

К. О. Городнича<sup>1</sup>  
В. А. Крісілов<sup>1</sup>  
Т. В. Оніщенко<sup>1</sup>

## МЕТОДИКА ОЦІНКИ ТА ПІДВИЩЕННЯ ВАЛІДНОСТІ ПЕДАГОГІЧНОГО ТЕСТУ НА БАЗІ КРИТЕРІЇВ ЯКОСТІ ТЕСТУ

<sup>1</sup>Одеський національний політехнічний університет

*Розглядається методика підвищення валідності тесту за рахунок аналізу та зміни значень критеріїв якості контролю знань в навчальному процесі. Запропонована методика дає можливість підвищити якість перевірки знань і умінь студентів з вивченого матеріалу за певний період часу, без залучення експертів.*

**Ключові слова:** змішане навчання, дистанційне навчання, педагогічні тести, оцінка педагогічних тестів, критерії якості тесту, підвищення якості педагогічних тестів, автоматизований контроль знань, валідність тесту, експертиза тестових завдань.

### Вступ

Одним з найперспективніших шляхів розвитку сучасних інформаційних технологій у сфері освіти є впровадження змішаного навчання. При цьому найважливішою складовою навчання залишається контроль знань. Питання якості навчання невіддільний від питання якості контролю знань, оскільки саме контроль є одним з факторів, що мотивують до безперервного і глибокого вивчення матеріалу.

Педагогічні тести дозволяють провести об'єктивну оцінку досягнутого рівня знань у разі масової перевірки. Розробка загальних принципів оцінки якості тестів — досить актуальна проблема, тому що важко здійснити послідовне зіставлення цілей освіти з результатами навчання, які можуть бути досягнуті [1], і для цього необхідно формалізувати методику оцінки тесту [2].

Експертиза тестових завдань є обов'язковою умовою для створення тестів незалежно від сфери їх застосування. Експертизу тестових завдань можна перевіряти за такими критеріями якості тесту, як репрезентативність, складність, надійність і валідність тесту.

Існують два підходи оцінки якості тесту: апріорний і апостеріорний. За цією методикою аналіз репрезентативності проводиться за апріорними оцінками, а розрахунок таких критеріїв якості тесту, як складність і надійність проводиться за результатами пробного тестування, тобто за апостеріорними оцінками.

Єдиним способом оцінити якість тесту є аналіз його валідності. Пропонується визначати якість тесту за кількісними характеристиками його валідності, які значно залежать від репрезентативності, складності та надійності тесту.

*Метою роботи* є підвищення валідності тесту за рахунок аналізу критеріїв якості контролю знань у навчальному процесі, а саме: аналіз репрезентативності тесту, аналіз складності тесту і аналіз надійності тесту.

### 1. Аналіз репрезентативності тесту

Одним з недоліків сучасних інформаційних систем навчання є те, що тестові питання не контролюються системою з точки зору репрезентативності. Особливо ця проблема стає важливою в умовах проведення комплексного тестування: вступних іспитів, державних іспитів тощо. Це істотно знижує якість тестування. На практиці вибір тестових завдань дуже часто виконується довільним чином і залежить від особи, що розробляє тест. Навіть в існуючих на сьогодні системах дистанційного навчання, таких як Moodle, ILIAS, WebTutor, CourseLab, Blackboard Learning System, оцінка навчання не контролюється системою з точки зору репрезентативності. Через це тест носить

суб'єктивний характер і є слабо формалізованим.

Запропонований аналіз репрезентативності заснований на аналізі пропорції кількості питань, які охоплюють різні розділи і навчальні цілі курсу.

Для достовірної перевірки знань студентів необхідно створювати репрезентативні тести, тобто набір питань, що охоплює всю необхідну інформацію із заданого фрагмента дисципліни: теми, навчального модуля, підсумкового контролю, вступного іспиту тощо.

Пропонований аналіз містить два основних етапи:

1. Розрахунок кількісних характеристик репрезентативної вибірки питань;
2. Перевірка та підвищення репрезентативності тесту.

