

УДК 378.26:004.94(045)

КОНСТРУЮВАННЯ ТЕСТІВ НАВЧАЛЬНИХ ДОСЯГНЕНЬ СТУДЕНТІВ — МАЙБУТНІХ ВЧИТЕЛІВ ІНФОРМАТИКИ

Івлієва Ольга Михайлівна,

*доцент кафедри математики, інформатики та інформаційної діяльності
Ізмаїльського державного гуманітарного університету, канд. пед. наук,
olgaiivlieva@mail.ru.*



Анотація. Представлено метод критеріально-орієнтованого тестування як інструмент визначення навчальних компетенцій майбутніх учителів інформатики. Розглянуто особливості створення тестів, процедури їх проведення у формі комп'ютерного тестування. Висвітлено процедуру вивчення валідності тесту. Узагальнено результати перевірки якості тестів.

Ключові слова: тестування, конструювання тестів, оцінка якості тестів, валідність, надійність, навчання вчителів інформатики.

На сучасному етапі розвитку вітчизняної вищої школи та її інтеграції в європейський освітній простір проблема діагностики навчальних досягнень студентів у закладі вищої освіти набуває особливої актуальності. Великою перевагою тестового іспиту є уніфікація вимог, застосування єдиного критерію і норм оцінювання, економія часу як тих, хто складає іспит, так і викладачів.

На формування української тестологічної школи значний вплив мали роботи А. Анастасі [1] й П. Клайна [2], які, по суті, стали вже класикою. Ці праці дали поштовх для інтеграції психологічної науки і методів тестування. Дослідники, в основному, дійшли згоди щодо алгоритму створення тестів досягнень, виділяючи окремим етапом роботу над оцінкою якості тестів.

Метою розпочатого дослідження є побудова методичної схеми конструювання і визначення якості тестів, створених для вивчення навчальних досягнень студентів — майбутніх учителів інформатики, з математичних дисциплін; підбір інструментарію для характеристики валідності й надійності тестів.

Історично склалися дві основні теоретичні теорії конструювання й використання педагогічних тестів.

- Перший на основі класичної теорії тестів (Classical Test Theory — СТТ).
- Другий у рамках теорії латентно-структурного аналізу, сучасна теорія конструювання тестів (Item Response Theory — IRT), яку правильніше було б назвати методологією, оскільки це теорія розробки і застосування методів дослідження емпіричних даних (теорія методів).

Очевидно, виникає проблема вибору того чи іншого підходу до розробки і застосування тестів. Теорія LSA або правильніше оснований на ній IRT (сучасна теорія тестування) більш ефективна, але складніша у використанні. Необхідно застосовувати потужний математичний апарат, розробляти дорожче і більш складне програмне забезпечення. Використання IRT потребує високої кваліфікації як від розробників, так і від користувачів тестів. У IRT не ставляться і не вирішуються фундаментальні проблеми валідності і надійності тесту: тест там заздалегідь вважа-

ється надійним і валідним. Обчислення IRT зводяться до отримання оцінок параметрів складності завдання і до вимірювання рівня знань тих, що беруть участь у випробуванні.

До досягнень IRT відносять використання однієї шкали у вимірах значень параметрів випробуваних і завдань тесту. Це дозволяє співвідносити рівні компетентності випробуваних з мірою складності тестових завдань. На цій можливості спільних оцінок параметрів випробуваних і завдань побудований застосовуваний більшістю тестологів математичний апарат. Додамо також, що застосування IRT (сучасної теорії тестування), більш ефективною на великих масивах досліджуваних, на наш погляд, не є доцільним на невеликих групах студентів: необхідно застосовувати потужний математичний апарат, розробляти дорожче і більш складне програмне забезпечення. Використання IRT потребує високої кваліфікації як від розробників, так і від користувачів тестів.

Отже, в умовах невеликих груп і для досягнення мети перевірки сформованості навчальних компетенцій студентів, доцільним є використання класичних тестів.

