

УДК 303.714, 681.5.01

DOI: <http://dx.doi.org/10.20535/2219-380415201686314>**В. В. Шведова¹**, доцент, к. т. н.

ОЦІНЮВАННЯ РОЗРІЗНЯЛЬНОЇ ЗДАТНОСТІ ТЕСТОВОГО ПРОСТОРУ КОМП'ЮТЕРИЗОВАНОЇ СИСТЕМИ ТЕСТУВАННЯ

En

The concept of information education involves the use of new learning technologies and leads to the development of new forms of knowledge using modern computer training aids. However, nowadays the simple computerization and informatization are insufficient to ensure the quality and effectiveness of the educational process. Learning outcomes, defined, in particular, as the level of students training, are considered as some measurable value for assessment which should involve scientifically based methods, and the values are accompanied by an assessment of their accuracy. The quality of the learning measurement is directly determined by the quality of the tools that were involved in the study and evaluation of the training level. Nowadays, the computerized testing system (CTS) is considered to be the most effective and objective means of the learning assessment, as evidenced by their prevalence in various fields of learning validation. At the same time, only the CTS that was scientifically justified under the development and testing quality indicators can give objective characteristics of the students training levels. A lot of publications are devoted to the issues such as the selection list of indicators that should characterize the CTS quality, assessment methods, as well as their practical value. However, despite the diversity of opinions and approaches the general concept of these issues has not been created. According to the author computerized testing system quality is laid on the stages of project development and CTS strategic calculation and verified on the stage of a CTS pilot experiment and its monitoring. This control verifies the CTS quality score by criteria reliability and validity of the test objectives and test difficulty and distinctiveness test tasks - based on statistics - test results. The evaluation of these indicators should be made using the developed information-measuring system that implements the task of collecting statistical data, their processing, combination, check with the established norms, issuance of reasonable recommendations to improve CTS.

The article deals with the existing and proposed by the author evaluating methods of one of the computerized test quality indicator - indicator of the distinctive character test space problems. The indicator of the distinctive character test and CTS space test tasks is particularly important by applying the legal-oriented tests used in the classroom. Using the criterion-oriented tests (in the final control) the figure is not always decisive. However, by creating the criterion-oriented tests distinctiveness test objectives should be taken into account at the design stage of the test space, because it determines the sensitivity of the testing procedure. Information about the importance of the distinctive character can increase the sensitivity of the testing procedures on a range of scales investigated the students properties (the level of students preparedness) by choosing the appropriate test tasks. Therefore, along with the test difficulty index the distinctiveness tasks can be used for the CTS testing in adaptive mode.

¹ Національний технічний університет України «Київський політехнічний інститут», кафедра інформаційно виміральної техніки

The article examined, systematized and proposed methods for evaluating the distinctive character of CTS test space tasks, submitted recommendations on the choice of one or another metrics that improve the efficiency of information-measuring system for CTS quality control indicators.

Ru

Подтверждение качества проведения и результатов учебного процесса в значительной мере обуславливается качественными показателями инструментария, при помощи которого это качество оценивается. Поэтому создание алгоритмического та информационно-измерительного обеспечения для оценивания показателей качества компьютеризированных систем тестирования является неотъемлемой частью программ информатизации образовательной отрасли. В статье рассмотрены, систематизированы и предложены методы и способы оценивания одного из наиболее важных показателей качества компьютеризированной системы тестирования (КСТ) – различительной способности заданий тестового пространства КСТ, представлены рекомендации относительно выбора той или иной метрики, что позволяет повысить эффективность информационно-измерительных систем контролю показателей качества КСТ.

Вступ

Основні засади сучасної освіти такі як доступність та наступність процесу здобуття освіти, інтеграція системи освіти України у світову систему, що прописані в Законах України «Про освіту» та «Про вищу освіту» [1] – [2] неможливі без впровадження сучасних засобів подання знань. Концепція розвитку та інформатизації освіти передбачає використання новітніх прогресивних технологій навчання і спонукає до розвитку нових форм подання знань з використанням сучасних комп'ютерних засобів навчання. За визначенням [3], комп'ютеризований засіб навчання – програмний засіб (програмний комплекс) або програмно-технічний комплекс, призначений для розв'язування визначених педагогічних задач, що має предметний зміст та орієнтований на взаємодію з учнем. Однак в наш час простоті комп'ютеризації та інформатизації недостатньо для забезпечення якості та ефективності освітнього процесу. Результат навчання, який визначається, зокрема, за рівнем навченості студентів розглядається як деяка вимірювана величина, для оцінювання якої мають бути залучені науково обґрунтовані методи, а отримані значення мають супроводжуватись оцінкою їх точності. При цьому якість вимірювання рівня навченості напряму визначається якістю того інструментарію, який було залучено під час навчання та оцінювання самого рівня навченості.

