

**Володимир Вікторович Kamiшин,**

кандидат технічних наук,
лауреат Державної премії УРСР
в галузі науки і техніки,
лауреат Державної премії України
в галузі освіти,
директор Інституту обдарованої дитини
НАПН України
м. Київ, Україна

УДК 37.091.2:004.415.539:005.6

ФОРМУВАННЯ АБСОЛЮТНОЇ ШКАЛИ КВАЛІМЕТРІЇ РІВНІВ НАВЧАЛЬНИХ ДОСЯГНЕНЬ З УРАХУВАННЯМ СКЛАДНОСТІ ТЕСТОВИХ ЗАВДАНЬ

Внедрение объективного тестового контроля в образовательных учреждениях всех уровней аккредитации, в том числе во время независимого внешнего оценивания знаний выпускников школ, увеличивает актуальность формирования абсолютных 100-бальной и 200-бальной шкал. Опираясь на распределение значений индекса дискриминативности заданий теста, предложена процедура установления его перцептивно-продуктивных, репродуктивных, конструктивно-вариативных и творческих составляющих, каждая из которых имеет соответствующий коэффициент значимости. И поскольку каждый вопрос может рассматриваться как отдельный показатель эффективности, то агрегированная оценка определяется с помощью аддитивной функции.

Ключевые слова: абсолютные шкалы, индекс дискриминативности вопросов теста и критерии их сложности, показатель эффективности, аддитивная функция.

Implementation of objective test monitoring in educational institutions of all accreditation levels including during independent external estimation of graduates knowledge of schools, increases the formation urgency of absolute 100 balls and 200 balls scales. Based on the distribution of values of the determination's index tests, a procedure to establish its perceptually-productive, reproductive, structurally variable and creative components, each of which with the help of the method of setting priorities matched up with the significance coefficient (desirability). Because each question of the test will be considered as separate performance indicator, aggregated score shall be determined by the usual additive functions. Because each question of the test will be considered as separate performance indicator, aggregated score shall be determined by the usual additive functions.

Key words: absolute scales, indexes of discriminatory test questions and criteria of their complexity, efficiency index, adaptive function.

Застосування об'єктивного тестового контролю рівнів навчальних досягнень тих, хто навчається, порушує питання педагогічного, ціннісного, етичного та психологічного характеру [1]. Педагогічний аспект стосується двох принципових напрямів у розв'язанні питань об'єктивності кваліметрії.

Перший напрям – традиційний, пов'язаний з намаганням досягти об'єктивності шляхом формування колегіальної оцінки (наприклад, державною екзаменаційною комісією), адже нормативно встановлено, що третє складання екзамену з певної навчальної дисципліни студентом, який двічі не зміг скласти його науково-педагогічному працівнику, відбувається на комісії [2]. Проте оцінки, що отримуються таким чином потрібно умовно вважати об'єктивними, адже відомо, що підсумовування декількох суб'єктивних думок не завжди відображає об'єктивний стан справ [3, 4]. Тому

такі оцінки названі інтерсуб'єктивними [1]: вони можуть бути більш чи менш наближені до об'єктивної залежно від складу комісії. При нерівноцінному складі тієї ж державної атестаційної комісії (ДЕК) думка одного фахівця може бути більш об'єктивною, ніж думки інших, але не прийнята ними, якщо у групі діє складова ірраціональної групової поведінки – конформізм. З іншого боку, більш впливовий член комісії (посада, наукові ступені, звання тощо), який за суб'єктивними причинами негативно ставиться до випробуваного, може сприяти необ'єктивній оцінці знань. Роль такого лідера у психологічній теорії прийняття рішень є одним з пояснень виникнення ризику у груповій поведінці [3, 4].

