

УДК 303.714

Ю. В. Бобков, О. М. Ліпіна

СИСТЕМА ТЕСТУВАННЯ ДИСТАНЦІЙНОГО КУРСУ «ЦИФРОВІ ВИМІРЮВАЛЬНІ ПРИЛАДИ»

Вступ

Використання в навчальному процесі нових інформаційних технологій і, зокрема, дистанційних форм навчання дозволяє не тільки істотно розширити коло студентів, а й отримати нові якісні можливості в порівнянні з традиційними видами занять. Особливо це характерно для технічних дисциплін, де істотне місце займають такі форми роботи студентів, як: лабораторні та практичні заняття, розрахункові і розрахунково-графічні роботи, курсові роботи та проекти. Однак, саме ці види робіт досить складно реалізуються в системах дистанційного навчання, оскільки в порівнянні з лекційними заняттями вимагають створення спеціального інструментарію для прищеплення учням реальних практичних умінь і навичок, аналогічних тим, які вони отримують при традиційних формах виконання завдань. Причому, саме ці види занять дозволяють сформувати повноцінного технічного фахівця, якого вимагають на ринку праці.

Ефективність дистанційних форм навчання залежить, не в останню чергу, від можливостей контролю засвоєння студентами навчального матеріалу, ступеня підготовленості до виконання лабораторних і практичних занять. Тому важливим елементом дистанційних курсів є різноманітні форми тестів, що використовуються для поточного, допускового, рубіжного, підсумкового та інших видів контролю.

Вказані функції тестів дозволяють реалізувати відповідні види педагогічних вимірювань рівня знань та умінь студентів. В зв'язку з цим на перше місце висувається визначення показників якості тестів, які дозволяють не тільки стандартизувати відповідні тести, але є також потужним інструментом вдосконалення тестових завдань.

На кафедрі «Інформаційно-вимірювальної техніки» Національного технічного університету України «Київський політехнічний інститут» підготовлена система дистанційного навчання з дисципліни "Цифрові вимірювальні прилади", яка є однією з базових у підготовці бакалаврів, спеціалістів і магістрів за напрямом 6.051001 "Метрологія та інформаційно-вимірювальні технології". Доступ до неї може здійснюватися або через локальну мережу, або через мережу Інтернет за допомогою спеціального сайту дисципліни. Система дистанційного навчання включає до свого складу лекції, лабораторні роботи, практичні заняття, курсове проектування та бібліотеку. Невід'ємною складовою частиною дистанційного курсу є система тестування.

Постановка задачі

Враховуючи важливість тестування в системах дистанційного навчання (СДН) як для вимірювання рівня знань і умінь студентів, так і для оцінки ефективності самих СДН, необхідно виробити показники та процедуру оцінки якості тестів.

На сьогодні існують різні підходи для вирішення задачі стандартизації тестів [2, 3, 4]. При цьому стандартизація показників якості тестів являє собою процедуру визначення показників якості тестових завдань та тестів в цілому в рамках прийнятої теорії педагогічного тестування за результатами проведення апробаційного тестування.

Таким чином, для стандартизації системи тестування, в тому числі з дисципліни «Цифрові вимірювальні прилади», необхідно розробити процедуру комплексної експертизи якості тестових завдань тестів, розрахувати значення показників якості за результатами апробаційного тестування та визначити шляхи вдосконалення тестів з метою досягнення встановлених норм.

Структура системи тестування з дисципліни «Цифрові вимірювальні прилади»

Структура системи тестування відповідає структурі навчальної дисципліни «Цифрові вимірювальні прилади» та структурі дистанційного курсу з цієї дисципліни.

Відповідно до основних видів навчальних занять: лекції, лабораторні роботи, практичні заняття – були розроблені три незалежні системи тестування, виконанні у окремих незалежних базах в середовищі Moodle.