Найбільш ефективними і об'єктивними засобами оцінювання рівня навченості в наш час вважаються комп'ютеризовані системи тестування (КСТ), про що свідчить їх поширеність в різних галузях перевірки результатів навчання. В той же час лише КСТ, що пройшли науково обґрунтовані стадії розробки та перевірки якісних показників [4], можуть дати об'єктивні характеристики щодо рівнів навченості студентів.

Вирішенню таких питань як вибір переліку показників, якими повинна характеризуватись якість КСТ, способів їх оцінювання, а також визначення їх практичної цінності присвячено досить багато публікацій [5] - [11]. Однак не дивлячись на різноманіття думок і підходів загальної концепції щодо названих питань так і не створено.

На думку автора якість комп'ютеризованої системи тестування закладається на стадіях розробки проекту та стратегічного розрахунку КСТ, а перевіряється на стадії пілотного експерименту КСТ та її моніторингу [12].

При цьому контролю підлягають такі показники якості КСТ як надійність та критеріальна валідність тесту та завдань тесту, трудність та розрізняльна здатність завдань тесту – на основі статистичних даних – результатів тестування. При цьому оцінювання даних показників має здійснюватись за допомогою розробленої інформаційно-вимірювальної системи, яка реалізує задачі від збору статистичних даних, їх опрацювання, поєднання, перевірки із встановленими нормами до видачі обґрунтованих рекомендацій щодо покращення КСТ.

Постановка задачі

В статті розглядаються вже існуючі та запропоновані автором способи оцінювання одного з показників якості систем комп'ютеризованого тестування – показника розрізняльної здатності задач тестового простору.

Огляд існуючих рішень постановки задачі

Розрізняльна здатність – показник якості комп'ютеризованої системи тестування, що характеризує чутливість процедури тестування до рівня підготовленості учнів. Даний показник в літературі висвітлюється в двох аспектах [6], [8], [13]. Перший пов'язаний з тим, що розрізняльну здатність розглядають як характеристику кожного завдання тесту. Такий підхід використовується як в класичній теорії обробки результатів тестування, так і в теорії моделювання та параметризації тестів (ТМПТ). Другий підхід пов'язаний з оцінюванням розрізняльної здатності цілого тесту. За цим підходом в рамках класичної теорії розглядають емпіричний розподіл результатів тестування і висновки щодо розрізняльної здатності тесту отримують на основі аналізу форми кривої розподілу. За ТМПТ розрізняльна здатність тесту трактується як ціна поділки на шкалі латентних параметрів.

Актуальність поставленої задачі

Показник розрізняльної здатності тесту та завдань тестового простору КСТ має особливе значення при використанні нормативно-орієнтованих

тестів, яких є більшість з тих, що використовуються в навчальному процесі. При використанні критеріально-орієнтованих тестів (при підсумковому контролі) даний показник не завжди є визначальним. Однак і при створенні критеріально-орієнтованих тестів розрізняльна здатність завдань тесту повинна враховуватись на етапі розробки тестового простору, оскільки вона визначає чутливість процедури тестування. Інформація про значення розрізняльної здатності дозволяє збільшити чутливість процедури тестування на певних діапазонах шкали досліджуваної властивості учнів (рівень підготовленості учнів) за рахунок підбору відповідних завдань тесту. Тому поряд з показником трудності завдань тесту розрізняльна здатність може бути використана при реалізації тестування в КСТ в адаптивному режимі.

Вибір методів, підходів, моделей та інструментів розв'язання поставленої задачі

Для оцінювання розрізняльної здатності в рамках класичного підходу [14] використовують способи оцінювання, що ґрунтуються на аналізі узгодженості (або контрастності) експериментальних даних, отриманих за результатами процедури тестування. При цьому розрізняльна здатність може бути оцінена як за всією вибіркою опитуваних, так і за її частинами.