Другий напрям забезпечення об'єктивності кваліметрії рівня навчальних досягнень студентів є більш ефективним, тому що стосується використання



стандартного тестового контролю на персональній електронно-обчислювальній машині (ПЕОМ) і відповідає пріоритетним завданням Міністерства освіти і науки України [5]. Застосування таких тестів, ефективно розв'язують питання *етичного* та *ціннісного* аспектів педагогічної кваліметрії рівнів навчальних досягнень тих, хто навчається. Проведення об'єктивного тестового контролю передбачає застосування 100-бальної чи 200-бальної (на етапі зовнішнього незалежного оцінювання – ЗНО) шкал кваліметрії, яким притаманні властивості абсолютної шкали вимірювань [6–13]. Дослідження [14–17] підтверджують, що ця властивість є основою створення оціночних систем та придатною для 100-бальної шкали.

Сьогодні більшість досліджень об'єктивного тестового контролю присвячена питанням створення тестів (формулювання питань та варіантів відповідей, оцінювання їхньої надійності та валідності) та встановлення закону розподілу абсолютних даних тестування [7; 18] і не стосується проблем формування 100-бальної чи 200-бальної шкал, не зважаючи на те, що до Болонських домовленостей 100-бальна шкала застосовувалася для оцінювання знань студентів у Києво-Могилянській академії та Донбаському гірничо-металургійному інституті [19]. Проведений аналіз свідчить про те, що завдання МОН України на врахування складності окремих модулів навчальних дисциплін в інтегральній оцінці рівня навчальних досягнень студентів виконуються не повною мірою [5].

Таким чином, проблема формування абсолютних шкал кваліметрії рівня навчальних досягнень тих, хто навчається, є актуальною.

З аналізу наукових досліджень на сьогодні в Україні оцінки за 100-бальною шкалою формуються при об'єктивному тестовому контролі рівня навчальних досягнень тих, хто навчається, шляхом співвіднесення правильних відповідей і загальної кількості тестових запитань:

$$n = \left[\frac{N^+}{N} \cdot 100 \right], \quad (1)$$

де N^+ – кількість правильних відповідей;
 N – загальна «довжина» тесту.

Вважається, що правильність відповіді на кожне питання оцінюється за бінарною системою «правильно/неправильно», внаслідок чого респондент отримує відповідні оцінки «1» або «0»:

$$x_{ij} = \begin{cases} 1 - & \text{якщо відповідь правильна} \\ 0 - & \text{у протилежному випадку} \end{cases}, \quad (2)$$

де x_{ij} – оцінка відповіді j -го випробуваного на i -те запитання тесту.

Відповідно до вимог МОН України, кількість модулів навчальних дисциплін коливається в межах 2–4. Якщо маємо максимальне їх число (тобто, 4), тоді «довжина» тесту утворюється через «довжину» питань, що діагностують знання кожного модулю:

$$N = N_I + N_{II} + N_{III} + N_{IV}, \quad (3)$$

де $N_I, N_{II}, N_{III}, N_{IV}$ – кількість запитань, що оцінюють знання студентів з відповідного модуля навчальних дисциплін.

Якщо студент правильно відповів на певну кількість питань, нескладно визначити загальний рівень його навчальних досягнень за 100-бальною шкалою, провівши обчислення:

$$\begin{aligned} PND(N) &= \frac{1}{4} (PND_I + PND_{II} + PND_{III} + PND_{IV}) = \\ &= \left[\frac{N_I^+}{N_I} + \frac{N_{II}^+}{N_{II}} + \frac{N_{III}^+}{N_{III}} + \frac{N_{IV}^+}{N_{IV}} \right] \cdot \frac{100}{4} \end{aligned} \quad (4)$$

Необхідність врахування складності окремих питань і модулів навчальних дисциплін декларується, але ж не розв'язується.

У працях [15; 17] проблема формування 100-бальної шкали при об'єктивному тестовому контролі розглядається з позицій системного аналізу та теорії прийняття рішень як однокрокова задача прийняття рішень з векторним показником ефективності. Кожне питання тесту розглядається, як показник ефективності. Оскільки «довжина» тесту $N \gg 2$, то за методологією системного аналізу маємо задачу прийняття рішення з векторним показником ефективності. За таких умов отримання інтегральної оцінки, якій належить *системна властивість емерджентності*, буде відбуватися за допомогою адитивної функції агрегування, де враховується складність модуля і питання його меж:

$$PND(N) = \sum_{i=1}^m \alpha_i \cdot PND(N_i) = \sum_{i=1}^{m=4} \alpha_i \cdot \left(\sum_{j=1}^{N_i} \beta_{ij} \cdot x_{ij} \right) \quad (5)$$

