
УДК 303.714, УДК 681.5.01

В.В. Шведова

Національний технічний університет України «Київський політехнічний інститут», Київ

ОЦІНЮВАННЯ ВАЛІДНОСТІ ТЕСТОВОГО ПРОСТОРУ КОМП'ЮТЕРИЗОВАНОЇ СИСТЕМИ ТЕСТУВАННЯ: ВИДИ ВАЛІДНОСТІ, МЕТОДИ ТА СПОСОБИ ОЦІНЮВАННЯ ПОКАЗНИКА ТА ЙОГО НЕВИЗНАЧЕНОСТІ

Автором систематизовано види валідності та проаналізовано методи і способи оцінювання цього показника якості комп'ютеризованої системи тестування (КСТ); виділено види валідності, що варто оцінювати на різних стадіях життєвого циклу КСТ; обрано метрику для оцінювання валідності КСТ на стадії пілотного експерименту і дослідної експлуатації та проведено їх удосконалення з огляду на особливості шкали, в якій оцінюються вхідні дані для розрахунку цього показника. Запропоновано аналітичні вирази для оцінювання невизначеності показника валідності тестового простору КСТ. Наведено динаміку зміни невизначеності показника валідності в залежності від об'єму опитуваних, що залучені до роботи з КСТ.

Ключові слова: системи дистанційного навчання, комп'ютеризована система тестування, тестовий простір, валідність тестового простору, невизначеність.

Вступ

Дистанційне навчання є альтернативним до очного та заочного навчання засобом забезпечення

навчальних потреб і є досить поширеним в розвинутих країнах світу. На території України дистанційне навчання часто застосовується як допоміжний елемент традиційного очного або заочного

навчання [1], що значно поліпшує комунікацію між викладачем та студентами, дозволяє забезпечити студентів матеріалами для самостійного опрацювання, проводити контрольні заходи в автоматизованому режимі тощо.

За своєю ідеологією дистанційне навчання передбачає мінімізацію участі викладача в початковому процесі. Для реалізації цієї ідеї створюються системи дистанційного навчання (СДН) на базі програмних платформ таких як Lotus Notes, Moodle тощо. В процесі реалізації навчання за допомогою СДН викладач виступає як так званий тьютор, тобто людиною, яка покликана з одного боку допомогти студентові працювати з системою, з іншого боку контролювати хід проходження навчального процесу. При цьому все смислове та методологічне опрацювання, наповнення системи навчальним матеріалом та тестування системи (пілотне експериментальне дослідження) від ідеї до реалізації [2-3] передбачає залучення висококваліфікованих працівників як в сфері галузі навчання, так і спеціалістів тестологів.

Система дистанційного навчання як і будь-який готовий продукт, що може бути використаний, зокрема, з комерційною метою, має пройти перевірку якості і підтвердити свою ефективність.

Не дивлячись на досить значний період часу існування та розвитку різних систем дистанційного навчання [2, 4 – 9] питання оцінювання якості таких систем залишається відкритим, і, як правило, зводиться до використання експертного оцінювання.

В той же час будь-яка СДН передбачається наявність як складової комп'ютеризованої системи тестування (КСТ). Часто КСТ виступають і як автономні допоміжні засоби, які використовуються при традиційних формах навчання, що значно полегшує діяльність викладачів та підвищує ефективність їх праці [10, 11]. Оцінювання якості КСТ часто в значній мірі дозволяє судити про якість і всієї СДН, адже недоліки в поданні навчального матеріалу безумовно відобразяться в результатах оцінювання студентів, хоча причини цих недоліків буває і не так легко ідентифікувати.

Серед кількісних показників якості КСТ, що характеризують якість процедуру тестування та окремих завдань тесту найчастіше згадують наступні [2, 12-15]: надійність – характеризує точність процедури тестування та її інваріантність до випадкових чинників, що впливають або можуть впливати на результат дослідження; валідність – характеризує адекватність процедури тестування поставленій меті; трудність – характеризує адекватність процедури тестування вибірці опитуваних за складністю змістовного матеріалу; розрізняльну здатність – характеризує чутливість процедури тестування до індивідуальних особливостей (відмінностей) учнів.

Оцінювання показника валідності тестового простору КСТ

Валідність є одним з головних показників якості тестового простору КСТ, що характеризує його адекватність поставленій меті тестування. Показник валідності є комплексним та багатоплановим. Поряд з цим потрібно зазначити певну нечіткість в систематизації видів валідності та способів її оцінювання. Найбільш повний та ґрунтовний огляд видів валідності наведено в роботі [13]. Автором статті пропонується систематизувати види валідності за такими критеріями [16]: за способом отримання даних для аналізу (емпірична та експертна валідність); за підходами, що використовують при аналізі валідності.

