунок показників якості тестових завдань

Було розроблено прототип веб-додатку аналізу якості показників тестових завдань та оцінювання

Список літератури

1. Аванесов, В. С. Научные проблемы тестового контроля знаний. Москва : ИЦПКПС, 1994. 86 с.
2. Аванесов, В. С. Форма тестовых заданий. Москва : Центр тестирования, 2006. 156 с.
3. Омаров, М. А., Мурадова, В. Х. Інфологічне забезпечення інформаційної технології управління дистанційною освітою. *Сучасний стан наукових досліджень та технологій в промисловості*. 2018. № 4 (6). С. 146-153. DOI: <https://doi.org/10.30837/2522-9818.2018.6.146>
4. Ковальчук Ю. О. Теорія освітніх вимірювань. Ніжин : Видавець ПП Лисенко М. М., 2012. 200 с.
5. Лісова Т. В. Моделі та методи сучасної теорії тестів. Ніжин : Видавець ПП Лисенко М. М., 2012. 112с.
6. An, X., Yung, Y. (2014), "Item Response Theory: What it is and how you can use the IRT procedure to apply it", Paper SAS364-2014, Cary, NC, USA : SAS Institute Inc. 14 p. URL: <http://support.sas.com/resources/papers/proceeding> (дата відвідування 11.02.2020).
7. Razak, N. A., Razak, A. Z. K., Lei, M. T. (2012), "Examining Quality of Mathematics Test Items Using Rasch Model: Preliminary Analysis", *Procedia – Social and Behavioral Sciences*, Vol. 69, P. 2205–2214. DOI: <https://doi.org/10.1016/j.sbspro.2012.12.187>
8. Майоров А. Н. Теория и практика создания тестов для системы образования. Москва : Интеллект-Центр, 2002. 267 с.
9. Малеева Ю. А., Косенко В. В., Малеева О. В., Світличний Д. В. Створення колаборативного середовища розробки в системі дистанційного навчання. *Сучасний стан наукових досліджень та технологій в промисловості*. 2019. № 2 (8). С. 62–71. DOI: <https://doi.org/10.30837/2522-9818.2019.8.062>
10. Крокер, Л., Алгина Дж. Введение в классическую и современную теорию тестов. Москва : Логос, 2010. 668 с.

результатів тестування, який використовує вищевикладений підхід. За його допомогою було проведено аналіз результатів тестів, що застосовувалися для проведення модульного контролю студентів третього курсу. Вибірка містить результати тестування 32 випробовуваних, тест включає 30 завдань у формі запитань закритого типу.

Зокрема, були розраховані критерії якості тестових завдань, показники рівня складності завдань і рівня підготовленості випробовуваних, які потім були приведені до єдиної шкали. Були розраховані і побудовані емпіричні характеристичні криві тестових завдань (рис. 2).

Висновки

В результаті проведеного дослідження було проведено аналіз процесу підготовки та проведення тестування; проведено огляду статистичних методів оцінювання якості тестових завдань та показників якості тестових завдань, на основі якого обрано показники якості тестового завдання та методи їх розрахунку. Запропоновано модифікацію процесу підготовки та проведення тестування з урахуванням особливостей тестування студентів ІТ-спеціальностей. Надано інтерпретацію значень показників якості тестових завдань, яка є основою для формування поради щодо поліпшення якості тесту. Розроблено прототип веб-додатку для аналізу якості тестів, і з його допомогою проведено аналіз тестових завдань і оцінювання результатів модульного контролю.

Напрямок подальших досліджень є удосконалення методів аналізу, зокрема для тестів з випадковою вибіркою тестових завдань з банку завдань та тестів з недихотомічною схемою оцінювання тестових завдань, застосування методів дистракторного аналізу, а також інтеграція веб-додатку для аналізу результатів тестування і якості тестових завдань з існуючими системами підготовки тестових завдань та тестування.