Система тестування охоплює перші три основні види занять та включає:

- систему тестування з теоретичного матеріалу (лекцій), що охоплює 18 тем (7 розділів) теоретичного навчання і включає 289 запитань;
- систему тестування з практичних занять, що охоплює 8 тем і включає 116 запитань;
- систему тестування з лабораторних робіт, що охоплює 19 віртуальних лабораторних робіт і включає 193 запитання.

Для тестових запитань обраний їх єдиний уніфікований вид – закрита форма з вибором однієї правильної відповіді з деякої множини відповідей. Для кожного контрольного запитання в залежності від його складності та змісту формувалась різна кількість дистракторів – від 4-х до 8-ми. Цей вид дозволяє достатньо глибоко перевіряти знання студентів, легко адаптується для сприйняття студентами, а результати легко обробляються та інтерпретуються. Тестова система виконана в середовищі Moodle та дозволяє реалізувати різні стратегії тестування та підготовку різнобічних видів звітів.

Оцінювання показників якості тестів

Для визначення ефективності розробленої системи при оцінюванні знань і вмінь студентів, а також шляхів її удосконалення були проведені дослідження показників якості системи тестування. При цьому були визначені показники якості системи тестування, запропонована процедура їх розрахунків та проведені необхідні обчислення числових значень.

На підставі аналізу відомих методів оцінювання показників якості тестової інформації [2, 3, 4] була розроблена процедура комплексної експертизи якості тестових завдань і тестів.

Система комплексної експертизи якості тестових матеріалів складається з чотирьох основних етапів:

- 1) попередня експертиза якості тестових матеріалів, метою якої є відбір авторського матеріалу, який би відповідав вимогам тестів;
- 2) експертне оцінювання (внутрішня експертиза) якості тестів, що проводиться після складання першого варіанта тесту;
- 3) тестологічна експертиза, що полягає у проведенні апробаційного тестування і розрахунку на основі його результатів статистичних характеристик якості тестів і тестових завдань;
- 4) комплексне оцінювання якості тестових завдань і тестів.

Важливим етапом комплексної експертизи якості тестів і тестових завдань (ТЗ) є проведення апробаційного тестування з метою встановлення, перевірки та оцінки його вимірювальних можливостей шляхом апробації на репрезентативних вибірках. У зв'язку з цим було проведено апробаційне тестування в групах ВВ–51 (14 осіб), ВВ– 52 (16 осіб), ВВ–61 (15 осіб), ВВ–62 (12 осіб). Після апробаційного тестування проводиться розрахунок ряду параметрів. Розглянемо основні етапи розрахунку показників якості.

Етап 1. Формування матриці відповідей, випробовуваних на завдання, в якій міститься інформація про відповіді в дихотомічній або політомічній шкалі, номер варіанта, виконаного тестованим, інформація про вибір варіантів відповідей на кожне тестове завдання із запропонованих тестованому. В даному випадку використовується дихотомічна шкала оцінок результатів тестування, тобто коли безліч можливих оцінок складається всього з двох елементів (0; 1): 0 – завдання виконано невірно, 1 – завдання виконано правильно [1, 2].

Позначимо через x_{ij} числову оцінку успішності виконання j -го завдання, виконаного i -м випробуваним. Результати тестування подаються у вигляді матриці (x_{ij}) з N рядками і стовпцями L ($i=1, \dots, N; j=1, \dots, L$).

Після формування матриці відповідей з неї видаляються всі рядки та стовпці, які містять тільки нулі або тільки одиниці, оскільки вони не несуть інформації, яку можна використовувати для оцінювання рівня підготовленості учнів.

Якщо завдання вирішено всіма випробуваними, значить воно занадто просте для даної вибірки досліджуваних, якщо завдання невирішене жодним випробуваним в групі, то воно занадто складне для даної вибірки випробовуваних. Отже, такі завдання не слід враховувати при оцінюванні випробовуваних, і, ймовірно, їх необхідно виключити з тесту.