Оцінювати розрізняльну здатність завдань тестового простору КСТ в рамках класичної теорії обробки результатів тестування, на думку автора, можна з використанням наступних способів (метрик):

1. За коефіцієнтами зв'язку, пов'язаними з критерієм Пірсона χ^2 (коефіцієнт Чупрова, коефіцієнт Крамера);
2. За коефіцієнтами асоціації (коефіцієнт Юла), контингенції та колігації;
3. За долею опитуваних, що виконали завдання в контрастних групах.
4. За ранговою бісеріальною кореляцією;
5. За «альтернативними» мірами зв'язку двох груп експериментальних даних.

В табл. 1 автором систематизовані зазначені способи оцінювання розрізняльної здатності тестового простору за обмеженнями їх застосування.

Таблиця 1.

Способи оцінювання показника розрізняльної здатності тестового простору КСТ

Способи оцінювання	Обмеження за об'ємом вибірки	Обмеження за видом розподілу дослідних даних
За долею опитуваних, що виконали завдання в контрастних групах	Більше 300 осіб	Розподіл первинних балів близький до нормального

Способи оцінювання	Обмеження за об'ємом вибірки	Обмеження за видом розподілу дослідних даних
За коефіцієнтами зв'язку, пов'язаними з критерієм Пірсона χ^2 (коефіцієнт Чупрова, коефіцієнт Крамера)	Більше 100 осіб	Обмежень немає
За ранговим коефіцієнтом кореляції		
За коефіцієнтами асоціації (коефіцієнт Юла), контингенції та колігації	Більше 20 осіб, але не менше 5 осіб в кожній клітинці кореляційної матриці	Обмежень немає
За «альтернативними» мірами зв'язку (коефіцієнти Гудмана, близькості розбивок, Сомерса)		

Метрики згаданих в табл. 1 мір зв'язку є адекватними для оперування з величинами (в контексті дослідження – це первинні бали опитуваних та завдань тесту), що відображаються шкалою порядку. Потрібно зазначити, що метод контрастних груп є традиційним для дослідження тестового простору, однак, як бачимо в таблиці, він накладає значні обмеження, зокрема щодо об'єму досліджуваної вибірки та виду її розподілу. Тому за умов малих вибірок, що розглядаються в роботі, даний метод не може бути застосований. Визначення показника розрізняльної здатності за допомогою інших метрик не вимагає значних обсягів дослідних даних. Поряд з цим точність оцінювання показника може бути підвищена шляхом накопичення та спільної обробки дослідних даних.

Оцінювання розрізняльної здатності завдання тесту за долею опитуваних, які виконали завдання в контрастних групах

Розрізняльна здатність оцінюється за формулою [6], [15]:

$$D = n_{n_max} \cdot n_{max}^{-1} - n_{n_min} \cdot n_{min}^{-1}, \quad (1)$$

де n_{n_max} , n_{n_min} – кількість опитаних, які правильно відповіли на завдання в «слабкій» та «сильній» групах;

n_{min} , n_{max} – чисельність групи «слабких» та «сильних» опитуваних відповідно.

Оцінювання розрізняльної здатності завдання тесту за ранговим коефіцієнтом кореляції

При оцінюванні розрізняльної здатності на основі всієї вибірки можна використати ранговий бісеріальний коефіцієнт [16]:

$$r_{bis} = 2 \cdot n^{-1} \cdot (\bar{r}_B - \bar{r}_H), \quad (2)$$

де \bar{r}_B, \bar{r}_H – середні ранги серед первинних балів опитуваних, які правильно та неправильно відповіли на завдання тесту;

n – кількість опитуваних.

З урахування того, що ранги в (2) є номерами елементів досліджуваного ряду і являють собою елементи порядкової шкали, в якості центральних тенденцій обрано медіани ряду.

$$r_{bis} = 2 \cdot n^{-1} \cdot (\text{med}\{r_B\} - \text{med}\{r_H\}) \frac{2}{N}, \quad (3)$$

де $\text{med}\{r_B\}, \text{med}\{r_H\}$ – медіани рангів первинних балів опитаних, які правильно та неправильно відповіли на завдання тесту.

Коефіцієнти D та r_{bis} набувають значень від -1 до 1: -1 – зв'язок повний зворотній; 0 – відсутність зв'язку; 1 – повний прямий зв'язок.