11. Rasch, G. (1980), *Probabilistic Models for Some Intelligence and Attainment Tests (Expanded edition)*, Copenhagen : Danish Institute of Educational Research, 324 p.
12. Ким В. С. Анализ результатов тестирования в процессе Rasch measurement. *Педагогические измерения*. 2005. № 4. С. 39–45.
13. Rusch, T., Lowry, P. B., Mair, P., Treiblmaier, H. (2017), "Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory", *Information & Management*, Vol. 54, Issue 2, P. 189–203. DOI: <https://doi.org/10.1016/j.im.2016.06.005>
14. Wilson, M. (2005), *Constructing Measures: An Item Response Modeling Approach*, NJ : Lawrence Erlbaum Associates, 228 p. DOI: <https://doi.org/10.4324/9781410611697>
15. Михеев О. В. Математические модели педагогических измерений. *Педагогические измерения*. 2004. № 2. С. 75–88.
16. Диховичний А. А., Дудко А. Ф. Автоматизована система аналізу результатів комп'ютерного тестування з вищої математики. *Наукові праці ДонНТУ. Серія: "Педагогіка, психологія і соціологія"*. 2013. № 2(14). С. 103–110.
17. Suparman, U. (2016), "Analyzing the quality of english test items of daily, mid semester and final school examinations in bandar lampung: (assessment and evaluation in language teaching)", *International Conference On Educational Research & Evaluation, Yogyakarta*, 14 p. URL: <http://repository.lppm.unila.ac.id/id/eprint/3121> (дата відвідування 11.02.2020).
18. Quagrain, K., Arhin, A. K. (2017), "Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation", *Cogent Education*, Vol. 4 (1). P. 1–11. DOI: <https://doi.org/10.1080/2331186X.2017.1301013>.

References

1. Avanesov, V. S. (1994), *Scientific problems of test knowledge measurement [Nauchnye problemy testovogo kontrolya znani]*, Moscow, ITsPKPS, 86 p.
2. Avanesov, V. S. (2006), *Form of the tests [Forma testovykh zadani]*, Moscow, Tsentr testirovaniya Publ., 156 p.
3. Omarov, M., Muradova, V. (2018), "Infological support of the information technology for management of distance education", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (6), P. 146–153. DOI: <https://doi.org/10.30837/2522-9818.2018.6.146>
4. Koval'čuk, Ju. O. (2012), *Educational Measurement Theory [Teoriya osvity v vymirjuvan]*, Nizhyn, Publisher PP Lysenko M. M., 200 p.
5. Lisova, T. V. (2012), *Models and methods of the modern test theory [Modeli ta metody suchasnoyi teoriiy testiv]*, Nizhyn, Publisher PP Lysenko M. M. 112 p.
6. An, X. and Yung, Y. (2014), "Item Response Theory: What it is and how you can use the IRT procedure to apply it", Paper SAS364-2014 – Cary, NC, USA : SAS Institute Inc., 14 p., available at : <http://support.sas.com/resources/papers/proceeding> (last accessed 11.02.2020).
7. Razak, N. A., Razak, A. Z. K. and Lei, M.T. (2012), "Examining Quality of Mathematics Test Items Using Rasch Model: Preliminary Analysis", *Procedia - Social and Behavioral Sciences*, Vol. 69, P. 2205–2214. DOI: <https://doi.org/10.1016/j.sbspro.2012.12.187>
8. Maiorov, A. N. (2002), *Theory and practice of tests creating for the education system [Teoriya i praktika sozdaniya testov dlya sistemy obrazovaniya]*, Moscow, Intellect-Tsentr Publ., 267 p.
9. Malieieva, Ju., Kosenko, V., Malyeyeva, O., Svetlichnyj, D. (2019), "Creation of collaborative development in the system of distance learning", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (8), P. 62–71. DOI: <https://doi.org/10.30837/2522-9818.2019.8.062>
10. Croker, L., Algina, J. (2010), *Introduction to Classical and Modern Test Theory. [Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov]*, Moscow, Logos Publ., 668 p.
11. Rasch, G. (1980), *Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Danish Institute of Educational Research. (Expanded edition)*, 324 p.
12. Kim, V. S. (2005), "Test results analysis during Rasch measurement" ["Analiz rezul'tatov testirovaniya v protsesse Rasch measurement"], *Pedagogicheskie izmereniya – Pedagogical Measurements*, Vol. 4, P. 39–45.
13. Rusch, T., Lowry, P. B., Mair, P., Treiblmaier, H. (2017), "Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory", *Information & Management*, Vol. 54, Issue 2, P. 189–203. DOI: <https://doi.org/10.1016/j.im.2016.06.005>
14. Wilson, M. (2005), *Constructing Measures: An Item Response Modeling Approach*, NJ : Lawrence Erlbaum Associates, 228 p. DOI: <https://doi.org/10.4324/9781410611697>
15. Mikheev, O. V. (2004), "The mathematical model of pedagogical measurements" ["Matematicheskie modeli pedagogicheskikh izmerenii"], *Pedagogicheskie izmereniya – Pedagogical Measurements*, Vol. 2, P. 75–88.
16. Dykhovichnyi, A. A., Dudko, A. F. (2013), "Automated system for computer testing in higher mathematics results analysis" ["Avtomatizirovannaya sistema analiza rezul'tatov komp'yuternogo testirovaniya po vysshei matematike"], *Naukovi pratsi DonNTU. Seriya: "Pedagogika, psikhologiya i sotsiologiya"*, Vol. 2 (14), P. 103–110.
17. Suparman, U. (2016), "Analyzing the quality of english test items of daily, mid semester and final school examinations in bandar lampung: (assessment and evaluation in language teaching)", *International Conference On Educational Research & Evaluation, Yogyakarta*, 14 p. URL: <http://repository.lppm.unila.ac.id/id/eprint/3121> (last accessed 11.02.2020).
18. Quagrain, K., Arhin, A. K. (2017), "Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation", *Cogent Education*, Vol. 4 (1). P. 1–11. DOI: <https://doi.org/10.1080/2331186X.2017.1301013>.