Розглядаючи класифікацію тестів досягнень, можна виділити два підходи, які в даний час склалися в тестуванні: тести, орієнтовані на критерій (критеріально-орієнтовані), і тести, орієнтовані на норму (нормативно-орієнтовані). Ці два підходи виділяють зарубіжні й вітчизняні дослідники в галузі розробки тестів досягнень (Аванесов, 1998; Анастасі, Урбіна, 2005) [1, 3].

1. Критеріально-орієнтований підхід. Завдання — перевірити, наскільки учень (студент) засвоїв необхідний обсяг знань, умінь і навичок, який виступає як заданий стандарт або критерій засвоєння. За даного підходу головним завданням є прагнення з'ясувати, — які елементи змісту навчальної дисципліни засвоєні тим чи іншим випробуваним. При цьому визначається — що з генеральної сукупності завдань ... випробуваний знає і що не знає. Інтерпретація результатів ведеться педагогами, мовою навчальної дисципліни. [3]. Вважається, що достатньо не надто великої кількості завдань з галузі, обмежених конкретним стандартом або рівнем (критерієм) знань. Теоретичне під-

ґрунтя застосування критеріально-орієнтованих тестів у навчанні складають роботи В. П. Беспалька, у яких висвітлено питання про класифікацію рівнів навчання (засвоєння) залежно від характеру діяльності, що виконується [4].

2. Нормативно-орієнтований підхід. За такого підходу вирішується завдання не стільки оцінити повноту знання, як визначити місце (рейтинг) кожного з досліджуваних. До тесту відбирається така мінімально достатня кількість завдань, яка дозволяє порівняно точно визначити, образно кажучи, не «хто що знає», а «хто знає більше». Інтерпретація результатів тестування ведеться переважно на мові тестології [3]. Вважається, що цей тест доцільно використовувати для створення рейтингового списку під час вступу до вищих навчальних закладів, нагородження успішних у навчанні тощо.

На думку вчених, для педагогічного контролю більш звичними і доцільними є завдання, які розв'язуються саме за критеріально-орієнтованого підходу (наприклад, [5]). У започаткованому дослідженні використовували саме критеріально-орієнтовані тести. Вважалося, що створення тестів у межах критеріально-орієнтованого підходу ґрунтується на визнанні планомірного характеру навчання, його протіканні відповідно із заздалегідь сформульованими цілями. За точку відліку у визначенні успішності в таких тестах приймається не характеристика виконання завдань всіма учнями, а конкретна галузь змісту навчальної діяльності. Це дозволяє, як зауважує А. Анастасі, здійснювати інтерпретацію з акцентом на те, «що індивід може зробити і що він робить, а не те, як він виглядає на тлі інших» [3, с. 93].

Отже, у започаткованому дослідженні ми виходили з розуміння критеріально-орієнтованих тестів як типу тестів, призначених для визначення рівня індивідуальних досягнень відносно деякого критерію, здійснюваного на основі логіко-системного аналізу змісту завдань. Критеріально-орієнтовані тести задовольняють таким вимогам:

- конкретний і детальний опис цілі навчання;
- наявність для кожної цілі адекватного набору завдань;
- розподіл завдань за ступенями складності/важкості.

Зважаючи на суттєве зауваження А. Анастасі про те, що критеріально-орієнтоване тестування, якщо воно проводиться за всіма правилами, краще всього пристосоване для визначення основних навичок у математиці [3, с. 93], ми застосовували цей різновид тестів під час вивчення студентами вузівських дисциплін математичного циклу.

Рішенням Вченої Ради Ізмаїльського державного гуманітарного університету з 2012/13 навчального року семестрові екзамени проводяться у формі комп'ютерного тестування.