Якщо випробуваний не вирішив жодного завдання із запропонованого тесту, то для оцінки рівня його підготовленості необхідно використовувати тест з більш простими завданнями, і навпаки, якщо випробуваний вирішив всі завдання тесту, то для точної оцінки його рівня підготовленості необхідно використовувати тест з більш складними завданнями [3, 4].

Після цього здійснюється упорядкування матриці результатів тестування. Для цього змінюють місцями рядки матриці так, щоб верхній рядок відповідав випробуваним з мінімальним індивідуальним балом. Значення X_i розташовують згори вниз у порядку зростання.

Далі проводиться підрахунок індивідуальних балів випробовуваних і кількість правильних відповідей випробовуваних за кожне завдання тесту:

1) кількість випробовуваних, які правильно відповіли на завдання тесту (число правильних відповідей); R_j ;

2) індивідуальний бал i -го випробовуваного:
$$X_i = \sum_{j=1}^L X_{ij}.$$

Етап 2. Виконується графічна інтерпретація емпіричних даних. Емпіричні результати тестування можна представити у вигляді полінома, гістограми, згладженої кривої тощо. Для побудови кривих необхідно впорядкувати результати експерименту. Їх можна записати у вигляді не згрупованого ряду довільної форми, рангового ряду, частотного розподілу або розподілу згрупованих частот.

Етап 3. Обробка матриць відповідей з метою отримання та інтерпретації кількісних показників ТЗ:

1) кількість випробовуваних $W_j = N - R_j$, які неправильно відповіли на завдання тесту; N – загальна кількість випробовуваних;

2) середнє арифметичне індивідуальних балів випробовуваних за весь тест:
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i;$$

3) відносна кількість студентів $p_j = \frac{R_j}{N}$, які відповіли на завдання тесту або індекс легкості. Індекс легкості тестових завдань є відносним, тобто він характеризує складність завдань для конкретної групи випробовуваних і залежить від рівня їх підготовки. Важливість отримання кількісних характеристик складності пропонованих випробуваним завдань полягає в тому, що для того, щоб ці завдання мали здатність диференціювати випробовуваних за рівнем їх підготовки, тобто служити засобом виміру цього рівня, їх складність повинна відповідати рівню підготовки випробовуваних. Тест в цілому має включати в себе комплекс завдань різної складності – від легких до важких, проте очевидно, що занадто прості завдання, на які правильно відповідають усі досліджувані, і занадто складні завдання, на які не може відповісти ніхто з випробовуваних, не мають здатності

диференціювати випробуваних за рівнем їх підготовки і в цьому сенсі не є тестовими завданнями;

4) відносна кількість студентів $q_j = \frac{W_j}{N}$, які не відповідали на завдання тесту;

5) дисперсія завдання тесту: $p_j \cdot q_j$;

6) середньоквадратичне відхилення тесту: $\sqrt{p_j \cdot q_j}$;

7) сума квадратів відхилень індивідуальних балів студентів:

$$8) V = \sum_{i=1}^N V_i^2 = \sum_{i=1}^N (x_i - \bar{x})^2;$$

9) дисперсія індивідуальних балів випробуваних:

$$10) S_x^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1};$$

11) варіація або стандартне відхилення індивідуальних балів випробуваних:

$$12) S_x = \sqrt{S_x^2} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}};$$

Етап 3. Розрахунок і оцінювання показників надійності тесту, що характеризує якість тесту.

Надійність (reliability) – характеристика тесту, яка відображає точність педагогічних вимірювань, ступінь постійності результатів тестування, а також стійкість результатів тесту до дії сторонніх випадкових факторів.

У даній роботі надійність тесту розраховувалась за формулою Кьюдера – Річардсона [2].

$$KR_{20} = \frac{L}{L-1} \left(\frac{S_x^2 - \sum_{j=1}^L p_j \cdot q_j}{S_x^2} \right) = \frac{L}{L-1} \left(1 - \frac{\sum_{j=1}^L p_j \cdot q_j}{S_x^2} \right).$$

Етап 4. Розрахунок і оцінювання показників валідності тесту, що характеризує якість тесту.