Репрезентативність — відповідність характеристик вибірки характеристикам генеральної сукупності в цілому [3]. У цьому випадку репрезентативна вибірка — це вибірка тестових питань, в якій кількість питань представлена в тій же пропорції, що матеріал в певній дисципліні.

Занадто мала кількість питань не дасть отримати повну характеристику знань студентів, а неправильний розподіл кількості питань за темами не дасть можливість отримати об'єктивну оцінку знань студент, тому необхідно створювати репрезентативну вибірку тестових питань.

Якщо вибірка для тесту не буде репрезентативною, то матиме місце зміщення оцінок, що приведе до неправильної оцінки знань студентів. Перед аналізом репрезентативності викладачеві необхідно визначити загальну кількість питань у тесті, виходячи з виду тестового контролю: проміжний контроль, модульний контроль, екзаменаційний контроль тощо.

### 1.1. Розрахунок кількісних характеристик репрезентативної вибірки питань

Обчислюючи репрезентативність вибірки, необхідно спиратися на одну з таких статистичних характеристик курсу:

1. Нормативний час, відведений на вивчення навчального фрагмента;
2. Кількість навчальних цілей, сформульованих в рамках навчального процесу;
3. Коефіцієнт важливості навчального фрагмента, суб'єктивний коефіцієнт, який визначає викладач самостійно виходячи зі змісту теми, кількості навчальних цілей в темі, часу вивчення теми та інше.

Кількість питань, які потрібно включити в тест за  $i$ -ю темою розраховується за формулою

$$n_i = n \frac{X_i}{\sum_{i=1}^m X_i}, \quad (1)$$

де  $n_i$  — кількість питань, які потрібно включити в тест по  $i$ -й темі;  $n$  — загальна кількість питань тесту;  $m$  — кількість тем;  $X_i$  — обрана статистична характеристика для курсу в  $i$ -й темі.

### 1.2. Перевірка та підвищення репрезентативності тесту

Якщо тест вже був розроблений, до підрахунку необхідної кількості питань для тесту відповідно до пункту 1.1, то рекомендується перевірити його репрезентативність і, за необхідності, виправити тест.

Для визначення репрезентативності тесту необхідно виконати такі дії:

1. За технологією обчислити статистичні характеристики репрезентативної вибірки питань;
2. Розрахувати міру відповідності розрахованих характеристик тесту та кількості питань в аналізованому тесті;
3. Якщо тест нерепрезентативний, то рекомендується відкоригувати кількість тестових питань, відповідно до пункту 1.1.

Для визначення репрезентативності, вводимо міру відповідності розрахованих характеристик тесту та кількості питань в аналізованому тесті, яка розраховується за формулою

$$M = \frac{\sqrt{\sum_{i=1}^m (n_i - n_{i\text{вкл}})^2}}{m}, \quad (2)$$

де  $M$  — міра відповідності розрахованих характеристик тесту та кількості питань в аналізованому тесті;  $n_{i\text{вкл}}$  — кількість питань, які включили в тест від  $i$ -ї теми до розрахунку;  $m$  — кількість тем.

Репрезентативність тесту досягається у разі рівності значення міри відповідності нулю. Міра

відповідності дорівнює нулю за умови, що кількість питань включених до  $i$ -ї теми навчального курсу збігається з кількістю питань, які необхідно включити в  $i$ -у тему. Чим вище міра відповідності, тим гірше репрезентативність тесту.

### 1.3. Застосування підвищення репрезентативності на практиці

Аналіз репрезентативності використано для оцінки репрезентативності модульних тестів з дисциплін «Структура та організація даних в ЕОМ», «Людино-машинна взаємодія», «Моделювання та аналіз ПЗ» кафедри «Системне програмне забезпечення» для студентів, що навчаються за напрямом 6.050103 «Програмна інженерія».

З дисципліни «Структура та організація даних в ЕОМ» був обраний пріоритет за часом вивчення теми. Час вивчення першої теми був заданий 6 годинам, другої теми — 3, третьої теми — 6. Перевірка репрезентативності тесту показала, що міра відповідності дорівнює 1,25. Тест з такою кількістю питань має низьку репрезентативність, тому кількість питань необхідно переглядати. Для цього за запропонованою методикою розрахували необхідну кількість питань (рис. 1а).

