

УДК 303.2,37.091.26

МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ПРАКТИЧНИХ ЗАНЯТЬ ЗА ЗМІСТОВИМ МОДУЛЕМ “НАДІЙНІСТЬ ТЕСТУ”

Янчукова Н.В., Авраменко О.В.

У статті проведено аналіз теоретичних підходів до поняття надійності тесту та проведено класифікацію відповідних практичних завдань.

Ключові слова: надійність тесту, практичні завдання, освітні вимірювання.

В статье проведен анализ теоретических подходов к понятию надежности теста и проведено классификацию соответствующих практических заданий.

Ключевые слова: надежность теста, практические задания, образовательные измерения.

Analysis of theoretical approaches to the concept of test reliability is carried out and appropriate practical tasks are categorized.

Key words: reliability of test, practical tasks, educational measurements.

Вступ. Обрахунок надійності тестової оцінки посідає основну позицію в практичному створенні тестових завдань. Надійність як величина вказує на те, що учасник тестування з плином певного проміжку часу може хоча б приблизно відтворити свій результат і що його отриманий бал з тесту максимально відповідає оцінці реальних знань з деякої галузі. Огляд поняття надійності тесту з різних літературних джерел представлено нижче.

Надійність тесту – це характеристика методики тестування, що показує точність вимірів і стабільність результатів до впливу сторонніх випадкових факторів. Це означає, що невелика зміна умов тестування і стану учасників тестування повинно привести до невеликих змін кінцевого результату. Надійність тесту пов'язана з поняттям стандартної помилки: чим вища надійність, тим менша стандартна помилка вимірів. Іноді помилка вимірів трактується як статистична величина, що показує ступінь відхилення отриманого бала від істинного бала учасника.

Під надійністю (реліабільністю) вимірів розуміється ступінь надійності, або точності, з якою може бути виміряний та чи інша конкретна ознака. Надійність тесту характеризує відтворюваність його результатів. Надійність характеризується коефіцієнтом надійності. Коефіцієнт надійності – це кореляційний коефіцієнт, який показує ступінь збігу результатів тестування, що здійснюється в однакових умовах одним тестом.

З точки зору Л.Крокер і Дж.Алгини [4], надійність тесту розглядається як гарантія того, що результати можуть бути хоча б приблизно відтворені при повторному тестуванні тих самих тестованих у подібних умовах. В практичних умовах надійність – це ступінь, у якому відхилення оцінок тестованих залишаються відносно стабільними при повтор-

ному представленні тестованим того ж самого тесту чи його паралельних форм.

Поняття надійності і валідності педагогічного тесту надзвичайно важливі, оскільки саме вони характеризують тест як інструмент для вимірювання. Тест з невідомими надійністю і валідністю не придатний для вимірів [3].

Надійність та методи її обрахунку розглядаються в багатьох роботах [1–2; 7–8; 12–14], де показаний математичний апарат обчислення коефіцієнтів надійності та представлено його обґрунтування.

Метою нашої статті є висвітлення елементів математичного обґрунтування поняття надійності в теорії освітніх вимірювань та класифікація основних типів завдань з теми “Надійність тесту”, а також вироблення рекомендацій до використання при підготовці магістрів спеціальності 8.18010022 “Освітні вимірювання”.

1. Надійність тесту та методи її визначення. Наведемо основні теоретичні позиції змістового модуля “Надійність тесту” [3–6; 9–11].

Якщо невелика зміна умов тестування і стану випробовуваних призводить до несуттєвої зміни кінцевих результатів, то відповідна характеристика тесту називається його **надійністю**. Надійність тесту з точки зору статистики є величиною, що характеризує, ефективність оцінок тесту.

Для отримання даних для оцінки надійності існують такі методи: 1-й метод – тестування двома паралельними тестами; 2-й метод – повторне тестування одним тестом; 3-й метод – розщеплювання. Метод 1 потребує доведення паралельності тестів, методи 1 та 2 складно реалізуються на практиці, метод 3 потребує дослідження на паралельність частин тесту після розщеплення.

Дисперсійна методика визначення коефіцієнта надійності полягає у виділенні “факторної дисперсії”,

що породжується в даному разі впливом завдань та учасників тестування і "залишкової дисперсії", обумовленої випадковими причинами.