На підставі аналізу літературних джерел і власного багаторічного досвіду розробки і проведення оціночних тестових процедур, було запропоновано такий список етапів розробки тестів досягнень: 1) визначення цілей тестування; 2) визначення ресурсних можливостей розробників; 3) вибір організаційно-техно-

логічного рішення; 4) відбір змісту навчального матеріалу; 5) конструювання технологічної матриці; 6) складання тестових завдань або банку завдань; 7) експертиза, можлива доробка завдань, які не пройшли експертизу; 8) формування вибірки для апробації завдань і тестів; 9) компоновання завдань для апробації; 10) апробація тестових завдань і варіантів; 11) визначення і розрахунок показників якості тестових завдань; 12) відбраковування невдалих завдань і складання тесту (робочого банку завдань); 13) можлива доробка (корекція) вибракуваних завдань; 14) апробація варіантів тесту; 15) визначення і розрахунок показників якості тесту; 16) складання остаточного варіанта тесту (банку й алгоритму формування варіантів на його основі); 17) стандартизація тесту (комплекс процедур і заходів, що дозволяє створити для всіх випробовуваних рівні умови). 19) нормування тесту (процес отримання кількісних або якісних характеристик інструментарію, на основі порівняння з якими можна дати оціночне судження); 20) оснащення тесту (видання в буклетній або електронній формі з усіма супутніми інструктивними матеріалами).

Було складено набір тестових завдань з навчальних дисциплін, зокрема, з математичного аналізу, теорії ймовірностей та математичної статистики, для студентів, що здобувають освітній рівень бакалавра галузі знань 0403 Системні науки та кібернетика напряму підготовки 6.040302. Інформатика*.

Уведення підсумкового тестування потребувало певних змін у викладанні: студентів необхідно готувати до такого іспиту вже в процесі навчання, і тому, паралельно з підготовкою підсумкового тестового іспиту, проводився тестовий контроль знань відповідних тем.

Особливості тестових програм, складених спеціалістом кафедри математики, інформатики та інформаційної діяльності Ізмаїльського державного гуманітарного університету О. В. Коваленком, передбачали використання завдань з вибором однієї правильної відповіді з чотирьох запропонованих (одноваріантний вибір), з вибором кількох правильних відповідей з чотирьох запропонованих (поліваріантний вибір), тест з відкритою відповіддю, яка представлена числом і запитанням з двома варіантами відповідей («так», «ні»). Під час проведення тестування студент отримує варіант тесту, складеного з 30 запитань, отриманих генерацією випадкових чисел.

Так, вибір запитань здійснюється за поданими нижче принципами.

1. Існує двовимірний масив таблиці (масив) запитань, кожне з яких має свій унікальний номер. Попередня версія програми працювала в припущенні про однакову «вагу» різних типів запитань для оцінювання. Зараз зусилля розробників спрямовані на забезпечення можливості встановлення «вагового коефіцієнту» завдань.

2. Для створення індивідуального списку завдань з масиву вибирається випадковий номер запитання і записується в таблицю із завданням. Наступний номер повинен бути унікальним, тому його номер порівнюється з попередніми, вже обраними, і якщо він не співпадає з обраними раніше, записується таким.

Як результат, для кожного студента є набір номерів запитань, які йому були задані, і список його відповідей. Кількість правильних відповідей автоматично не підраховується. Відсутність автоматичного підрахунку балів або відсотків підвищує стійкість системи до зовнішнього впливу для спотворення результатів. Підрахунком успішності займається окремий модуль, який і формує звіт про результати кожного зі студентів академічної групи.

Аналіз запропонованих видів тестових завдань дозволяє стверджувати, що тим самим забезпечені основні потреби кількості завдань, вибору як форми відповіді студента, так і можливостей використання запитань різного ступеня складності.

Однією з основних вимог критеріально-орієнтованого тестування є включення до тестів завдань, різних за ступенем складності. Взнявши за основу шкалу рівнів засвоєння В. П. Беспалька, ми конструювали завдання трьох рівнів складності.

Як правило, завдання першого рівня передбачали відтворення означень, властивостей математичних об'єктів, застосування елементарних алгоритмів, правил розрахунків, наприклад:

1. Зв'язок між поняттями **події** та **випробуванням** є таким: (оберіть правильну відповідь)

- а) подія — результат випробування.
- б) випробування — результат події.
- в) зв'язку між цими поняттями немає.
- г) подія — вид випробування.