Надійшло (Received) 26.02.2020

Відомості про авторів / Сведения об авторах / About the Authors

Смідович Леонід Сергійович – кандидат технічних наук, доцент, Національний аерокосмічний університет ім. М. С. Жуковського "ХАІ", доцент кафедри комп'ютерних наук та інформаційних технологій, Харків, Україна; email: lsonlinals@gmail.com; ORCID: <https://orcid.org/0000-0001-6156-9506>.

Смидович Леонид Сергеевич – кандидат технических наук, доцент, Национальный аэрокосмический университет им. Н. Е. Жуковского "ХАИ", доцент кафедры компьютерных наук и информационных технологий, Харьков, Украина.

Smidovych Leonid – PhD (Engineering Sciences), Associate Professor, National Aerospace University "Kharkiv Aviation Institute", Associate Professor of the Department of Computer Science and Information Technologies, Kharkiv, Ukraine.

Артиух Олеся Володимирівна – Харківський радіотехнічний коледж, викладач іноземної мови, Харків, Україна; email: olesia.artiuh@gmail.com; ORCID: <https://orcid.org/0000-0002-4929-7541>.

Артиух Олеся Владимировна – Харьковский радиотехнический колледж, преподаватель иностранного языка, Харьков, Украина.

Artiukh Olesia – Kharkiv School of Radio Engineering, Foreign Language Teacher, Kharkiv, Ukraine.

Косенко Наталія Вікторівна – кандидат технічних наук, Харківський національний університет міського господарства імені О. М. Бекетова, доцент кафедри управління проектами у міському господарстві і будівництві, Харків, Україна; email: kosnatalja@gmail.com; ORCID ID: <https://orcid.org/0000-0002-5942-3150>.

Косенко Наталья Викторовна – кандидат технических наук, Харьковский национальный университет городского хозяйства имени А. Н. Бекетова, доцент кафедры управления проектами в городском хозяйстве и строительстве, Харьков, Украина.

Kosenko Nataliia – PhD (Engineering Sciences), O. M. Beketov Kharkiv National University of Urban Economy, Associate Professor at the Department of Project Management in Urban Economy and Construction, Kharkiv, Ukraine.