Коефіцієнт надійності визначає долю дисперсії "істинного" балу в загальній дисперсії

$$r = \frac{D\{b^*\}}{D\{b\}} = \frac{D\{b\} - D\{\Delta\}}{D\{b\}},$$

де b^* – "істинний бал"; випадкові величини b – результат тестування, Δ – випадкова помилка, $b^* = b + \Delta$.

Обчислимо величини:

$$S_{\text{загал}} = \sum_{i=1}^n \sum_{j=1}^k b_{ij}^2 - \frac{b^2}{nk},$$

$$S_{\text{учасн}} = \frac{1}{k} \sum_{i=1}^n b_i^2 - \frac{b^2}{nk},$$

$$S_{\text{завд}} = \frac{1}{n} \sum_{j=1}^k b_j^2 - \frac{b^2}{nk},$$

$$S_{\text{ост}} = S_{\text{загал}} - S_{\text{завд}} - S_{\text{учасн}},$$

де

$$b_i = \sum_{j=1}^k b_{ij}, \quad b_j = \sum_{i=1}^n b_{ij}, \quad b_{ij} = \sum_{i=1}^n b_{ij} \sum_{j=1}^k b_{ij}$$

$$\text{Оцінимо дисперсії } D_{\text{учасн}} = \frac{S_{\text{учасн}}}{n-1},$$

$$D_{\text{завд}} = \frac{S_{\text{завд}}}{k-1}, \quad D_{\text{ост}} = \frac{S_{\text{ост}}}{(n-1)(k-1)}.$$

Значення $\sqrt{D_{\text{ост}}}$ дає середню квадратичну помилку вимірів.

– Якщо усі завдання *однакові* для усіх учасників, то дисперсія помилки

$$D_{\varepsilon}\{b^*\} = D_{\text{учасн}} - D_{\text{ост}}, \quad \text{а коефіцієнт}$$

$$\text{надійності } r = \frac{D_{\text{учасн}} - D_{\text{ост}}}{D_{\text{учасн}}}.$$

– Якщо учасникам пропонувалися *різні* завдання,

$$S_{\varepsilon} = S_{\text{загал}} - S_{\text{учасн}}; \quad \hat{D}_{\varepsilon} = \frac{S_{\varepsilon}}{n(k-1)} = 5,58; \quad \hat{\sigma}\{\Delta\} = \sqrt{\hat{D}_{\varepsilon}}$$

$$\text{а коефіцієнт надійності } \hat{r} = \frac{D_{\text{учасн}} - \hat{D}_{\varepsilon}}{D_{\text{учасн}}}.$$

Середнє значення коефіцієнта надійності при всіляких розщеплюваннях тесту на дві половини може бути обчислене формулою К'юдера-Ричард-

$$\text{сона: } r = \frac{k}{k-1} \cdot \frac{D\{b\} - \sum_{j=1}^k p_j q_j}{D\{b\}}, \quad \text{при цьому}$$

саме розщеплювання практично не потрібно. Формула К'юдера-Ричардсона застосовна, якщо елементи a_{ij} матриці відповідей A можуть набувати тільки двох значень 0 та 1. Існує її узагальнення у вигляді так званого коефіцієнта

альфа Кронбаха. Для підвищення надійності можна збільшити кількість завдань, кратність збільшення

$$\text{обчислюється } K = \frac{r_{tk}(1-r_t)}{r_t(1-r_{tk})}, \quad \text{де } r_t - \text{коефіцієнт}$$

надійності до зміни довжини тесту; r_{tk} – коефіцієнт надійності після зміни.

Кореляційна методика оцінювання надійності полягає у вимірюванні коефіцієнту кореляції між результатами двократного тестування (або розщеплення). Коефіцієнтом надійності виступає коефіцієнт кореляції.

Для встановлення відповідності між варіантами тесту або двох частин розщепленого тесту аналізуються матриці відповідей, елементи яких можна інтерпретувати як реалізації дихотомної випадкової величини. Коефіцієнт дихотомної кореляції:

$$\hat{\varphi} = \frac{\bar{xy} - \bar{x}\bar{y}}{(\hat{p}_x \hat{q}_x \hat{p}_y \hat{q}_y)^{1/2}} = \frac{\hat{p}_{xy} - \hat{p}_x \hat{p}_y}{(\hat{p}_x \hat{q}_x \hat{p}_y \hat{q}_y)^{1/2}}$$

Для перевірки гіпотези про відсутність кореляції між даними двох тестів у припущенні нормального розподілу використовують критерій Стюдента.