2. Послідовність $\{x_n\}$ називається **обмеженою**, якщо існують такі числа m і M , що для $\forall n \in \mathbb{N}$ виконується:

- а) $m \leq \{x_n\} \leq M$;
- б) $m \leq x_n \leq M$;
- в) $m \leq \{x_n\} \leq M$;
- г) $m \leq |x_n| \leq M$.

Завдання другого рівня вимагали вмінь, інтерпретації, порівняння, узагальнення, застосування теоретичних знань у стандартних ситуаціях. Інтерпретуюча діяльність студента спрямована на те, щоб глибше проникнути у суть явища, пояснити його, осмислити з різних точок зору. На цьому рівні діяльність є продуктивною, яку можна характеризувати через ступінь оволодіння уміннями застосовувати засвоєну інформацію у практичній сфері для розв'язання деякого класу задач, зокрема на основі використання зразка. Прикладами завдань другого рівня можуть бути:

1. Скільки елементів містить така множина:

$\{\{x\}, x, \{x, \{x\}\}\}$ _____

2. Границі

- а) $\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = e$;
- б) $\lim_{x \rightarrow 0} (1 + x)^{\frac{1}{x}} = e$;
- в) $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$;

називають відповідно:

а) друга особлива границя; б) друга особлива границя; в) перша особлива границя;

а) перша особлива границя; б) перша особлива границя; в) друга особлива границя;

а) друга особлива границя; б) перша особлива границя; в) перша особлива границя;

а) перша особлива границя; б) друга особлива границя; в) перша особлива границя.

Творча активність студентів була спрямована на створення нових продуктів діяльності. Найбільш характерна риса творчої діяльності — вироблення принципово відмінних від засвоєних програм розв'язань і дій. Завдання третього рівня складності вимагали від студентів творчого підходу до розв'язання.

Відтак, робота зі студентами організовувалася так, щоб була представлена можливість перевірки рівня сформованості всіх видів діяльності — від репродуктивної до діяльності продуктивної.

Як уже доповідалося раніше, під час підготовки до тестування був складений набір тестових завдань з курсу математичного аналізу (180 запитань) та теорії ймовірностей (200 запитань).

Після виконання перших етапів роботи зі складання тестів з математичних дисциплін, була організована робота з їх апробації. Пробне тестування переслідувало такі цілі:

- виявлення завдань, у яких є недоліки (невідповідність складності завдань рівню підготовленості випробовуваних; незрозумілі або двозначні формулювання завдань, виявлення непрацюючих дистракторів у завданнях закритої форми та ін.);
- визначення статистичних характеристик тестових завдань і тесту в цілому.

Для проходження тесту на етапі первинної апробації для кожного з учасників (випробовуваних) генерувався варіант, що складається з 30 запитань. У дослідженні на етапі апробації взяли участь 42 осіб (студенти 2–4-их курсів факультету управління, адміністрування та інформаційної діяльності ІДГУ), у другій серії в цілому 135 студентів.

Необхідно зазначити, що вже на етапі апробації якість складених тестових матеріалів була досить високою: так, наприклад, значення коефіцієнта кореляції r_{xy} , (де x — кількість балів за результатами поточного контролю, що включав оцінку роботи студента на практичних заняттях, виконання поточних і модульних контрольних робіт, самостійну роботу; y — результат тестування — кількість балів, набраних під час виконання тестів), склало 0,76 ($r_{xy} = 0,76$).

Після повного циклу апробації первинного тесту приблизно 14 відсотків завдань мали текстові і змістові зміни. Наступна версія банку завдань складала 320 завдань з математичного аналізу і 312 запитань з курсу теорії ймовірностей та математичної статистики. З банку тестових запитань з теорії ймовірностей було складено два паралельних тести.

Далі проведена серія тестувань і статистична обробка результатів. Практика показує, що під час роботи над створенням педагогічного тесту найбільш трудомісткими є етапи 17–20, оскільки збір статистичних даних, зазвичай, займає багато часу, а вивчення яко-

сті тесту вимагає проведення додаткових обчислень. Як і в попередніх дослідженнях, згідно з класичним підходом до визначення якості тестів, основними її показниками виступали валідність і надійність.