Систематизація видів валідності проведена на основі зробленого автором аналізу методів оцінювання показника, використовуваного при цьому математичного апарату та шкал вимірювання [17, 18], до яких належить отримуваний показник. Результат зазначеної систематизації представлено у вигляді табл. 1.

З наведеної в табл. 1 інформації видно певна неадекватність використовуваного математичного апарату властивостям величин, що підлягають статистичній обробці. Зокрема недоцільність використання таких мір, як коефіцієнта кореляції Пірсона [19], для оцінювання зв'язків між величинами, що мають властивості шкали порядку [17]. Тому однією з задач дослідження було вибір адекватних способів оцінювання валідності при пілотному експерименті та дослідній експлуатації КСТ.

На основі запропонованого автором поділу життєвого циклу КСТ [3, 20] проведено систематизацію різновидів валідності, що доцільно визначати на різних стадіях життєвого циклу системи – табл. 2.

Експертну валідність (змістовну, очевидну, ілюзорну) отримують на основі експертних оцінок, які виступають критерієм валідації самого тесту. Даний метод не дозволяє отримати беззаперечних і виключно об'єктивних оцінок. Однак він завжди має використовуватися на стадії проектування тестового простору КСТ. До емпіричної валідності автором віднесено конструктну та критеріальну валідності. Конструктна валідність має характер теоретичного дослідження тестового простору КСТ з перевіркою висунутих гіпотез щодо здатності тестів оцінювати ті чи інші конструкти в структурі знань опитуваних, що закладені в неї на стадії розробки.

Визначення конструктної валідності проводиться на стадії розробки тестового простору КСТ та на стадіях, що передують стандартизації КСТ. Через певну складність в дослідженні цього виду валідності тестового простору КСТ часто визначення конструктної валідності пропускають, оцінюючи на попередніх стадіях лише змістовну валідність тестового простору КСТ.

Таблиця 1

Види валідності тестового простору КСТ,
методи та математичний апарат, який використовується для оцінювання показника

Вид валідності		За методом оцінювання	За використаним математичним апаратом	За шкалою вимірювання	Обмеження на застосування метода	
Експертна	Змістовна	Метод експертних оцінок (критерієм валідації при цьому виступає сам тест (або його завдання))	а) визначення колективної оцінки експертів; б) методом середньозваженої оцінки; в) метод ранжування; г) метод попарного порівняння.	Шкала порядку		
	Очевидна					
	Ілюзорна					
Емпірична	Конструктивна	Конвергентна	Кореляційний аналіз (критерієм валідації виступають оцінки учнів, отримані за іншим тестом, валідність якого вже встановлено) (можливий в поєднанні з методом контрастних груп)	а) коефіцієнт кореляції моментів Пірсона;	Інтервальна шкала	
		Дискримінативна		б) коефіцієнт кореляції моментів Спірмена;	Шкала порядку	
		Валідність за віковою диференціацією		в) коефіцієнт кореляції моментів Кендала.	Шкала порядку	
		Факторна		Факторний аналіз	Шкала порядку	
	Критеріальна	Консенсусна (ретроспективна, поточна, прогностична, інкриментна)	Кореляційний аналіз (критерієм валідації виступають експертні оцінки учнів) (можливий в поєднанні з методом контрастних груп, при цьому в якості критерію виступає різниця об'єму викладеного матеріалу в контрастних групах)	а) коефіцієнт кореляції моментів Пірсона;	Інтервальна шкала	
				б) коефіцієнт кореляції моментів Спірмена;	Шкала порядку	
				в) коефіцієнт кореляції моментів Кендала.	Шкала порядку	
		Порівняльна (ретроспективна, поточна, прогностична, інкриментна)	Кореляційний аналіз (критерієм валідації виступають оцінки учнів, отримані за іншим тестом, валідність якого вже встановлено) (можливий в поєднанні з методом контрастних груп, при цьому в якості критерію виступає різниця об'єму викладеного матеріалу в контрастних групах)	а) коефіцієнт кореляції моментів Пірсона;	Інтервальна шкала	
				б) коефіцієнт кореляції моментів Спірмена;	Шкала порядку	
				в) коефіцієнт кореляції моментів Кендала.	Шкала порядку	
		Валідність окремих завдань тесту	Кореляційний аналіз	Коефіцієнт бісеріальної кореляції	Інтервальна шкала	Нормальний розподіл первинних даних
				Коефіцієнт точково бісеріальної кореляції	Інтервальна шкала	
Коефіцієнт рангової бісеріальної кореляції	Шкала порядку					