Оцінювання розрізняльної здатності завдання тесту за коефіцієнтами зв'язку, пов'язаними з критерієм Пірсона χ^2 (коефіцієнт Чупрова, коефіцієнт Крамера)

Для розрахунку цих метрик первинні бали опитуваних, які відповідали на дане завдання, розділяють:

- на три частини (якщо опитуваних більше 300 осіб): «слабкі» – з найменшими первинними балами за тест, «середні» та «сильні» – з найбільшими первинними балами за тест. Чисельність груп має становити 27 – 33 % від всієї вибірки опитуваних, яка відповідала на дане завдання тесту.
- на дві частини (якщо опитуваних від 100 до 300 осіб) – по 50 % кожна.

Далі заповнюють кореляційну таблицю вигляду 2. В таблиці позначено: N_{ij} – частоти потрапляння у відповідну групу, $i = \overline{1, k}$, $j = \overline{1, l}$,

$N(y_j) = \sum_{i=1}^k N_{ij}$, $N(x_i) = \sum_{j=1}^l N_{ij}$, N – загальна кількість опитаних, що відповідали на завдання тесту.

Таблиця 2.

	Дали правильну відповідь на завдання	Дали неправильну відповідь на завдання	Маргінальні суми, $N(x_i)$
Підгрупа «сильних» опитуваних	N_{11}	N_{12}	$N(x_1)$
Підгрупа «середніх» опитуваних	N_{21}	N_{22}	$N(x_2)$

Розділ 1. Інформаційні системи

	Дали правильну відповідь на завдання	Дали неправильну відповідь на завдання	Маргінальні суми, $N(x_i)$
Підгрупа «слабких» опитуваних	N_{31}	N_{32}	$N(x_3)$
Маргінальні суми, $N(y_j)$	$N(y_1)$	$N(y_2)$	N

Коефіцієнт Крамера можна розрахувати для не прямокутної кореляційної таблиці: $k \neq l$, k - кількість рядків, l - кількість стовпчиків за формулою [16]:

$$T_{cr} = \sqrt{\frac{\chi^2}{N\sqrt{(k-1)(l-1)}}} : \sqrt[4]{\frac{\min(k-1, l-1)}{\max(k-1, l-1)}}, \quad (4)$$

де χ^2 визначається за однією з відомих формул, зокрема:

$$\chi^2 = N \left(\sum_{i=1}^k \sum_{j=1}^l N_{ij}^2 \cdot [N(x_i) \cdot N(y_j)]^{-1} - 1 \right). \quad (5)$$

Для випадку $k=3$, $l=2$ (табл. 2) маємо: $T_{cr} = \sqrt{\chi^2 \cdot N^{-1}}$.

Коефіцієнт Чупрова розраховується за формулою [16]:

$$T_{ch} = \sqrt{\chi^2 \cdot [N\sqrt{(k-1)(l-1)}]^{-1}}. \quad (6)$$

Коефіцієнт Чупрова нормований до 1 лише у випадку прямокутної кореляційної таблиці: $k=l$. При $k=2$, $l=2$ (табл. 2) маємо: $T_{ch} = \sqrt{\chi^2 \cdot N^{-1}}$.

Наведені коефіцієнти набувають значень від 0 до 1 в напрямку посилення зв'язку. Значущість коефіцієнтів T_{cr} та T_{ch} визначається за критерієм χ^2 . Тому попередньо потрібно провести порівняння χ^2 та $\chi_{кр}^2$ при заданому рівні значущості q та числі степенів свободи $f = (k-1)(l-1)$. Якщо $\chi^2 > \chi_{кр}^2$, то кореляційний зв'язок двох факторів суттєвий, а його сила відображається коефіцієнтами Крамера та Чупрова відповідно.

Визначення розрізняльної здатності завдання тесту за коефіцієнтами асоціації (коефіцієнт Юла), контенгенації та колігації

Названі коефіцієнти визначають тип та «силу» зв'язку на основі значень частот, що потрапляють у відповідні клітинки кореляційної таблиці виду 2.

Коефіцієнт Юла оцінюють наступним чином [16] - [17]:

$$Q = (N_{11}N_{22} - N_{12}N_{21}) \cdot (N_{11}N_{22} + N_{12}N_{21})^{-1}. \quad (7)$$

Коефіцієнт Юла набуває значень від -1 до 1: -1 – зв'язок повний зворотній; 0 – відсутність зв'язку; 1 – повний прямий зв'язок.