Критеріальна валідність тесту розраховується як бісеріальний коефіцієнт зв'язку кожного завдання з усіма завданнями за формулою [4]:

$$(r_{bis})_j = \frac{(\bar{X}_1)_j - (\bar{X}_0)_j}{S_x} \cdot \sqrt{\frac{R_j \cdot W_j}{N \cdot (N-1)}} = \frac{(\bar{X}_1)_j - (\bar{X}_0)_j}{S_x} \cdot \sqrt{p_i \cdot q_i},$$

де $(\bar{X}_1)_j$ - середнє значення індивідуальних балів випробуваних, які вірно відповіли на $j - e$ завдання тесту; $(\bar{X}_0)_j$ - середнє значення індивідуальних балів випробуваних, які невірно відповіли на $j - e$ завдання тесту.

Етап 5. Аналіз отриманих результатів обчислень. Формування висновків у вигляді рекомендацій.

Аналіз результатів апробаційного тестування статистичним методом та за допомогою Moodle

Враховуючи, що система тестування з дисципліни «Цифрові вимірювальні прилади» розміщена в середовищі Moodle, доцільним було б застосування для обробки результатів його обчислювальних можливостей. Система аналізу Moodle [6, 7] дозволяє розраховувати індекс легкості, середньоквадратичне відхилення, коефіцієнт диференціації та індекс диференціації. Основною перевагою обробки результатів Moodle є можливість визначати відносну частоту вибору випробуваними тих чи інших дистракторів при виконанні конкретного тестового завдання, що дозволяє покращувати внутрішню структуру тестових завдань в закритій формі. Проте аналіз показав наявність відмінностей в алгоритмах обробки результатів у системі Moodle і запропонованих вище.

Головна відмінність в тому, що в системі Moodle не видаляються рядки та стовпці, які містять тільки нулі або тільки одиниці, як це робиться в статистичному методі на першому кроці та в класичній теорії.

У зв'язку з вказаними відмінностями обробки результатів статистичним методом та Moodle було проведено порівняльний аналіз результатів розрахунків.

З цією метою були обрані тести з практичних занять № 2 та № 3:

- в результатах тестування з практичного заняття № 2 відсутні нульові та одиничні рядки та стовпці;

- у результатах тестування з практичного заняття № 3 присутні неінформативні параметри.

Результати показали, що відмінності в отриманих двома способами показників якості з практичного заняття №2 знаходяться в межах похибки розрахунків (округлення), а з практичного заняття №3 – значні відмінності.

Для прикладу в таблиці 1 наведено результати розрахунку індексу легкості результатів тестування груп ВВ–51, ВВ–52 з практичного заняття №3 трьома способами за допомогою:

- статистичного методу з видаленням неінформативних параметрів з матриці результатів тестування ($P_j, \%$);

- статистичного методу без видалення неінформативних параметрів з матриці результатів тестування ($P_{jП}, \%$);

- Moodle (ІЛ,%).

Таблиця 1

Відхилення значень індексу легкості розрахованого статистичним методом і Moodle

Показники, %	Завдання теста										
	1	2	3	4	5	6	7	8	9	10	11
P_j	100	69,57	86,96	39,1	26,1	82,61	91,3	47,8	34,78	52,17	78,26
P_{jII}	100	74,07	88,89	48,1	37	85,19	92,59	55,6	44,44	59,26	81,4
Індекс легкості (ІЛ) Moodle	100	74	89	48	37	85	93	56	44	59	81
$P_{jII} - ІЛ$	0	-0,07	0,11	-0,15	-0,04	-0,19	0,41	0,44	-0,44	-0,26	-0,48
$P_j - ІЛ$	0	4,43	2,04	8,87	10,9	2,39	1,696	8,17	9,2	6,8	2,74

Як видно з табл. 1 максимальне відхилення обчислення за допомогою Moodle досить значне $\Delta_{\max} = 10,9\%$.