АВТОМАТИЗАЦИЯ АНАЛИЗА КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ И РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ СТУДЕНТОВ ИТ-СПЕЦИАЛЬНОСТЕЙ

Предметом исследования статьи являются статистические модели и методы оценки качества тестовых заданий, а также процессы и технологии информационной поддержки оценивания результатов тестирования и качества тестов. **Целью** является повышение качества тестовых заданий и точности оценки результатов тестирования в условиях постоянных изменений содержания учебных программ и содержания тестовых заданий, и сокращение времени на обработку результатов тестирования за счет автоматизации, применение разработанных алгоритмов и информационной системы. В работе решаются следующие **задачи**: анализ процесса подготовки и проведения тестирования; проведение аналитического обзора статистических методов оценки качества тестовых заданий и показателей качества тестовых заданий; разработка обобщенного процесса подготовки и проведения тестирования; автоматизация анализа результатов тестирования студентов ИТ-специальностей. Для решения указанных задач были использованы **методы** статистического анализа, модель Rasch IRT (Item Response Theory), технологии разработки программных приложений. Получены следующие **результаты**. Проанализирован процесс подготовки и проведения тестирования, предложено его модификацию с учетом особенностей тестирования студентов ИТ-специальностей. Выбраны показатели качества тестовых заданий, модели и методы их расчета, в частности модель Rasch IRT. Приведена интерпретация значений показателей качества тестовых заданий, и расчет коэффициентов аномалии качества тестового задания, которые являются основой для формирования советующего решения по улучшению качества теста. Разработан прототип информационной системы поддержки процесса анализа качества тестовых заданий, оценки результатов тестирования и принятия решений о модификации теста. Проведено тестирование системы при оценке результатов модульного контроля студентов третьего курса. **Выводы**: автоматизация и информационная поддержка расчета и анализа показателей качества тестовых заданий и результатов тестирования позволит повысить точность оценки и сократить время на обработку результатов тестирования.

Ключевые слова: оценка качества тестовых заданий; тестовое задание; процесс тестирования; статистические методы; модели IRT; валидность; надежность; дифференцирующая способность.

AUTOMATION OF TEST ITEMS QUALITY ANALYSIS AND IT STUDENTS TEST RESULTS ASSESSMENT

The **subject** of this article is statistical models and methods for test items quality assessment, as well as processes and information support technologies for test results and test quality evaluating. The **goal** is to improve the quality of test items and the accuracy of the evaluation of test results in the context of constant changes in the content of study programs and the content of test tasks, and to reduce the time for test results processing through automation, and developed algorithms and an information system usage. The following **tasks** was solved: analysis of the test preparation and testing process; an analytical review of statistical methods for the test items quality assessment and the test items quality indicators; generalized process of the test preparation and conduct of testing development; and automation of IT students test results assessment. To solve these problems, the following **methods** was used: statistical analysis methods, the Rasch IRT (Item Response Theory) model, and software application development technologies were used. The following **results** are obtained. The process of the test preparation and testing was analyzed; its modification is proposed taking into account the peculiarities of IT students testing. The test task quality indicators, models and methods for it calculation were selected, in particular, the Rasch IRT model. The interpretation of the test item quality indicators values, and the calculation of the coefficients of the anomalies in the test item quality, what are the basis for the generation of an advisory solution to improve the test

quality, are given. A prototype of an information system for support of processes of test items quality indicators evaluating, test results assessment and decision making about test modification has been developed. The system was tested during the module control of third-year students results evaluating. **Conclusions:** automation and information support of the test items quality indicators calculating and analyzing, and test results assessment will improve the accuracy of the assessment and reduce the time for test results evaluating.

Keywords: statistic methods; IRT models; Rasch model; test quality; test items; knowledge testing; validity; reliability; discriminant ability.

Бібліографічні описи / Bibliographic descriptions

Смідович Л. С., Артюх О. В., Косенко Н. В. Автоматизація аналізу якості тестових завдань та результатів тестування студентів ІТ-спеціальностей. *Сучасний стан наукових досліджень та технологій в промисловості*. 2020. № 1 (11). С. 19–27. DOI: <https://doi.org/10.30837/2522-9818.2020.11.019>.

Smidovych, L., Artiukh, O., Kosenko, N. (2020), "Automation of test items quality analysis and IT students test results assessment", *Innovative Technologies and Scientific Solutions for Industries*, No. 1 (11), P. 19–27. DOI: <https://doi.org/10.30837/2522-9818.2020.11.019>.