Існує кілька видів валідності, обумовлених особливостями діагностичних методик. У багатьох роботах (напр., А. Анастасі), називаються найчастіше валідність за змістом за одночасністю, прогностична і ретроспективна валідність.

Для тестів вимірювання навчальних досягнень на перший план висувається вимога валідності за змістом. *Валідність за змістом* закладалася у тести вже під час вибору завдань. Тести склалися відповідно до діючих навчальних програм підготовки фахівців. Попри це, отримана оцінка експертів на відповідність запропонованого варіанту тесту меті, завданням та обсягу перевірки, а також вимогам, що висуваються до якості тестових матеріалів. У роботі взяли участь 3 експертів — співробітники кафедри математики, інформатики та інформаційної діяльності ІДГУ.

Валідність за одночасністю, або *поточна валідність*, зазвичай, визначається за допомогою зовнішніх критеріїв, за якими інформація збиралася одночасно з експериментом. Зовнішніми критеріями виступали показники успішності результатів:

- поточного контролю студента, виражений оцінкою (бали) — x_1 ;
- підсумкового контролю — x_2 ;
- академічної успішності з предмету (поточний контроль+тест) — x_3 ;
- середній бал залікової книжки — x_4 .

Критерієм валідності, аналогічно процедурі попереднього оцінювання якості тестів, виступав коефіцієнт кореляції $r_{x_1x_2}$, де x_1 — кількість балів за результатами поточного контролю, що включав оцінку роботи студента на практичних заняттях, виконання поточних і модульних контрольних робіт, самостійну роботу; x_2 — результат тестування — кількість балів, набраних під час виконання тестів.

Були опрацьовані матеріали дослідження тестової і поточної успішності 42 студентів, що взяли участь у тестуванні. Наприклад, проаналізовано результати тестування з теорії ймовірностей та математичної статистики (22 студенти). Результати представлені на рис. 1.

Як видно з рисунка, значення коефіцієнта кореляції $r_{x_1x_2}=0,74$, що є непоганим показником валідності (зазначимо, що, на думку багатьох дослідників, результат тестування можна вважати валідним при значенні коефіцієнта кореляції, більшому ніж 0,6). Аналіз кореляційного зв'язку результатів виконання тесту показав, що найкраще змінна x_2 — результат тестування, кількість балів, набраних під час виконання тестів, корелює зі змінною x_3 — академічна успішність з предмету: $r_{x_2x_3}=0,84$.

Як показник *прогностичної валідності* тесту був обраний коефіцієнт кореляції між кількістю балів, отриманих під час вивчення дисципліни (поточна успішність+тестовий бал) та інтегральним

D24 =КОРРЕЛ(С3:С23;D31+D3:D23)							
перевірка якості тестів.xls [Режим совместимости]							
	A	B	C	D	E	F	G
1			Кільк.бал.пот	Кільк.бал.підс	Загальна	Сер.бал зал.кн	
2			x1	x2	x3	x4	
3	Арсеньев І		42	12	54	78,8	
4	Арутюнян А		42	18	60	76,6	
5	Бринза А.		64	22	86	87,3	
6	Бринзой А.		47	18	65	73,3	
7	Веремешко А.		47	19	66	81,8	
8	Гавріляк А.		63	23	86	89,5	
9	Головльов В.		42	18	60	70,9	
10	Гребенюк Н.		50	21	71	79,1	
11	Жарко О.		49	18	67	78,3	
12	Івлєв О.		66	24	90	84,7	
13	Карпенко С.		42	19	61	77,1	
14	Кірюглю м.		42	20	62	72,4	
15	Кудрявський В.		43	18	61	73,2	
16	Рагнев А.		42	19	61	75,9	
17	Репіна А.		65	22	87	89,6	
18	Рошу Н.		68	25	93	84,6	
19	Рябчук В.		43	24	67	78,4	
20	Сафта В.		42	19	61	71,4	
21	Трикуліч С.		42	20	62	68,7	
22	Чернева Ю.		47	19	66	77,7	
23	Шаріфова Р.		69	26	95	92,2	
24		$r_{x_1x_2}$		0,736145932			
25		$r_{x_2x_3}$		0,837993755			
26		$r_{x_2x_4}$		0,602019213			
27		$r_{x_3x_4}$		0,860130829			