З дисципліни «Людино-машинна взаємодія» був обраний пріоритет за кількістю навчальних цілей. Кількість навчальних цілей першої теми був заданий 1, другої теми — 2, третьої теми — 3, четвертої теми — 3, п'ятої теми — 1, шостої теми — 5, сьомої теми — 1, восьмої теми — 6. Перевірка репрезентативності тесту показала, що міра відповідності дорівнює 1,3. Тест з такою

кількістю питань має низьку репрезентативність, тому кількість питань необхідно коригувати. Для цього за запропонованою методикою розрахували необхідну кількість питань (рис. 1б).

З дисципліни «Моделювання та аналіз ПЗ» був обраний пріоритет за коефіцієнтом важливості. Важливість вивчення першої теми було задано 0,1, другої теми — 0,1, третьої теми — 0,3, для четвертої теми — 0,5. Перевірка репрезентативності тесту показала, що міра відповідності дорівнює 3,6. Тест з такою кількістю питань має низьку репрезентативність, тому кількість питань необхідно коригувати. Для цього за запропонованою методикою розрахували необхідну кількість питань (рис. 1с).

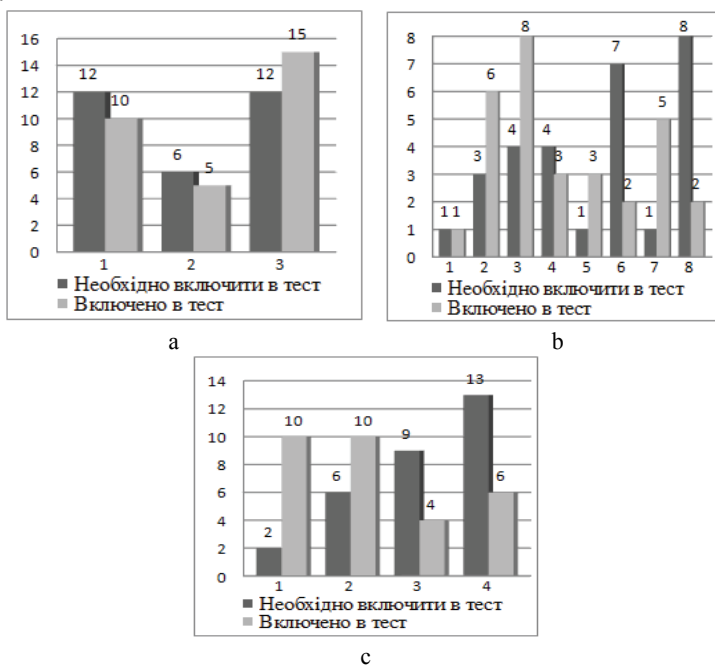


Рис. 1. Приклад відхилення кількості обраних питань від необхідної кількості з дисципліни: а — «Структура та організація даних в ЕОМ»; б — «Людино-машинна взаємодія»; с — «Моделювання та аналіз ПЗ»

## 2. Аналіз складності тесту

Для аналізу складності тесту беремо до уваги відомі у педагогічній практиці рекомендації:

— якщо після проходження тестування більше 10...15 % студентів набрали відмінні оцінки, то тест потрібно ускладнити;

— якщо більше 5...10 % студентів набрали за тест незадовільні оцінки, то тестові питання слід спростити. [4]

Під складністю тестового питання будемо розуміти показник, що характеризує частку студентів, які неправильно відповіли на тестове питання. В педагогічній практиці існують рекомендації з оптимальної відповідності, у процентному співвідношенні, успішних і неуспішних студентів [4]. Тому вважаємо, що складність питання є задовільною, якщо вона знаходиться в межах від 0,15 до 0,9. Питання, в яких складність знаходиться в межах від 0,9 до 1 вважаються складними і рекомендується їх спростити. Питання, в яких складність знаходиться в межах від 0 до 0,15 вважаються легкими і рекомендується їх переробити.