де α_i – коефіцієнт відносної «вагомості» (складності, ваги) i -го модуля тесту ($i=1, \dots, 4$), що охоплюється частиною, «довжиною» N_i запитань;

x_{ij} – оцінка правильності відповіді на j -те запитання i -го модуля тесту, що виставляється згідно (2);

β_{ij} – коефіцієнт відносної «вагомості» j -го запитання ($j=1, \dots, N_i$) i -го модуля тесту;

Коефіцієнти α_i, β_{ij} мають бути «зважені», тобто, задовольняти такі умови:

$$\begin{cases} 0 \leq \alpha_i \leq 1, & \sum_{i=1}^4 \alpha_i = 1 \\ 0 \leq \beta_{ij} \leq 1, & \sum_{j=1}^{N_i} \beta_{ij} = 1 \end{cases} \quad (6)$$

При застосуванні виразу (5) виникають певні труднощі. *По-перше*, визначено, що для оцінювання рівня навчальних досягнень студентів «довжина» екзаменаційного тесту з курсу навчальної дисципліни має складатись із 380–420 завдань. Встановлення коефіцієнтів «ваги» для кожного модуля не викликає труднощів, а знаходження коефіцієнтів «ваги» для приблизно 100 запитань всередині кожного модуля викликає труднощі. *По-друге*, не існує методів визначення складності завдань за критеріями перцептивно-продуктивності, репродуктивності,



конструктивно-варіативності та творчості, що мають оцінюватись диференційовано. Йдеться про встановлення не суб'єктивно-психологічної, а об'єктивної (статистичної) складності завдань. *По-третє*, може скластися ситуація, де перевірка рівня навчальних досягнень за важливістю модуля буде здійснюватися за допомогою менш складних запитань.

Метою статті є розробка науково обгрунтованого методу і рекомендацій формування абсолютних шкал тестової кваліметрії знань з чітким визначенням критеріїв завдань перцептивно-продуктивного, репродуктивного, конструктивно-варіативного та творчого характеру.

З аналізу наукових джерел випливає, що ефективним для визначення можливості завдань тесту диференціювати студентів за рівнем навчальних досягнень є застосування для аналізу точково-бісеріального коефіцієнта кореляції, розповсюдженого на практиці створення психологічних тестів, де він названий *індексом дискримінативності* і обчислюється за такою формулою:

$$r_i = \frac{\bar{x}_i^+ - \bar{X}}{\sigma_i} \sqrt{\frac{N_i^+}{N - N_i^+}}, \quad (7)$$

де r_i – індекс дискримінативності i -го завдання тесту, $i=1, 2, \dots, n$;

N – загальна кількість випробуваних за допомогою тесту студентів;

N_i^+ – кількість студентів, які успішно виконали i -те завдання тесту;

σ_i – середнє квадратичне відхилення результатів випробування N студентів за n завданнями тесту;

\bar{x}_i^+ – середнє значення позитивних результатів виконання студентами i -го завдання тесту здійснюється за бінарною системою (2) і обчислюється:

$$\bar{x}_i^+ = \frac{N_i^+}{N}, \quad (8)$$

де \bar{X} – середнє значення показників успішності виконання N студентами n завдань тесту:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n \bar{x}_i^+ = \frac{1}{n \cdot N} \sum_{i=1}^n N_i^+ \quad (9)$$

Індекс дискримінативності r_i змінюється у межах інтервалу $[-1, +1]$, оцінює міру відповідності кожного завдання тесту і є показником дискримінативності. Позитивні значення r_i свідчать про ефективність певного тестового завдання та демонструють різницю результатів тих, хто проходить випробування, а негативні – про непридатність певного завдання тесту, його невідповідність підсумковому результату. Отже, індекс дискримінативності r_i є показником критеріальної валідності завдання, оскільки визначається стосовно зовнішнього критерію – підсумкового результату. Тому можна зробити висновок, що існують функції визначення складності кожного завдання.