Рис. 1. Результати перевірки якості тестів

показником успішності студента — загальною академічною успішністю. Вибір показника обумовлений доведенням О. О. Чумакова про їх найвищу кореляцію (гіпотеза підтверджена на рівні значущості 0,01) [6, с. 22]. Експериментальне значення коефіцієнта кореляції $r_{x_3x_4}=0,86$.

Ретроспективна валідність тесту не визначалась. (Вона визначається на основі критерію, що відображає події або стан якості в минулому).

У традиційній тестології термін «надійність» означає відносну сталість, стійкість, узгодженість результатів тесту при первинному і повторному його застосуванні на одних і тих же випробовуваних. При цьому певною мірою можуть збігатися як самі результати, так і порядкове місце (ранг), займане випробуваним у групі. І в тому, і в іншому випадку в разі повторення дослідів можливі деякі розбіжності, але важливо, щоб вони були незначними, у межах однієї групи. Тобто, надійність методики — це критерій, який дозволяє судити про те, наскільки вселяють довіру отримані результати.

Надійність тесту також залежить від кількості тестових завдань. За рекомендаціями спеціалістів, для забезпечення достатньої надійності підсумкового контролю тест має містити не менше 40 завдань. Наступна робота із забезпечення якості тестів буде проводитися в напрямку збільшення кількості завдань.

Після проведення серії тестувань планується статистична обробка результатів для з'ясування числових характеристик надійності тестів.

Особливо важливою характеристикою в розробці педагогічних тестів є дискримінативна здатність завдання, так як від неї значною мірою залежить валідність тесту. Ця характеристика показує, наскільки ефективно тестове завдання розрізняє студентів, які опанували і не опанували навчальний ма-

теріал. Існує чимало показників для розрізнення здібностей, деякі з них досить складні для обчислення. Ці показники, у своїй більшості, вимагають для розрахунку дві серії вимірювань — повторно-тестування однієї групи учнів або проведення тесту на двох різних групах.

Під час розробки тесту для однієї або невеликої кількості груп учнів найзручніше отримати дві серії вимірювань шляхом формування контрастних груп. Викладач вибирає з групи студентів тільки тих учнів, про яких він може точно стверджувати, що вони оволоділи або не оволоділи навчальним матеріалом. Ті, що опанували матеріал, складають «високу» контрастну групу, а не опанували — «низьку». Студенти, що знаходяться в проміжному положенні, не включаються до контрастних груп. Найпростіший і широко відомий показник розрізняльної здатності завдання по відношенню до навчання D обчислюється як різниця між часткою випробовуваних з «високої» групи, що правильно виконала завдання, і часткою випробовуваних з «низької» групи, які також правильно виконали завдання.

Надалі було запропоновано використовувати інший показник розрізняльної здатності завдання $P(X)$, який вважається більш досконалим, ніж D з ряду теоретичних міркувань. Показник $P(X)$ можна розглядати як ймовірність узгодженості між результатом виконання випробовуваним завдання і віднесенням випробованого до «високої» або «низької» контрастної групи. Він розраховується за формулою: $P(X) = n_1/N_1 + n_2/N_2$, де N_1 і N_2 — кількість випробовуваних, які потрапили відповідно в «високу» і «низьку» контрастні групи; n_1 — кількість досліджуваних з «високою» групи, що правильно виконали завдання; n_2 — кількість досліджуваних з «низької» групи, що неправильно виконали завдання. Найкращі завдання матимуть значення $P(X)$ рівні одиниці. Мінімальне значення показника досягається в тому випадку, якщо між віднесенням випробованого до однієї з груп і виконанням ним завдання не існує ніякого зв'язку. Обчислене нами значення показника на основі характеристик виконання всього тесту 22 студентами експериментальної групи становить 0,44.