Коефіцієнт Юла є показником одностороннього зв'язку. В нашому випадку приналежність до певної групи («сильних» чи «слабких» студентів) є первинним, а відповідь на завдання (правильна чи неправильна) залежним фактором.

Коефіцієнт контингенції оцінюють наступним чином [16] – [17]:

$$\Phi = (N_{11}N_{22} - N_{12}N_{21}) \cdot [N(x_1)N(x_2)N(y_1)N(y_2)]^{-0.5}, \quad (8)$$

або для таблиці 2×2

$$\Phi^2 = \chi^2 \cdot N^{-1}. \quad (9)$$

Розширення коефіцієнту контингенції для кореляційних таблиць розмірністю більше ніж 2×2 , зокрема для таблиці виду табл. 2, дає альтернативну метрику для визначення середньої квадратичної спряженості ϕ^2 за формулою [16] – [17]:

$$\phi^2 = \chi^2 \cdot N^{-1}. \quad (10)$$

Тобто ϕ^2 є узагальненням Φ^2 .

Середня квадратична спряженість є мірою розходження, що припадає на одне спостереження. Оскільки значення $0 < \phi^2 < \infty$, то визначають коефіцієнт середньої квадратичної спряженості Пірсона:

$$C = \sqrt{\phi^2 \cdot (1 + \phi^2)^{-1}}. \quad (11)$$

Оскільки C має значення від 0, але не досягає 1, то проводять нормування:

$$C' = \sqrt{\phi^2 \cdot (1 + \phi^2)^{-1} \cdot [1 + \min(k-1, l-1)] \cdot [\min(k-1, l-1)]^{-1}}. \quad (12)$$

Для випадку $k = 3, l = 2$ маємо:

$$C' = \sqrt{2 \cdot \phi^2 \cdot (1 + \phi^2)^{-1}} = \sqrt{2 \cdot \chi^2 \cdot (N + \chi^2)^{-1}}.$$

Коефіцієнт контингенції набуває значень від -1 до 1. Максимальне абсолютне значення коефіцієнту контингенції для різних кореляційних таблиць різне, але не перевищує 1 по модулю. Значення близькі до -1 свідчать про повний зворотній зв'язок; 0 – відсутність зв'язку; близькі до 1 – про повний прямий зв'язок.

Коефіцієнт контингенції є показником двостороннього зв'язку. В нашому випадку приналежність до певної групи («сильних» чи «слабких» студентів) може бути первинним, а відповідь на завдання (правильна чи

неправильна) залежним фактором. Зворотнє твердження також буде справедливим.

Для оцінювання міри зв'язку оцінюють також коефіцієнт колігації [17]:

$$Y = \left[(N_{11}N_{22})^{1/2} - (N_{12}N_{21})^{1/2} \right] \cdot \left[(N_{11}N_{22})^{1/2} + (N_{12}N_{21})^{1/2} \right]^{-1}. \quad (13)$$

Оцінювання розрізняльної здатності завдання тесту за «альтернативними» мірами зв'язку

Для оцінювання розрізняльної здатності тестового простру КСТ можна використати «альтернативні» міри зв'язку, які у статистичних дослідженнях використовуються відносно рідко [16]. Такими мірами є наступні.

G-коефіцієнт Гудмана використовується для ординальних шкал і характеризує односторонній зв'язок.

$$G = \left[\sum_{i=1}^k \max N_{ij} - \max N(y_j) \right] \cdot \left[N - \max N(y_j) \right]^{-1}. \quad (14)$$

Коефіцієнт G набуває значень від 0 до 1: 0 – відсутність зв'язку; 1 - повний прямиий зв'язок.

γ – коефіцієнт Гудмана використовується для ординальних шкал і характеризує двосторонній зв'язок.

$$\gamma = (P - Q) \cdot (P + Q)^{-1}, \quad (15)$$

де P - число пар об'єктів, у яких обидві ознаки упорядковані в однаковій послідовності:

$$P = \sum_{i=1}^k \sum_{j=1}^l N_{ij} \left(\sum_{r=i+1}^k \sum_{s=j+1}^l N_{rs} \right), \quad (16)$$

Q - сума результатів суми добутоків частот кожної клітинки на суму частот, розташованих нижче і лівіше:

$$Q = \sum_{i=1}^k \sum_{j=1}^l N_{ij} \left(\sum_{r=i+1}^k \sum_{s=1}^{j-1} N_{rs} \right). \quad (17)$$

d - коефіцієнт Сомерса.

$$d = (P - Q) \cdot (p + Q + Y_0), \quad (18)$$

де P та Q визначають за формулами (16) та (17),

Y_0 - число пар об'єктів з однаковим значенням по одному критерію (але різним значенням по іншому критерію).