Маємо валідний тест, застосування якого для перевірки рівня навчальних досягнень студентів з навчальної дисципліни проведено на вибірці випробуваних. Маючи результати тестування, нескладно проранжувати відповідні завдання в порядку спадання r_i (рис. 1). Розбивши інтервал ранжування на чотири частини, які завдання тесту виявляють випробуваних, а які – ні,

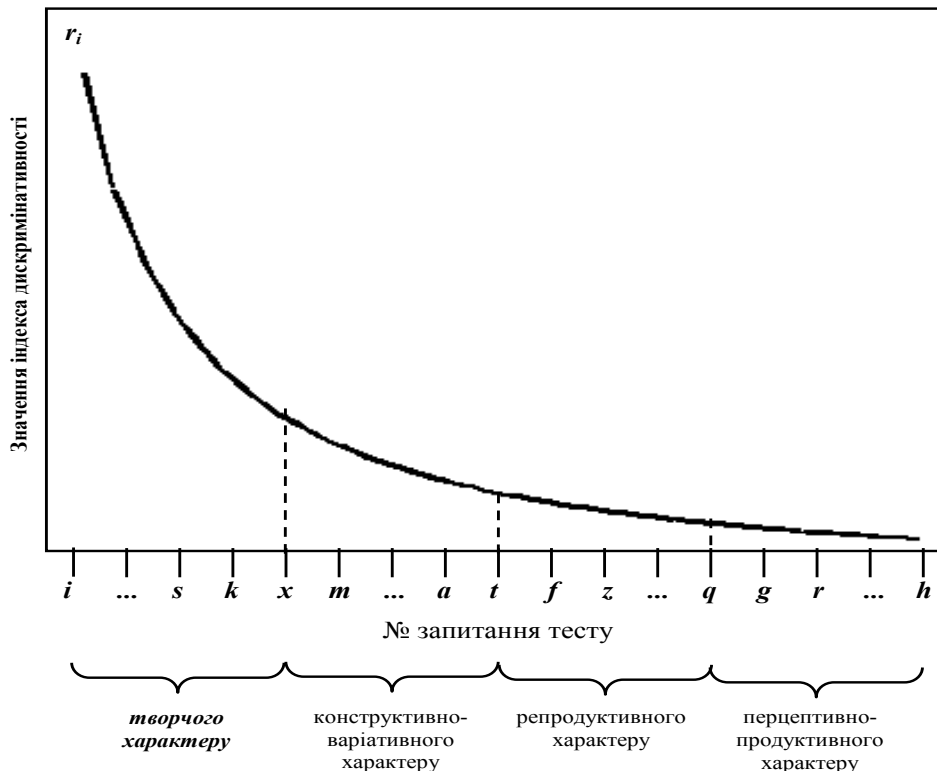


Рис. 1. Впорядкування завдань тесту за гіпотетичним графіком зміни індексу дискримінативності для різних запитань тесту



тобто, за графіком на рис. 1 покажемо їх складність. Характеристика стосується запитань тесту незалежно від модуля навчальної дисципліни. Таким чином, реалізується процедура встановлення статистичної (об'єктивної) складності завдань.

Для кожної категорії виявленої складності питань тесту, користуючись математичним методом, нескладно ранжувати пріоритети і обчислити відповідні коефіцієнти складності цих питань (табл. 1).

Для кожної категорії виявленої складності питань тесту, користуючись математичним методом, нескладно ранжувати пріоритети і обчислити відповідні коефіцієнти складності цих питань (табл. 1).

З табл. 1 бачимо, що залежно від поставлених задач об'єктивного тестового контролю можна обирати значення коефіцієнтів складності, обчислені з точністю до 0,01, 0,001 чи 0,0001, що відповідає третій, сьомій та сімнадцятій ітерації застосування методу ранжування пріоритетів.