Таблиця 2

Оцінювання різних видів валідності
на стадіях життєвого циклу КСТ

	Стадія життєвого циклу КСТ	Вид валідності, що визначається
1	Розробка проекту КСТ	Змістовна, очевидна, ілюзорна (закладаються в проект)
2	Стратегічний розрахунок та розробка КСТ	Змістовна, очевидна, ілюзорна (реалізуються та перевіряються в проект) Конструктивна (оцінюється)
3	Пілотний експеримент КСТ	Критеріальна (оцінюється)
4	Експлуатація КСТ	Критеріальна (уточнюється)

Оцінювання критеріальної валідності тесту та завдань тесту, що складають тестовий простір КСТ, проводять на стадії пілотного експерименту (тобто в процесі стандартизації, атестації КСТ). Оцінювання та контроль критеріальної валідності на стадії дослідної експлуатації КСТ на думку автора є невід'ємною умовою забезпечення якості подібних систем.

В той час як в психологічних дослідженнях оцінювання критеріальної валідності тестів дуже поширене через наявність численних варіантів тестів, що дублюють один одного, то в педагогічній практиці оцінювання критеріальної валідності часто обмежується визначенням консенсусної валід-

ності (тобто такої, де критерієм валідності виступають зовнішні експертні оцінки рівня підготовленості опитуваних). Іншими словами консенсусну валідність визначають за збіжністю результатів тестування з традиційними методами оцінювання. На думку автора при створенні тестового простору КСТ бажано знайти більш обґрунтовані кількісні показники для порівняння. Таким показником може бути порівняльна валідність. Однак розгляд методів та математично апарату, який використовується при її оцінюванні дозволяє висловити певні зауваження [16].

Найпоширенішим методом оцінювання критеріальної валідності є використання кореляційного аналізу. Зокрема в літературі [2 – 13] поширеним є коефіцієнт кореляції Пірсона. При цьому не враховується той факт, що даний коефіцієнт визначається на основі первинних балів опитуваних, що мають властивості шкали порядку, в той час як коефіцієнт кореляції Пірсона розрахований на обробку неперервних величин, визначених на шкалі відношень [20].

Тому вважаємо більш доцільним в подібних випадках використовувати показники рангової кореляції такі, як коефіцієнти рангової кореляції Спірмена, Кендала та коефіцієнт рангової бісеріальної кореляції [2, 19].

При проведенні пілотного експерименту та дослідної експлуатації КСТ важливо оцінити валідність, що характеризує зв'язок кожного завдання з усім тестом, адже всі завдання, що обрані для тестування мають відповідати загальній меті тестування. Характеристикою такої валідності виступає бісеріальна кореляція кожного завдання із зовнішнім критерієм – результатом виконання завдань тесту (первинними балами опитуваних). Використання цього показника для характеристики якості тестів наводиться в [2, 13]. Однак з огляду на властивостей шкали порядку, якою відображаються первинні бали опитуваних, автор вважає за доцільно як метрику критеріальної валідності завдань тесту обрати рангову бісеріальну кореляцію.

Коефіцієнт рангової бісеріальної кореляції визначають за наступною формулою [77]:

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

де \bar{r}_X , \bar{r}_Y – середні ранги за ознаками X та Y відповідно; n – кількість досліджуваних об'єктів.

З урахування того, що ранги в формулі 1 є номерами елементів досліджуваного ряду і являють собою елементи порядкової шкали, замість середніх арифметичних в оригінальній формулі пропонується використовувати медіани ряду. Тоді для розрахунку бісеріальної кореляції, як показника валідності тестових завдань формула набуває вигляд:

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

де $\text{med}\{r_B\}$, $\text{med}\{r_H\}$ – медіана рангів первинних балів опитаних, які правильно та неправильно відповіли на завдання тесту відповідно; n – кількість опитуваних, які взяли участь у тестуванні.

Показника критеріальної валідності завдань тесту (за формулою 2) може набувати значень від -1 до 1. Менші за нуль значення показника свідчать про незадовільність завдання тестового простору. Задовільними можна вважати завдання, для яких отримано $r_{bis} > 0.3$ [2].