Коефіцієнт γ набуває значень від -1 до 1: -1 – зв'язок повний зворотній; 0 – відсутність зв'язку; 1 – повний прямий зв'язок.

Коефіцієнт близькості розбивок:

$$\delta_{xy} = \frac{1}{N(N-1)} \left[\sum_{i=1}^k N^2(x_i) + \sum_{j=1}^l N^2(y_j) - 2 \sum_{i=1}^k \sum_{j=1}^l N_{ij}^2 \right], \quad (19)$$

Коефіцієнт близькості розбивок набуває значень від 0 до 1: 1 – відсутність зв'язку; 0 – повний прямий зв'язок.

Δ - коефіцієнт, що використовується для ординальних шкал і характеризує односторонній зв'язок:

$$\Delta = k \cdot [2(m-1)(2k-m)]^{-1} \sum_{i=1}^k \sum_{j=1}^l |n_{ij} - \bar{n}_j|, \quad (20)$$

де $m = \min(k, l)$, $n_{ij} = N_{ij} \cdot N^{-1}_{ij}$, $\bar{n}_j = l^{-1} \sum_{p=1}^l n_{jp}$.

Для випадку таблиці 2×2 маємо: $\Delta = 0.5 \sum_{i=1}^k \sum_{j=1}^l |n_{ij} - \bar{n}_j|$. Δ - коефіцієнт набуває значень від 0 до 1: 0 – відсутність зв'язку; 1 – повний прямий зв'язок.

Приклад оцінювання описаних характеристик для пілотної вибірки учнів

Розрахунки досліджуваних коефіцієнтів проілюстровані в таблиці 3 на прикладі визначення розрізняльної здатності 10 завдань, на які відповідали 20 учнів.

Всі метрики мір зв'язку, розглянуті в розділі, є адекватними для оперування з величинами (в контексті дослідження – це первинні бали опитуваних та завдань тесту), що відображаються шкалою порядку. В табл. 3 завдання розміщені в порядку зростання розрізняльної здатності. З таблиці видно, що рангові місця завдань для кожного коефіцієнта зберігаються.

Таблиця 3.

Показники розрізняльної здатності

	Номер завдання									
	6	1	4	3	9	7	2	8	10	5
Коефіцієнт Юла	0,00	0,20	0,40	0,56	0,60	0,69	0,71	0,71	0,71	0,81
Коефіцієнт контентації	0,00	0,10	0,20	0,30	0,31	0,40	0,41	0,41	0,41	0,50

Розділ 1. Інформаційні системи

	Номер завдання									
	6	1	4	3	9	7	2	8	10	5
Коефіцієнт колігації	0,00	0,10	0,21	0,30	0,33	0,40	0,42	0,42	0,42	0,51
Показник за методом контрастних груп	0	0,1	0,2	0,3	0,3	0,4	0,4	0,4	0,4	0,5
Показник за ранговою бісеріальною кореляцією	0,2	0,4	0,6	0,4	0,4	0,8	0,8	0,8	0,8	0,8
G - коефіцієнт Гудмана	0	0	0	0,22	0	0,4	0,25	0,25	0,25	0,44
γ - коефіцієнт Гудмана	0,33	0,46	0,58	0,69	0,7	0,77	0,8	0,8	0,8	0,86
d - коефіцієнт Сомерса	0,33	0,46	0,58	0,69	0,7	0,77	0,8	0,8	0,8	0,86
Коефіцієнт близькості розбивок	0,47	0,48	0,49	0,52	0,52	0,56	0,56	0,56	0,56	0,61
Δ - коефіцієнт	0	0,1	0,2	0,3	0,3	0,4	0,4	0,4	0,4	0,8

Показник розрізняльної здатності, визначений в рамках класичної теорії обробки результатів тестування, має бути більшим від 0,3.

В ТМПТ розрізняльна здатність розглядається в двох аспектах:

- розрізняльна здатність завдань тесту, як параметр двохпараметричної моделі Бірнбаума;
- розрізняльна здатність тесту в цілому, як ціна поділки на шкалі рівнів підготовленості.