Коефіцієнт надійності r повного тесту пов'язаний з коефіцієнтом r' надійності в

$$\text{результаті розщеплення тесту } r = \frac{2r'}{1+r'}, \quad \text{що є}$$

частковим випадком формули Спірмена-Брауна

$$r = \frac{\lambda r'}{1+(\lambda-1)r'}, \quad \text{де } \lambda - \text{відношення числа}$$

завдань тесту до числа завдань після розщеплення.

Ретестова надійність тесту – це ступінь збереження рангових позицій випробовуваних, визначається на порядковій і номінальній шкалах. Початковими даними слугують результати повторного тестування випробовуваних за допомогою одного і того ж тесту. Кількісною мірою є коефіцієнти рангової кореляції Спірмена або Кендалла.

Коефіцієнт рангової кореляції Спірмена

$$\hat{r} = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n d_i^2$$

є коефіцієнтом лінійної кореляції Пірсона між рангами $r\{x\}$ і $r\{y\}$, тут випадкові величини x

та y – результати повторного тестування або розщеплення тесту: $d_i = r\{x_i\} - r\{y_i\}$.

Рангова кореляція оцінюється також коефіцієнтом Кендалла τ , для чого треба підрахувати в

стовпчику ранг y_i кількість R_i рангів, що перевищують $r\{y_i\}$ та сума $R = \sum_{i=1}^{n-1} R_i$. Вибірковий коефіцієнт рангової кореляції Кендалла і його дисперсія знаходяться за формулами:

$$\hat{r} = \frac{4R}{n(n-1)} - 1, \quad D\{\hat{r}\} = \frac{2(2n+5)}{9n(n-1)}$$

При рівні значущості α та конкуруючій гіпотезі $\tau \neq 0$ критичне значення

$$\tau_{\text{крит.}} = U_{\text{крит.}} \sqrt{\frac{2(2n+5)}{9n(n-1)}}$$

де $U_{\text{крит.}}$ визначається умовою

$$\Phi(U_{\text{крит.}}) = \frac{1-\alpha}{2},$$

функція Лапласа $\Phi(u)$; n – об'єм вибірки. Якщо

$|\hat{\tau}| < \tau_{\text{крит.}}$, то немає підстав для відмови від нульової гіпотези, і рангова кореляція визнається незначущою. Якщо $|\hat{\tau}| > \tau_{\text{крит.}}$, то зв'язок слід вважати значущим.

2. Аналіз основних типів практичних завдань. Для модуля “Надійність тесту” дисципліни “Класичні тестові моделі”, що викладається майбутнім магістрам спеціальності 8.18010022 “Освітні вимірювання”, створено посібник практичних завдань, укладачі якого спирались на такі відомі підручники та посібники, як [4; 10–11]. Головною метою створення посібника є показати, що всі висновки, які випливають з теоретичних питань, можуть бути математично обґрунтовані або доведені, навчити правильної інтерпретації отриманих статистичних даних та надати основний теоретичний матеріал, математичний апарат та розв'язані теоретико-практичні типові задачі для навчання майбутніх фахівців з освітніх вимірювань різнобічної оцінки розроблених тестів або тих, що існують.

Серед практичних завдань можна виділити три основні типи: теоретичні завдання, математичні завдання на аналіз тесту, математичні завдання на виведення співвідношень.

Застосування практичних **завдань теоретичного типу** дозволяє систематизувати пройдений теоретичний матеріал, розвивати логічне мислення і розуміння ситуацій, що можуть виникати в процесі створення завдань та під час тестування і впливати негативно чи позитивно на результати. В цьому разі будь-який вплив відхиляє отримані результати від істинних оцінок і тому вважається негативним фактором. Наприклад, вплив систематичних і випадкових помилок на результати тестування або зміна коефіцієнту надійності залежно від того, якій репрезентативній вибірці він запропонований, або від зміни часу, виділеного на виконання тесту.

Наведемо два приклади теоретичних завдань.

Завдання 1. Для кожної ситуації вкажіть тип оцінки надійності, що найбільш підходить наступним умовам.