Отже, складені і застосовані тести з математичних дисциплін для перевірки навчальних досягнень майбутніх учителів інформатики вивчають сформованість репродуктивної, продуктивної та творчої складових набутих знань з предметів математичного циклу. Складені тести пройшли первинну апробацію, що дозволило покращити якість тестових завдань. Після проведення основної сесії тестування перевірена якість тестів. Основними показниками якості тестів виступали валідність та надійність. Окремо вивчалася дискримінативна властивість запропонованих тестів. У ході дослідження використовувалися методи експертних оцінок тестових завдань; порівняння результатів проходження тестування респондентами з їх об'єктивними показниками успішності; показник ретестової надійності; обчислення коефіцієнта лінійної кореляції Пірсона; метод контрастних груп.

Отже, робота з конструювання тестів з математичних дисциплін може проводитися згідно із запропонованою схемою (етапи 1–20). Доцільно використовувати тестові завдання різних форм, й обов'язково — різного рівня складності. Після створення першого варіанту тестів необхідним є проведення роботи з їх апробації, що дозволить відібрати якісні завдання і відбракувати або покращити якість інших. Про якість тестів можна робити висновки, обчисливши показники: коефіцієнт кореляції результатів виконання тесту з академічною успішністю з предмету — показник поточної валідності, коефіцієнт кореляції результатів виконання тесту із середнім балом залікової книжки — показник прогностичної валідності. Показник $P(X)$ виражає розподільну здатність тесту як засіб забезпечення валідності. Змістова валідність забезпечується грамотним відбором матеріалу й оцінками експертів.

Проведене дослідження дозволяє стверджувати, що розроблені тести є валідними, об'єктивними, надійними, мають значну розподільну властивість, значить, якісно вимірюють досліджувану якість і можуть бути використаними у подальшій роботі.

* * *

Ивлиева О. М. Конструирование тестов учебных достижений студентов — будущих учителей информатики

Анотация. Представлен метод критериально-ориентованного тестирования как инструмент определения учебных компетенций будущих учителей информатики. Рассмотрены особенности создания тестов, процедура их проведения в форме компьютерного тестирования. Освещена процедура изучения валидности теста. Обобщены результаты проверки качества тестов.

Ключевые слова: тестирование, конструирование тестов, оценка качества тестов, валидность, надежность, обучение учителей информатики.

* * *

Ivliyeva O. M. Constructing tests of educational achievements for students — future IT-teachers

The article deals with the method of criterion-oriented testing as the implementation for determining the educational competencies of future IT-teachers. The peculiarities of creation such tests and the procedure of their taking in the form of computer testing are considered. The procedure of defining the validity of the test is described. The results of appraisal the quality of the tests were generalized.

Key words: testing, constructing of tests, results of appraisal the quality of tests, validity, reliability, teaching IT-teachers.

Література

1. Аванесов В. С. Композиция тестовых заданий: учебная книга для преподавателей вузов, учителей школ, аспирантов и студентов педвузов / В. С. Аванесов. — [2 изд., испр. и доп.]. — М. : Адепт, 1998. — 217 с.
2. Клайн П. Справочное руководство по конструированию тестов. Введение в психометрическое проектирование : пер. с англ. — К. : Малое научн. внедр. предприятия «ПАН ЛТД», 1994. — С. 283.
3. Анастаси А. Психологическое тестирование / А. Анастаси, С. Урбина. — СПб. : Питер, 2005. — 688 с.
4. Беспалько В. П. Слагаемые педагогической технологии / В. П. Беспалько. — М. : Педагогика, 1989. — 192 с.
5. Булах І. Є. Створюємо якісний тест : навчальний посібник / І. Є. Булах, М. Р. Мруга. — К. : Майстер-клас, 2006. — 160 с.
6. Чумаков А. А. Методика конструирования тестов профессиональных достижений с использованием Интернет-технологий : автореф. дис. ... канд. псих. наук / Чумаков А. А. ; МГУ им. М. В. Ломоносова. — М, 2007. — 24 с.