Таким чином, формули (4) і (5) перетворюються на такі:

$$PHD(N) = \alpha_I (\beta_{nn} N_{nn}^I + \beta_p N_p^I + \beta_{\kappa \epsilon} N_{\kappa-\epsilon}^I + \beta_m N_m^I) + \alpha_{II} (\beta_{nn} N_{nn}^{II} + \beta_p N_p^{II} + \beta_{\kappa \epsilon} N_{\kappa-\epsilon}^{II} + \beta_m N_m^{II}) + \alpha_{III} (\beta_{nn} N_{nn}^{III} + \beta_p N_p^{III} + \beta_{\kappa \epsilon} N_{\kappa-\epsilon}^{III} + \beta_m N_m^{III}) + \alpha_{IV} (\beta_{nn} N_{nn}^{IV} + \beta_p N_p^{IV} + \beta_{\kappa \epsilon} N_{\kappa-\epsilon}^{IV} + \beta_m N_m^{IV}) \quad (10)$$

де α_i – коефіцієнт важливості i -го модуля, значення якого обираються з табл. 1;

$\beta_{nn}, \beta_p, \beta_{\kappa-\epsilon}, \beta_m$ – коефіцієнт складності (відповідно перцептивно-продуктивного, репродуктивного, конструктивно-варіативного та творчого характеру) j -го запитання i -го модуля, значення якого також вибираються з даних табл. 1;

$N_{nn}^i, N_p^i, N_{\kappa-\epsilon}^i, N_m^i$ – кількість запитань відповідної складності у i -му модулі навчальної дисципліни, з якої відбувається тестування.

Таким чином, можемо зробити висновки.

1. Інтегральна оцінка рівня навчальних досягнень з навчальної дисципліни або окремого модуля має важливу системну властивість емерджентності та дозволяє робити відповідні висновки щодо навченості студентів.

2. Процедура отримання інтегральної оцінки рівня навчальних досягнень студентів при застосуванні об'єктивного тестового контролю і переведення його у 100-бальну шкалу шляхом відповідного нормування розглянута з позицій системного аналізу і теорії прийняття рішень. Визначено, що такі завдання мають бути класифіковані як однокрокова задача прийняття рішення з векторним показником ефективності.

3. Уперше, використовуючи показники індексів дискримінативності завдань тесту, встановлено об'єктивні статистичні критерії виявлення їхньої складності, а саме: перцептивно-продуктивний, репродуктивний, конструктивно-варіативний та творчий зміст.

4. Користуючись математичним методом ранжування пріоритетів, встановлено «зважені» коефіцієнти важливості (складності) питань, а також коефіцієнти складності модулів навчальних дисциплін. Це надало змогу подати узагальнену оцінку тестування за 100-бальною шкалою, як розв'язання однокрокової задачі прийняття рішень з векторним показником ефективності. Розв'язання цієї задачі прийняття рішень відбувається за допомогою адитивної функції агрегування.

Виходячи з вищевказаного, можна зробити узагальнений висновок про розширення методології системно-інформаційних досліджень кваліметрії знань. Подальші дослідження питань формування 100-бальної шкали кваліметрії рівнів навчальних досягнень із застосуванням об'єктивного тестового контролю доцільно проводити з використанням поданих пропозицій щодо оцінювання знань тих, хто навчається, з конкретних навчальних дисциплін.

Таблиця 1

Коефіцієнти складності запитань тесту

№	Характер запитання тесту	Коефіцієнт складності, отриманий на ітерації №									
		I	II	III	IV	V	VI	VII	VIII	...	XVII
1	2	3	4	5	6	7	8	9	10	...	19
1.	Творчий	0,4375	0,5682	0,6563	0,7167	0,7599	0,7920	0,8168	0,8364	...	0,9171
2.	Конструктивно-варіативний	0,3125	0,2955	0,2604	0,2278	0,2007	0,1786	0,1605	0,1456	...	0,0783
3.	Репродуктивний	0,1875	0,1136	0,0729	0,05	0,0362	0,0273	0,0213	0,0171	...	0,0045
4.	Перцептивно-продуктивний	0,0625	0,0227	0,0104	0,0056	0,0033	0,0021	0,0014	0,0010	...	0,0001