Аналітичний вираз функції успіху $p(\theta - \delta, d)$ за двохпараметричною моделлю має вигляд [8]:

$$p(\theta - \delta, d) = \left[1 + e^{-d(\theta - \delta)} \right]^{-1}, \quad (21)$$

де параметр d - характеризує розрізняльну здатність завдання тесту.

Графічно параметр розрізняльної здатності відображається нахилом дотичної до функції успіху в точці перегину. Даний параметр може набувати значень від -1 до 1. Параметр оцінюється у відносних одиницях і не повинен бути меншим за 0. Рекомендацій щодо оцінювання даного параметра в рамках ТМПТ не виявлено. Зважаючи на статистичні характеристики показника, автор вважає можливим для його оцінювання використати рекомендації, наведені вище, зокрема, коефіцієнт Юла.

За ТМПТ розрізняльна здатність тесту в цілому трактується як ціна поділки на шкалі латентних параметрів. Числова характеристика показника розрізняльної здатності тесту, визначеного в рамках ТМПТ, залежить від

кількості завдань в тесті і не може бути меншою від $P3T = |\theta(b_k) - \theta(b_{k-1})|$ для двох сусідніх значень рівня підготовленості на шкалі логітів в даному діапазоні. Таким чином властивість – розрізняльна здатність – за ТМПТ визначається в логітах і має нелінійний характер. Числова характеристика показника розрізняльної здатності тесту, визначеного в рамках ТМПТ, може набувати значень в діапазоні $4 \cdot k^{-1} < P3T < +\infty$ логіт, хоча на практиці, як правило, обмежується діапазоном $4 \cdot k^{-1} < P3T < 11 \cdot k^{-1}$ логіт (k - кількість завдань в тесті). Менші значення показника свідчать про зростання якості тесту, як засобу педагогічного вимірювання.

Аналіз та пропозиції щодо вибору характеристик для оцінювання розрізняльної здатності КСТ

Аналіз розглянутих метрик дозволив вибрати ті з них, що можуть бути використані для калібрування завдань тестового простору під час пілотного експерименту та дослідної експлуатації КСТ. Розрізняльна здатність, яка визначається за методом контрастних груп вимагає значних обсягів дослідних даних (більше 300 осіб опитаних) та накладає обмеження на вид їх розподілу (розподіл має бути нормальним). В умовах пілотного експерименту КСТ з малими контингентами учнів цього виконати не можливо. Тому даний метод може бути використаний при накопиченні значних обсягів дослідних даних під час моніторингу КСТ. Метрики, що пов'язані з критерієм Пірсона, та ранговий бісеріальний коефіцієнт також вимагають наявності значних обсягів дослідних даних (хоча і менших ніж в попередньому) і не завжди можуть бути використані при пілотному експерименту КСТ в зазначених умовах.

Найкращими метриками для оцінювання розрізняльної здатності завдань тестового простору КСТ, на думку автора, є група метрик – коефіцієнт асоціації, контенгенації та колігації. Відмінність між коефіцієнтами асоціації (Юла), контенгенації та колігації полягає в наступному. Коефіцієнти асоціації та колігації мають різну чутливість в різних точках шкали. Коефіцієнт контенгенації, на відміну від коефіцієнтів асоціації та колігації, що характеризують односторонній зв'язок, характеризує двосторонній зв'язок. Оскільки емпіричні дані (первинні бали опитуваних) відображаються за шкалою порядку, то вибір за характеристикою чутливості в різних точках шкали не є визначальним. В той же час при дослідженні розрізняльної здатності є важливим односторонній зв'язок. Тому для розрахунку розрізняльної здатності завдань тестового простору КСТ під час пілотного експерименту та дослідної експлуатації обрано коефіцієнт Юла. Ще однією суттєвою перевагою метрики – коефіцієнта Юла є можливість розрахунку невизначеності оцінки цього показника, чого позбавлені «альтернативні» міри зв'язку. Тому коефіцієнт Юла варто обирати для оцінювання розріз-

няльної здатності тестового простору КСТ як при пілотному експерименті, так і в процесі експлуатації систем.