Оцінювання невизначеності показника валідності

Невизначеність показника критеріальної валідності завдань тесту пропонується оцінювати наступним чином. Для проранжованих рядів первинних балів опитаних, які правильно $b'_{B,i}$ та неправильно $b'_{H,i}$ відповіли на завдання тесту оцінюють інтервали, що характеризує точність їх медіанних оцінки при довірчій ймовірності P відповідно:

$$(b'_{Bu,i}, b'_{Bv,i}) \text{ та } (b'_{Hu,i}, b'_{Hv,i}), \quad i = \bar{1}, n, \quad (3)$$

де $b'_{Bu,i}$, $b'_{Bv,i}$, $b'_{Hu,i}$, $b'_{Hv,i}$ - елементи проранжованих рядів $b'_{B,i}$ та $b'_{H,i}$; u - номер елемента, найбільше ціле число, що є меншим за $(n + 1 - z_p \sqrt{n}) / 2$; v - номер елемента, найменше ціле число, що є більшим за $(n + 1 + z_p \sqrt{n}) / 2$; z_p – квантіль стандартного гаусівського розподілу? при $P = 0,95 \quad z_p = 1,96$.

Тоді для оцінювання розширеної невизначеності з рівнем довіри $P = 0,95$ за умови нормального розподілу показника критеріальної валідності завдань тесту, розрахованого за формулою (2), виваженно вираз:

$$U[r_{bis,i}] = n^{-1}[(b'_{Bv,i} + b'_{Hv,i}) - (b'_{Bu,i} + b'_{Hu,i})] \quad (4)$$

Динаміку зменшення максимального значення розширеної невизначеності показника валідності $U[r_{bis,i}]$ (за форм. 4 для теста з 10 завдань) залежно від кількості опитуваних n ілюструє табл. 3.

Таблиця 3

Динаміка зміни розширеної невизначеності показника валідності в залежності від кількості опитуваних

n	20	100	200	300	400	500
$U[r_{bis,i}]$	0,5	0,1	0,05	0,03	0,025	0,02

Таким чином розрахунки, наведені в таблиці, підтверджують необхідність проведення моніторингу КСТ в процесі експлуатації (під час якого здійс-

ноється збір та інтегральне опрацювання дослідних даних) з метою підвищення точності оцінювання показника валідності завдань тесту.