У зв'язку з цим доцільно проводити аналіз за допомогою статистичного методу. Розрахунки показників якості, обчислені в Moodle, можна застосовувати лише як орієнтовні, або в оптимізованих тестових системах, результати тестування в яких не містять неінформативних параметрів. На етапі апробаційного тестування використання обчислення показників якості за допомогою Moodle вважаємо недоцільним.

Подальші дослідження системи тестування з дистанційного курсу дисципліни «Цифрові вимірювальні прилади» проводилися за допомогою викладеного вище статистичного методу.

За результатами дослідження шістьох лабораторних робіт, трьох тем лекційного матеріалу, сімох тем практичних занять, були визначені:

- 1) для кожного тестового завдання: індекс легкості, розрізняльна здатність та критеріальна валідність;
- 2) для тесту в цілому – надійність.

Аналіз результатів тестування допоміг на перших етапах усунути технічні помилки і неточності в завданнях тесту та визначити завдання (загальною кількістю 56), що підлягають корегуванню.

Аналіз кореляції результатів апробаційного тестування

У роботі також було проведено дослідження кореляції загальної оцінки результатів тестування з результатами іспиту з дисципліни «Цифрові вимірювальні прилади», а також з середнім балом диплома бакалавра груп ВВ-51, ВВ-52.

На рис. 1 наведені в графічному вигляді результати тестування (крива 1), результати іспиту з дисципліни «Цифрові вимірювальні прилади» (крива

2) та середній бал диплома бакалавра груп ВВ–51, ВВ–52 (крива 3) за п'ятибальною шкалою.



Рис. 1. Результати тестування (1), іспиту (2) з дисципліни «Цифрові вимірювальні прилади» та середній бал диплома бакалавра (3) груп ВВ–51, ВВ–52

Було розраховано значення коефіцієнта кореляції, яке склало: 1) для загальної оцінки за тестуванням з результатами іспиту – 0,6; 2) для загальної оцінки за тестування з середнім балом диплома бакалавра – 0,5; 3) для середнього бала диплома бакалавра з результатами іспиту – 0,8.

Отримані значення коефіцієнта кореляції є достатньо високими, що вказує на ефективність розробленої системи тестування.

Висновки

Проведені дослідження дозволили запропонувати методіку оцінювання показників якості тестів в рамках класичної теорії тестування. За цією методикою за результатами апробаційного тестування в групах ВВ – 52, ВВ – 52, ВВ – 61, ВВ – 62 були визначені показники якості для системи тестування з дисципліни «Цифрові вимірювальні прилади». В ході дослідження встановлено, що застосування для обчислення значень показників якості системи Moodle можливе лише при відсутності неінформативних параметрів, тобто лише в оптимізованих тестових системах.

Проведені дослідження показників якості дозволяють зробити висновок про можливість застосування розробленої системи тестування дистанційного курсу дисципліни «Цифрові вимірювальні прилади» у навчальному процесі для оцінки рівня підготовки студентів при поточному, допусковому, рубіжному, підсумковому та інших видах контролю.

За результатами показників якості планується корегування 56 тестових завдань, які не відповідають встановленим критеріям, з подальшим

дослідженням показників якості скорегованої тестової системи за запропонованою методикою.

Список використаної літератури

1. *Алексеев А. Н., Волков Н. И., Майорова Т. А.* К вопросу о повышении достоверности оценки при тестовом контроле знаний // Открытое образование. – 2004. - №3. - С. 27-32.
2. *Хлебников В. А.* Теория и методы оценки эффективности систем обучения коллективного пользования. – Автореферат к диссертационной работе, - Москва, 2006.
3. *Нейман Ю. М., Хлебников В. А.* Введение в теорию моделирования и параметризации педагогических тестов. – Москва, 2000. –168 с.
4. *Чельщикова М. Б.* Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.
5. <http://docs.moodle.org/>
6. <http://docs.moodle.org/ru/>