А. Єдина форма стандартизованого тесту навчальних досягнень охоплює ділянки інформатики з допомогою дихотомічно оцінюваних завдань за трьома розділами: використання пакету MSOffice, адміністрування операційної системи Windows та програмування мовою Pascal. Питання, що краще: виставити єдину оцінку чи окремі оцінки з кожної ділянки змісту субтестів?

В. Шкільний учитель хоче перевірити на учнях свого класу надійність підсумкового тесту з біології, що містить завдання з множинним вибором.

С. Психолог-консультант розробляє два набори атитюдів, один з яких призначений для використання як попередній тест, а другий – як вихідний тест для програми, що триває один день.

Д. Соціальний психолог розробляє інструментарій для оцінки згоди з атитюдами студентів, що тільки прийняті до коледжу, з метою виявлення їхнього ставлення до сумісного навчання і проживання в гуртожитку. Кожна відповідь буде оцінюватися відповідно до шкали: відповідь “так” – 2 бали, “можливо” – 1 бал, “ні” – 0 балів.

Завдання 2. Порівняйте відмінності в методах поділу навпіл і методах підрахунку коваріації завдань для оцінювання надійності щодо вимог до даних і поясніть, чому вони дають різний результат.

Метою використання таких завдань є закріплення знань теоретичного матеріалу з розділу “Процедури для оцінювання надійності”, а також понятійного апарату та можливості аналізувати і робити висновки із поданих ситуацій. Передбачається володіння такими основними поняттями:

- процедури, що потребують двох пред'явлень тесту (метод взаємозамінюваних форм, що потребує конструювання двох подібних форм тесту і пред'явлення обох форм одній і тій же групі тестованих, метод тест-ретест (ретестовий метод), що потребує пред'явлення однієї форми тесту групі через певний проміжок часу, та комбінація двох методів – тест-ретестовий метод з застосуванням взаємозамінюваних форм (у цьому разі процедура оцінки надійності включає представлення першої форми тесту групі тестованих, а через деякий проміжок часу – другої форми));

- процедури, що потребують одноразового пред'явлення тесту (метою використання таких процедур є визначення узгодженості між собою результатів одних і тих же тестованих за всіма завданнями тесту);

- процедури, за якими розділяють тест на взаємозамінювані форми (метод оцінки кореляції (тобто оцінюється кореляція між двома формами тестових завдань і за її величиною формують висновок про паралельність даних тестових форм), метод, в основі якого лежить використання передбачуваних тестових оцінок (розробник тесту повинен виявити джерела помилки вимірів, які вносять найбільше перешкод у корисну інтерпретацію оцінки, і планувати дослідження надійності таким чином, щоб ефекти впливу цих помилок могли бути оцінені у вимірюваннях));

- помилки, що впливають на кінцевий результат: випадкові (помилки, що впливають на оцінку тестованого через різні непередбачувані фактори) та систематичні (помилки, що постійно впливають на оцінку індивіда через деякі специфічні характеристики тестованого чи тесту). Перший тип зменшує як відтворюваність, так і практичну повноцінність тестових оцінок, другий тип є причиною зниження точності тестових оцінок, тобто скорочує їх практичну корисність.

Другий тип завдань на аналіз надійності тесту є **завдання математичного типу**. У них завдання ставиться так, що спочатку потрібно розрахувати

певні величини, а потім інтерпретувати їх залежно від ситуації. Наприклад, розрахунок всіх представлених у теоретичній частині коефіцієнтів надійності і порівняльний аналіз отриманих результатів, вибір оптимального значення. Нижче наведемо характерні завдання такого типу.

Завдання 3. Початковий тест містить 40 тестових завдань і має коефіцієнт надійності (розрахований за формулою KR20), що дорівнює 0,85. Розробник тесту розглядає можливість зменшення кількості тестових завдань на 5. Оцініть, як це вплине на надійність тесту.

Завдання 4. Тест, що складається з 10 завдань, був пред'явлений 9-ти учням. Матриця, що об'єднує результати за завданнями і учнями, має такий вигляд:

Учень	Завдання										заг. оцінка
	1	2	3	4	5	6	7	8	9	10	
A	1	0	1	1	1	0	0	0	1	0	5
B	1	1	1	1	1	0	0	0	0	0	5
C	0	1	1	1	1	1	1	0	1	0	7
D	0	0	1	0	0	0	1	0	0	0	2
E	1	0	0	1	1	1	0	0	1	0	5
F	1	1	1	1	1	0	1	1	1	0	8
G	1	0	0	1	1	1	1	0	0	1	6
H	0	0	0	1	0	0	1	0	0	0	2
I	1	1	1	0	0	1	1	0	0	0	5
	6	4	6	7	6	4	6	1	4	1	

Потрібно дати відповіді на наступні питання:

A. Оцінити внутрішню узгодженість тесту, використовуючи метод поділу навіл (створити два субтести з парних і непарних завдань). Обрахувати дві оцінки, що були б отримані для кожного учня за субтестами.