### 3. Аналіз надійності тесту

Аналіз надійності проводиться за допомогою обчислення коефіцієнта кореляції тесту за формулою

$$R_j = \frac{\sum_{i=1}^n (x_{ij} \cdot y_i)}{S_j \cdot S_y} - \bar{x}_j \cdot \bar{y} \cdot \frac{n}{n-1}, \quad (3)$$

де  $R_j$  — коефіцієнт кореляції Пірсона балів між  $j$ -м питанням і сумарними балами усіх студентів;  $S_y$  — стандартне відхилення сумарних балів досліджуваних;  $S_j$  — стандартне відхилення балів усіх досліджуваних по  $j$ -му питанню;  $\bar{y}$  — середній результат по сумарних балах за тест всіх студентів;  $y_i$  — сумарний бал за тест  $i$ -го студента;  $\bar{x}_j$  — середній сумарний бал всіх студентів по  $j$ -му питанню;  $x_{ij}$  — бал, отриманий  $i$ -м студентом по  $j$ -му питанню;  $n$  — кількість студентів.

У разі високого коефіцієнта кореляції студенти, які мають високий бал за тест, відповідали на певне питання правильно, а більшість студентів з низьким балом за тест — відповідали неправильно. У разі низького коефіцієнта кореляції студенти з високим балом за тест помилялися частіше ніж студенти з низьким балом за тест. Таким чином, чим вище значення коефіцієнта кореляції, тим більше ймовірність включення певного завдання в тест. Завдання зі значенням коефіцієнта кореляції менше або рівного 0,4 пропонується виключати з тесту, як ті, що не витримали емпіричної перевірки.

### 4. Аналіз валідності тесту

За запропованою методикою аналізується конструктивна валідність, тобто валідність, спосіб перевірки якої полягає в проведенні продуманих експериментів: аналізу репрезентативності, складності та надійності тесту. Валідність перевіряє наскільки якісно створені питання до тесту.

За запропованою методикою для аналізу валідності тесту проводився розрахунок відхилення оцінок за тест від деякої опорної оцінки для презентативної групи. В якості опорної оцінки можна використовувати середній бал, або оцінки з певного предмету, але отримані в рамках інших форм навчального процесу (РГР, лабораторні роботи тощо). У цій методиці пропонується використовувати середній бал з усіх предметів за весь період навчання.

#### 4.1. Формування презентативної групи студентів

Використання середнього балу в якості опорної оцінки накладає деякі обмеження на склад презентативної групи студентів, для якої буде перевірятися валідність. До складу цієї групи пропонується включати тільки студентів зі стабільним рівнем знань, закон розподілу оцінок яких відповідає нормальному (унімодальному). Така функція має тільки одну моду (один екстремум).

Студенти, у яких спостерігається полімодальний закон розподілу оцінок, в презентативну групу не включаються, тобто рівень оцінок студентів має значний розкид і розрахунок арифметичного середнього для них не є коректним. Така функція має дві або більше мод (тобто два або більше екстремумів).

#### 4.2. Перевірка валідності тесту

За результатами тестування для студентів, обраних в презентативну групу перевіряється валідність тесту, яка проходить в два етапи:

1. Перевірка відхилення оцінок від середнього балу.
2. Розрахунок валідності тесту.

#### Перевірка відхилення оцінок за виконання тесту від середнього балу

Величина відхилення оцінок за виконання тесту від середнього балу для студента визначає якість навчання і дозволяє контролювати рівень знань студентів щодо тестованого предмету та розраховується за формулою

$$O_{ij} = RO_{ij} - RS_{ij}, \quad (4)$$

де  $O_{ij}$  — відхилення оцінок за предмет від середнього балу для кожного студента;  $RO_{ij}$  — оцінка

даного студента за даний предмет;  $RS_{ij}$  — середній бал для даного студента .

Для наочності результатів будується діаграма, що показує відхилення оцінок за тест від середнього балу кожного студента.