Висновки

Підтвердження якості проведення та результатів освітнього процесу значною мірою обумовлюється якісними показниками інструментарію, за допомогою якого ця якість оцінюється. Тому створення алгоритмічного та інформаційно-вимірювального забезпечення для оцінювання показників якості комп'ютеризованих систем тестування є невід'ємною частиною програми інформатизації освітньої галузі. Розглянуті, систематизовані та запропоновані в статті методи та способи оцінювання одного з важливих показників якості КСТ – розрізняльної здатності – свідчать про важливість чіткого зіставлення цільової спрямованості КТС та показника, який буде використано для оцінювання розрізняльної здатності тестового простору КСТ. В статті наведено результати дослідження та рекомендації, які допоможуть розробнику інформаційно-вимірювальної системи контролю показників якості КСТ у виборі найбільш адекватної для конкретної ситуації оцінки. Як перспективу наступних досліджень, потрібно зазначити необхідність визначення метрик для оцінювання точності показника розрізняльної здатності тестових просторів КСТ.

Список використаної літератури

1. Закон України «Про вищу освіту» від 01.06.14 № 1556-VII [Електронний ресурс]. – Режим доступу: <http://zakon0.rada.gov.ua/laws/show/1556-18>.
2. *Гуржій А. М.* Електронні освітні ресурси як основа сучасного навчального середовища загальноосвітніх навчальних закладів / А. М. Гуржій, В. В. Лапінський [Електронний ресурс] // Інформаційні технології в освіті. – 2013. – № 15. – С. 30–37. – Режим доступу: <http://ite.kspu.edu/issue-15/> р.30-37.
3. *Башмаков А. И.* Разработка компьютерных учебников и обучающих систем / А. Башмаков, И. Башмаков - М.: - Информационно-издательский дом «Филинь». - 2003. – 616 с.
4. *Шведова В. В.* Проблеми створення дистанційних курсів та підтримки їх якості в процесі експлуатації / В. В. Шведова // Техніка АПК. – 2008. - №11-12. – С. 55 - 56.
5. *Аванесов В. С.* Композиция тестовых заданий. Учебная книга. 3 изд., доп / В. С. Аванесов.– М.: Центр тестирования, 2002 г. – 240 с.
6. *Бурлачук Л. Ф.* Словарь-справочник по психодиагностике: 2 –е издание переработанное и дополненное / Л. Ф. Бурлачук, С. М. Морозов. - СПб.: Питер, 2004.-250с.

7. *Чельшкова М. Б.* Теория и практика конструирования педагогических тестов: Учебное пособие / М. Б. Чельшкова. - М.: Логос, 2002. – 432 с.
8. *Нейман Ю. М.* Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. – Москва, 2000. –168 с.
9. *Mudur S. P.* A Methodical Assessment of Integrative Model-Based E-Course Development / Mudur S. P., Prachi Gharpure, Parvati Rajan // IEEE Transaction on education - November 2005. - vol. 48, №4. – p. 607-611.
10. *Щербина О. А.* Оцінювання компетентностей засобами платформи Moodle. // Інформаційні технології і засоби навчання – 2015. – Том 45, №1. –С. 134–145. [Електронний ресурс]. –Режим доступу : journal.iitta.gov.ua/index.php/itlt/article/download/1169/885.
11. *Федорук П. І.* Адаптивна система дистанційного навчання та контролю знань на базі інтелектуальних Інтернет-технологій: монографія / П. І. Федорук. – Івано-Франківськ: Видавничо-дизайнерський відділ ЦІТ Прикарпатського національного університету імені Василя Стефаника, 2008. – 326с.
12. *Ціделко В. Д.* Моніторинг систем дистанційного навчання / В. Д. Ціделко, Н. А. Яремчук, В. В. Шведова // Вища освіта України. – 2006. - №2. – С. 54 – 61.
13. *Анастасі А.* Психологическое тестирование. / А. Анастасі, С. Урбина–СПб.: Питер, 2001. – 688 с.
14. *Ціделко В. Д.* Оцінювання якості контролюючих тестів / В. Д. Ціделко, Н. А. Яремчук, В. В. Шведова // Техніка АПК. – 2006. - №4. – С. 34 - 36.
15. *Гласс Дж.* Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стенли; - пер. с англ. - М.:Прогресс, 1976. – 498 с.
16. *Паніотто В. І.* Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Харченко. – К.: Вид. дім «КМ Академія», 2004.-270 с.
17. *Кендал М.* Статистические выводы и связи / М. Кендал, А. Стьюарт; - пер. с англ. - Главная редакция физико-математической литературы изд-ва «Наука», 1973, -502 с.