B. Знайти кореляцію між субтестами.

C. Яка оцінка надійності для тесту повної довжини?

D. Сказати, користуючись результатами візуального аналізу, який із субтестів, на Вашу думку, буде мати більш високу оцінку надійності, обґрунтувати відповідь.

E. Який з методів поділу тесту забезпечить найбільше наближення до оцінки коефіцієнта точності? Обґрунтувати відповідь.

F. Яка величина коефіцієнта надійності за формулами KR20 і KR21?

G. Пояснити, чому відрізняються отримані величини коефіцієнтів за формулами KR20 і KR21.

Завдання 5. Експерт цікавиться питанням: що більше впливає на рівень уваги студентів вищого навчального закладу: стать чи вік. Розглядаються студенти різних факультетів і форм навчання. Результати наведено в таблиці:

Студенти					
Чоловіки			Жінки		
від 18 до 25 років	від 26 до 35	старші 35	від 18 до 25 років	від 26 до 35	старші 35
16	15	13	17	15	12
17	16	14	15	17	10
15	13	13	16	14	10
16	14	10	16	14	9

Використавши двофакторний дисперсійний аналіз даних, визначити коефіцієнт генералізації, що вказує на систематичний вплив одного із факторів (стать, вік)

на рівень уваги студента. Проаналізувати отримані результати. Зробити висновки.

Завдання 6. Наступна таблиця має дані за двома гіпотетичними субтестами стандартизованої батареї тестів навчальних досягнень з використання різних типів програмного забезпечення, кожен з яких має дві взаємозамінювані форми. Потрібно дослідити дані, наведені в таблиці, і відповісти на питання.

Характеристики	Використання комерційного програмного забезпечення	Використання безкоштовного програмного забезпечення
Форма А		
Середнє	30	28
Стандартне відхилення	8	12
Число завдань	50	55
KR20	0,83	0,85
Форма В		
Середнє	32	28
Стандартне відхилення	6	12
Число завдань	50	55
KR20	0,8	0,85

A. Які стандартні помилки обрахунків для двох форм тестів? Чи завжди краще використовувати тестову форму з найменшою стандартною помилкою вимірів?

B. Нехай оцінка студента X – 20 балів за математичним тестом обрахунків за формою А. Побудуйте інтервал, який імовірно містить його істинну оцінку (з 68%-ним шансом достовірності). Які були б границі інтервалу при 95%-ній впевненості в тому, що інтервал містить його істинну оцінку?

C. За формою В математичного тесту обрахунків студент Y отримав оцінку 10 балів, а студент Z – 29 балів. Які значення їх можливих істинних оцінок?

В цих завданнях, в першу чергу, перевіряється знання студентом математичних формул, що відповідають теоретичному матеріалу, та вміння аналізувати обраховані результати. В наведених вище прикладах розглядаються різні види коефіцієнтів надійності (стабільності, еквівалентності, внутрішньої узгодженості), що вибираються залежно від основних джерел помилок (зміна тестованих з плином часу, вибірка перевірки змісту від форми до форми, вибірка перевірки змісту чи виявлення некоректних завдань), та проводиться їх обрахунок і аналіз доцільності використання в цій ситуації.

Завдання такого типу дають можливість підкріплювати свої твердження відповідними математичними показниками, що є важливим фактором для студентів з нематематичною базовою освітою.

До **математичних завдань на виведення співвідношень** віднесемо завдання на доведення тверджень чи виведення формул. Такі завдання необхідні для розвитку і розширення математично-статистичної бази студентів. Нижче наведені характерні завдання названого типу.

Завдання 7. Вказівка з тестування для індивідуального використання має оцінки надійності (метод поділу тесту), що розраховані за формулою:

$$\hat{\rho}_{XX} = \frac{4\hat{\rho}_{AB}\hat{\sigma}_A\hat{\sigma}_B}{\sigma_X^2}$$

де А, В — оцінки за половинами тесту. Довести, що ця формула еквівалентна формулі для коефіцієнта Рюлона.