Для більшості студентів позитивне значення  $O_{ij}$  означає, що тест є легким і вимагає ускладнення (рис. 2а);

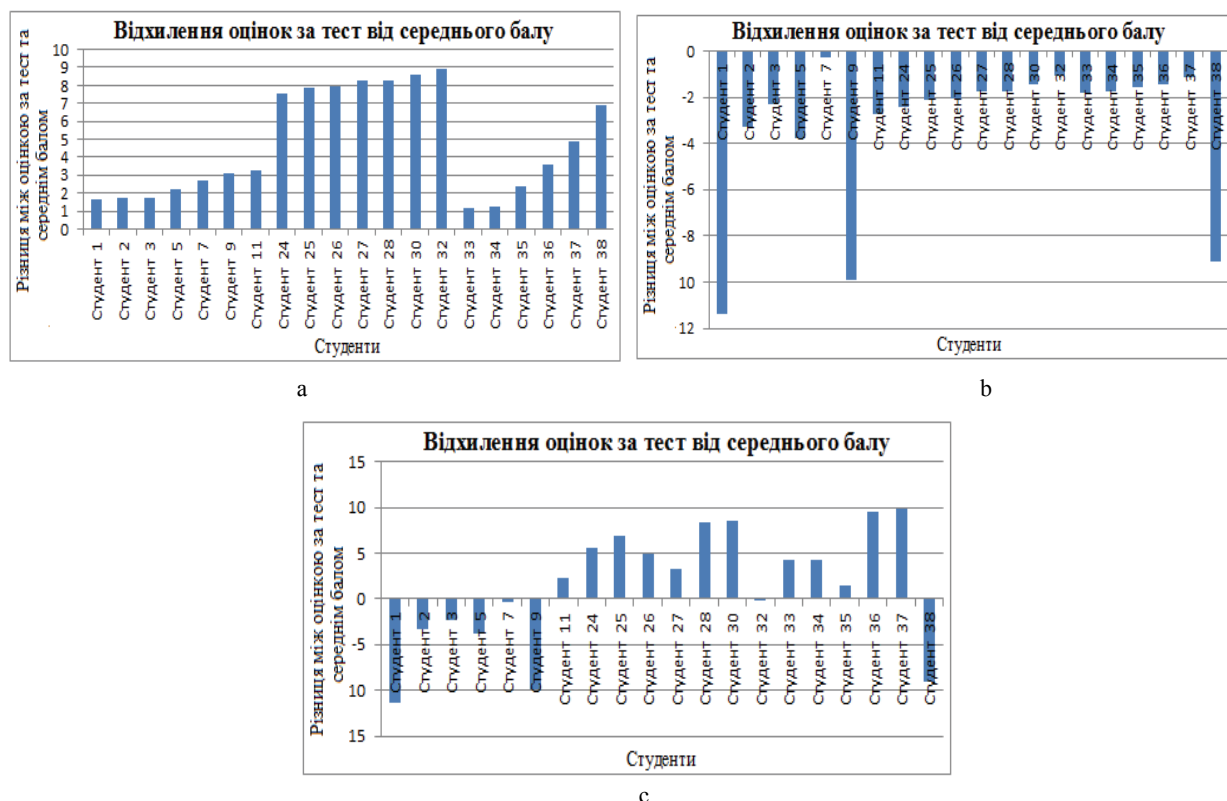


Рис. 2. Діаграма відхилення оцінок по тесту від середнього балу, які необхідно змінити, відповідно до результатів перевірки відхилення: а — позитивне відхилення; б — негативне відхилення; с — відхилення з аналізованого тесту

Для більшості студентів негативних значення  $O_{ij}$  означає, що тест є складним і потребує покращення тестових питань (рис. 2б);

Приблизно половину позитивних і половину негативних значень  $O_{ij}$  означає, що тест не дає надійних результатів і потребує зміни.

### Перевірка валідності тесту

Відхилення оцінок від середнього балу не може дати точну оцінку якості тесту для студентів, відхилення оцінок яких не знаходиться в критичних точках, тобто якщо відхилення у всіх студентів позитивне чи негативне.

Валідність тесту залежить від відхилення оцінок від середнього балу і розраховується за формулою

$$V = \sqrt{\frac{1}{n} \sum_{i=1}^n Q_{ij}^2}, \quad (5)$$

де  $V$  — валідність тесту,  $n$  — кількість студентів у репрезентативній групі.