Список літератури

1. Паращенко Л.І. Тестові технології в навчально-му закладі: Методичний посібник / Л.І. Паращенко, Л.І. Леонський, В.Д. Леонська. – К.: ТОВ «Майстерня книги», 2006. – 217 с.
2. Чельшикова М.Б. Теория и практика конструирования педагогических тестов: Учебное пос. / М.Б. Чельшикова. – М.: Логос, 2002. – 432 с.
3. Шведова В.В., Моніторинг систем дистанційного навчання / В.Д. Ціделко, Н.А. Яремчук, В.В. Шведова // Вища освіта України. – 2006. – №2. – С. 54 – 61.
4. Беспалько В.П. Программированное обучение (дидактические основы) / В.П. Беспалько. – М.: Высшая школа, 1979. – 300 с.
5. 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.
6. Колин К.К. Инновационное развитие в информационном обществе и качество образования / К.К. Колин // Открытое образование. – 2009. – № 3/ – С. 63-71.
7. Кондратьев И.Н. Основные составляющие образовательной среды дистанционного образовательного университетского комплекса / И.Н. Кондратьев // Открытое образование. – 2008. – № 1. – С. 11 – 18.
8. Вісмут Дж. Оцінювання для навчання: навч. посіб./ За ред. І.С. Булах/ – К.: Майстер-клас, 2007. – 170 с.
9. Ясулатіс В.А. Дистанційне навчання: метод. Рекомендації / В.А. Ясулатіс. – К.: МАУП, 2005. – 72 с.
10. Волков Н.И. Тестовый контроль знаний: учебное пособие / Н.И. Волков, А.Н. Алексеев, Н.А. Алексеев. – Сумы: ИТД «Университетская книга», 2004. – 109 с.
11. Алексеев А.Н. Дистанционное обучение инженерам специальностям: Монография / А.Н. Алексеев. – Сумы: ИТД «Университетская книга», 2005. – 333 с.
12. Алексеев Н.А. К вопросу о повышении достоверности оценки при тестовом контроле знаний / Н.А. Алексеев, Н.И. Волков, Т. А. Майорова // Открытое образование. – 2004. – № 3. – С. 27 – 32.
13. Бурлачук Л.Ф. Словарь-справочник по психодиагностике: 2-ое изд., перераб. и дополн. / Л.Ф. Бурлачук, С.М. Морозов. – СПб.: Питер, 2004. – 250 с.
14. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов / Ю.М. Нейман, В.А. Хлебников. – М., 2000. – 168 с.
15. Аванесов В.С. Композиция тестовых заданий. Учебная книга. 3 изд., доп. / В.С. Аванесов. – М.: Центр тестирования, 2002. – 240 с.
16. Шведова В.В. Оцінювання валідності тестового простору комп'ютеризованої системи тестування // Збірник статей учасників двадцять дев'ятої Міжнародної науково-практичної конференції «Інноваційний потенціал світової науки – ХХІ сторіччя». Т. 2. Природничі і точні науки. – Запоріжжя 2014.
17. Шведова В.В. Систематизація вимірюваних величин і складових невизначеності / В.Д. Ціделко, Н.А. Яремчук, В.В. Шведова // Метрологія та вимірювальна техніка. Метрологія – 2006: V Міжнародна науково-технічна конференція. – Х., 2006. – С. 48 – 50.
18. Шведова В.В. Застосування непараметричних оцінок при визначенні показників якості контролюючих тестів в системах дистанційного навчання / Н.А. Яремчук, В.В. Шведова // Гидротехнологии, навигация, управление и конструирование авиационно-космической техники: VII Міжн. науково-технічна конференція. – К., 23-24 квітня, 2009. – С. 261 – 265.
19. Гмурман В.Е. Теория вероятностей и математическая статистика: учебн. пособие для вузов - 10-е изд. стер. / В.Е. Гмурман. – М.: Высш. шк., 2004. – 479 с.
20. Шведова В.В. Оцінювання якості контролюючих тестів / В.В. Шведова, В.Д. Ціделко, Н.А. Яремчук // Техніка АПК. – 2006. – № 4. – С. 34 – 36.
21. Паніотто В.І. Статистичний аналіз соціологічних даних / В.І. Паніотто, В.С. Максименко, Н.М. Харченко. – К.: Вид.дім «КМ Академія», 2004. – 270 с.

Надійшла до редколегії 25.12.2014

Рецензент: д-р техн. наук, проф. І.В. Руженцев, Харківський національний університет радіоелектроніки, Харків.

ОЦЕНИВАНИЕ ВАЛИДНОСТИ ТЕСТОВОГО ПРОСТРАНСТВА КОМПЬЮТЕРИЗИРОВАННОЙ СИСТЕМЫ ТЕСТИРОВАНИЯ: ВИДЫ ВАЛИДНОСТИ, МЕТОДЫ И СПОСОБЫ ОЦЕНИВАНИЯ ПОКАЗАТЕЛЯ И ЕГО НЕОПРЕДЕЛЕННОСТИ

В.В. Шведова

Автором систематизированы виды валидности и проанализированы методы и способы оценивания этого показателя качества компьютеризированной системы тестирования (КСТ); выделены виды валидности, которые целесообразно оценивать на разных стадиях жизненного цикла КСТ; выбрана метрика для оценивания валидности КСТ на стадии пилотного эксперимента и опытной эксплуатации и проведено ее усовершенствование с точки зрения особенностей шкалы, в которой оцениваются входные данные для расчета этого показателя. Приведена динамика изменения неопределенности показателя валидности в зависимости от объема тестируемых, которые привлечены к работе с КСТ.

Ключевые слова: система дистанционного обучения, компьютеризированная система тестирования, тестовое пространство, валидность тестового пространства, неопределенность.

ASSESSMENT OF VALIDITY OF THE TEST AREA FOR COMPUTERIZED TESTING SYSTEM: TYPES OF VALIDITY, METHODS AND TECHNIQUES OF EVALUATION OF ITS INDEX AND UNCERTAINTY.

V.V. Shvedova

Author systematized and analyzed the types of validity, methods and techniques for evaluation of this quality index of the computerized testing system (CTS); allocated types of validity, which should be assessed at various stages of the life cycle of the CTS; selected metric for evaluating the validity of the CTS at the stage of a pilot experiment and trial operation and conducted its improvement in terms of features of the scale, which assesses the input data for this index. The dynamics of changes in the uncertainty of validity depending on the volume of the tested students, which involved working with the CTS.

Keywords: distance learning system, computerized testing system, the test space, validity of test space, uncertainty.