Завдання 8. Покажіть, що коваріація між отриманими за двома тестами оцінками еквівалентна коваріації між їх істинними оцінками, тобто доведіть, що $\sigma_{X_1 X_2} = \sigma_{T_1 T_2}$.

Завдання 9. На основі формули KR20 виведіть формулу для коефіцієнта надійності KR21.

Наведені вище приклади орієнтовані на якісне знання студентів математичного апарату обрахунку надійності, формул і означень елементарної статистики та класичної моделі істинної оцінки. Вони не є типовими і вимагають від магістранта логічного мислення, знання основних формул і методів їх виведення, аналізу отриманих результатів.

Кінцевою метою застосування всіх трьох типів завдань є, з одного боку, виробити у магістранта навички практичного аналізу тесту на надійність та закріпити теоретичні знання цього матеріалу, а з

іншого боку – виявити ступінь засвоєння ним поданого теоретичного матеріалу.

Висновки. Проведено аналіз теоретичних підходів до поняття надійності тесту та здійснено класифікацію відповідних практичних завдань. Серед практичних завдань виділено три основні типи: теоретичні завдання, математичні завдання на аналіз тесту, математичні завдання на виведення співвідношень. Наведено характерні завдання кожного із трьох названих типів.

Відмічено, що завдання першого типу спрямовані на покращення розуміння понятійного апарату та можливості аналізувати і робити висновки із реальних ситуацій. Завдання другого типу дають можливість підкріплювати свої твердження відповідними математичними показниками, що є важливим фактором для магістрантів з нематематичною базовою освітою. Завдання третього типу вимагають від магістранта логічного мислення, знання основних формул і методів їх виведення, аналізу отриманих результатів.

Література

1. Гарвілл Л. М. Стандартна похибка вимірювання (навчальний модуль Національної Ради США з оцінювання у сфері освіти) / Л. М. Гарвілл // Вісник. Тестування і моніторинг в освіті. – 2010. – № 3. – С. 40–48.
2. Горох В. П. Порівняльний аналіз шкалювання результатів ЗНО різними методами / В. П. Горох, О. Ю. Соколов // Вісник. Тестування і моніторинг в освіті. – 2010. – № 12. – С. 22–28.
3. Ким В. С. Тестирование учебных достижений : монография / В. С. Ким. – Уссурийск : Издательство УГПИ, 2007. – 214 с.
4. Крокер Л. Введение в классическую и современную теорию тестов : учебник / Л. Крокер, Дж. Алгина. – Логос, 2010. – 668 с.
5. 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, 2001. – 286 с.
6. Нейман Ю. М. Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. – Москва, 2000. – 168 с.
7. Раков С. А. Теоретичні засади шкалювання результатів ЗНО методом еквіпроцентної нормалізації / С. А. Раков // Вісник. Тестування і моніторинг в освіті. – 2010. – № 12. – С. 11–22.
8. Фрарі Р. Б. Обчислення балів за формулою для тестів множинного вибору (коригування вгадування) / Р. Б. Фрарі // Вісник. Тестування і моніторинг в освіті. – 2008. – № 9. – С. 32–40.
9. Чельшкова М. Б. Теория и практика конструирования педагогических тестов : учебное пособие / М. Б. Чельшкова. – М. : Логос, 2001. – 432 с.
10. Baker F. V. The Basics of Item Response Theory / F. V. Baker. – Portsmouth NH : Heinemann Educational Books, 1985. – 131 p.
11. Hambleton R. K. Fundamentals of Item Response Theory / R. K. Hambleton, H. Swaminathan, H. J. Rogers. – Newbury Park, CA : Sage, 1991.
12. Rogosa D. R. Demonstrating the reliability of the difference score in the measurement of change / D. R. Rogosa, J. B. Willet // Journal of Educational Measurement , 20, 335–343.
13. Stanley J. C. Reliability / J. C. Stanley // Educational Measurement Reliability / R. Thorndike (ed.). – 2-nd ed. – Washington, D.C : American Council on Education, 1971. – P. 356–442.
14. Traub R. E. Understanding reliability. Educational measurement: Issues and practice / R. E. Traub , G. L. Rowley. – 1991. – № 10 (1). – P. 37–45.