Чим нижче значення відхилення оцінок кожного студента по тесту, тим нижче, тобто краще значення валідності. Валідність тесту досягає значення нуль, у випадку коли для кожного студента, відхилення оцінки за тест від середнього балу дорівнює нулю. Чим більше значення валідності тесту, тим гірше тест.

Для тесту, діаграма відхилення оцінок яких зображена на рис. 2а, валідність тесту дорівнює 5,5. Для тесту, діаграма відхилення оцінок яких зображена на рис. 2б, валідність тесту дорівнює 4,33. Для тесту з діаграмою відхилення оцінок показаною на рис. 2с, валідність дорівнює 6,4.

## 5. Методика підвищення валідності педагогічного тесту. Рекомендації щодо підвищення якості тесту

Виходячи з вищевикладеного, розрахунок критеріїв якості тесту проводиться з використанням методів теорії ймовірностей та математичної статистики, і проводиться за наступною методикою:

1. Аналіз репрезентативності тесту. Розрахунок кількісних характеристик репрезентативної вибірки питань проводиться у відповідності з обраними статистичними характеристиками курсу за допомогою формули (1). Перевірка репрезентативності тесту розраховується за формулою (2).

2. Аналіз складності тесту і тестових питань проводиться відповідно до пункту 2.

3. Аналіз надійності тесту розраховується за формулою (3).

4. Аналіз валідності тесту. Формування презентативної групи студентів: побудова функції розподілу балів за весь період навчання та вибір студентів, успішність яких підпорядковується унімодальному (нормальному) закону розподілу, проводиться відповідно до пункту 4.1. Розрахунок відхилення тестових оцінок з певної дисципліни від середнього балу проводиться за формулою (4). Валідність тесту розраховується за формулою (5).

Проаналізувавши значення критеріїв якості тесту, визначених за запропонованою методикою, автори для підвищення якості тесту рекомендують:

1. За аналізом репрезентативності тесту — змінити кількість питань, якщо після перевірки ступеня відповідності, тест є нерепрезентативним;

2. За аналізом складності рекомендується переробити тест, якщо більше 10...15 % студентів набрали за тест відмінні оцінки, або якщо більше 5...10 % студентів набрали незадовільні бали. Рекомендується замінити або доопрацювати тестові питання, у яких складність виходить за межі (0,15; 0,9);

3. За аналізом надійності тесту рекомендується прибрати або доопрацювати тестові питання, у яких коефіцієнт кореляції менше або дорівнює 0,4;

4. За аналізом валідності тесту рекомендується для студентів, які входять до презентативної групи, по тестованій дисципліні на підставі діаграми, що показує відхилення оцінок за тест від середнього балу кожного студента, зробити висновки про складність тесту і можливу зміну низки питань.

### Висновки

Розроблено методику оцінки та підвищення валідності педагогічних тестів на базі критеріїв якості тесту, яка складається з аналізу репрезентативності, складності та надійності тесту.

Для перевірки правильності зміни попередніх характеристик використано аналіз валідності тесту.

Методика дозволяє проаналізувати якість розроблених тестів, замінити або виключити проблемні питання, правильно розставити акценти на головних темах в досліджуваному матеріалі, правильно розподілити питання за темами: додати або зменшити кількість питань у тестових темах.

Проаналізовано використання методики підвищення валідності педагогічних тестів для аналізу якості тесту з предмету «Моделювання та аналіз ПЗ» кафедри «Системне програмне забезпечення» для студентів, що навчаються за напрямом 6.050103 «Програмна інженерія».

Створено програмну систему для аналізу педагогічних тестів. Після чого проведено пробне тестування та проаналізовано тест за методикою. Аналіз валідності тесту показав, що з 20 студентів, що потрапили в презентативну групу, троє студентів отримали приблизно на 10 балів нижче їх середнього балу за період навчання, а четверо приблизно на 10 балів вище, ніж їх середній бал. При цьому валідність дорівнювала 6,4. Відповідно до рекомендацій аналізу репрезентативності було змінено кількість питань. За рекомендаціями аналізу складності, були ускладнені 12 питань, і полегшено одне питання. За рекомендаціями аналізу надійності було змінено 15 питань.

Після зміни показників, аналіз валідності показав, що всі студенти, які потрапили в презентативну групу, набрали бали, що не відрізняються від середнього більш ніж на 5 балів. При цьому валідність дорівнювала 3,05.

Таким чином, можемо зробити висновок, що після використання запропонованої методики, модельний тест з дисципліни «Моделювання та аналіз ПО» покращився приблизно в 2 рази і дозволяє оцінити знання із засвоєного матеріалу якісніше.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Чельшкова М. Б., Теория и практика конструирования педагогических тестов уч. пос. / М. Б. Чельшкова. — М : Логос, 2002. — 432 с.

2. Майоров, А. Н. Теория и практика создания тестов для системы образования: Как выбирать, создавать и использовать тесты для целей образования / А. Н. Майоров — М. : Интеллект-Центр, 2002. — 296 с.
3. Теория вероятностей и математическая статистика в задачах : учеб. пос. для вузов / В. А. Ватутин, Г. И. Ивченко, Ю. И. Медведев, В. П. Чистяков. — 2-е изд., испр. — М. : Дрофа, 2003. — 328 с.
4. Мелединек А. Международное общество по инженерной педагогике / А. Мелединек, В. Приходько, В. Жураковский // Высшее образование в России : науч. пед. журн. — 2004. — № 3. — С. 53—60.

Рекомендована кафедрою вищої математики ВНТУ

Стаття надійшла до редакції 9.04.2014

**Городничка Катерина Олександрівна** — магістр кафедри системного програмного забезпечення, e-mail: katherine.gorodnichaya@ukr.net;

**Крисілов Віктор Анатолійович** — д-р техн. наук, професор, завідувач кафедри системного програмного забезпечення;

**Онищенко Тетяна Вікторівна** — старший викладач кафедри системного програмного забезпечення.  
Одеський національний політехнічний університет, Одеса

**K. O. Gorodnycha<sup>1</sup>**  
**V. A. Krisilov<sup>1</sup>**  
**T. V. Onishchenko<sup>1</sup>**

## **Method of estimation and increasing validity of pedagogical test based on the criteria of test quality**

<sup>1</sup>Odessa National Polytechnic University

*Method of increasing the validity of the test by analyzing and changing values of quality control criteria in the learning processes. The proposed method makes it possible to improve the quality of checking the knowledge and skills of students learned material over a certain period of time, without the involvement of experts.*

**Keywords:** blended learning, distance learning, educational tests, assessment of pedagogical tests, criteria of test quality, improving the quality of pedagogical tests, automated control of knowledge, validity of the test, examination of test tasks.

**Gorodnycha Kateryna O.** — M. S. of the Chair of System Software, e-mail: katherine.gorodnichaya@ukr.net;  
**Krisilov Victor A.** — Dr. Sc. (Eng.), Professor, Head of the Chair of System Software;  
**Onishchenko Tetiana V.** — Senior Lecturer of the Chair of System Software.

**Е. А. Городничая<sup>1</sup>**  
**В. А. Крисилов<sup>1</sup>**  
**Т. В. Онищенко<sup>1</sup>**

## **Методика оценки и повышения валидности педагогического теста на базе критериев качества теста**

<sup>1</sup>Одесский национальный политехнический университет

*Рассмотрена методика повышения валидности теста за счет анализа и изменения значений критериев качества контроля знаний в учебном процессе. Предлагаемая методика дает возможность повысить качество проверки знаний и умений студентов по изученному материалу за определенный период времени, без привлечения экспертов.*

**Ключевые слова:** смешанное обучение, дистанционное обучение, педагогические тесты, оценка педагогических тестов, критерии качества теста, повышение качества педагогических тестов, автоматизированный контроль знаний, валидность теста, экспертиза тестовых заданий.

**Городничая Екатерина Александровна** — магистр кафедры системного програмного обеспечения, e-mail: katherine.gorodnichaya@ukr.net;

**Крисилов Виктор Анатольевич** — д-р техн. наук, профессор, заведующий кафедрой системного програмного обеспечения;

**Онищенко Татьяна Викторовна** — старший преподаватель кафедры системного програмного обеспечения