Використані літературні джерела

1. *Петровский А. В.* Основы педагогики и психологии высшей школы [Текст] / А. В. Петровский, В. М. Ковалева, А. А. Крашеников [и др.]; Под ред. А. В. Петровского. – М.: МГУ, 1986. – 304 с.
2. Положення про організацію навчального процесу у вищих навчальних закладах [Текст]. Затверджено наказом Міністерства освіти України від 2 червня 1993 р. № 161.
3. *Козелецкий Ю.* Психологическая теория решений [Текст]: Пер. с польск. Г. Е. Минца, В. Н. Поруса / Ю. Козелецкий; Под ред. Б. В. Бирюкова. – М.: Прогресс, 1979. – 504 с.
4. *Рева О. М.* Колективні рішення у невеликій групі авіаційних операторів [Текст]: конспект лекцій з курсу «Основи теорії прийняття рішень» / О. М. Рева. – Кіровоград: ДЛІАУ, 1998. – 34 с.
5. Модернізація вищої освіти України і Болонський процес [Текст] / Уклад. М. Ф. Степко, Я. Я. Боллобаш, К. М. Левківський, Ю. В. Сухарніков; Відп. ред. М. Ф. Степко. – К.: Освіта України, 2004. – 60 с.
6. *Супес П.* Основы теории измерений [Текст] / П. Супес, Р. Зинес // Психологические измерения. – М.: Мир, 1967. – С. 9–110.
7. *Гласс Дж., Стенли Дж.* Статистические методы в педагогике и психологии [Текст]: Пер. с англ. Л. И. Харусовой / Дж. Гласс, Дж. Стенли; общ. ред. Ю. П. Адлера. – М.: Прогресс, 1976. – 496 с.
8. *Пфанцагль И.* Теория измерений [Текст]: Пер. с англ. / И. Пфанцагль. – М.: Мир, 1976. – 248 с.
9. Надежность и эффективность в технике: Справочник в 10 т. [Текст] / Под общ. ред. В. Ф. Уткина, Ю. В. Крючкова. – М.: Машиностроение, 1988. – Т. 3. Эффективность технических систем. – 328 с.
10. *Перегудов Ф. И.* Введение в системный анализ [Текст]: Учеб. пособ. / Ф. И. Перегудов, Ф. П. Тарасенко. – М.: Высшая школа, 1989. – 367 с.
11. *Циба В. Т.* Основы теории кваліметрії [Текст]: Навч. посібн. / А. Т. Циба. – К.: ІЗМН, 1997. – 160 с.
12. *Федієнко В. В.* Шляхом Болонського процесу: Порівняльний аналіз ефективності шкал вимірювання і оцінювання знань [Текст] / В. В. Федієнко // Наукові праці академії. – Вип. ІХ. – Кіровоград: ДЛІАУ, 2005. – С. 212–232.
13. *Камишин В. В.* Системно-інформаційний аналіз ефективності шкал кваліметрії академічної обдарованості / В. В. Камишин // Проблеми інформатизації та управління: Збірник наукових праць. – Вип. 2 (42). – К.: НАУ, 2013. – С. 45–55.
14. *Рева О. М.* 12 балів: український компроміс європейської «полегшеної шкали оцінювання» [Текст] / О. М. Рева, О. Ф. Штанько, І. А. Добрянський // Вища школа: Наук.-практ. видання. – К., 2005. – № 4. – С. 40–55.
15. *Рева О. М.* Формування 100-бальної шкали кваліметрії знань студентів ВНЗ як однокрокова задача прийняття рішень з векторним показником ефективності [Текст] / О. М. Рева, В. В. Федієнко // Педагогіка. Психологія. Медико-біологічні проблеми фізичного виховання і спорту: Наукова монографія / Под ред. проф. С. С. Єрмакова. – Х.: Харківська державна академія дизайну і мистецтв, 2006. – № 11. – С. 98–102.
16. *Рева О. М.* Шляхом Болонського процесу: 100-бальна шкала – універсальна основа створення різноманітних оціночних систем [Текст] / Н. О. Василенко, О. М. Рева, В. В. Федієнко // Наукові записки: Зб. наук. статей Кіровоградського державного педагогічного університету імені Володимира Винниченка. – Вип. 74. – Кіровоград: КДПУ імені В. Винниченка, 2007. – С. 43–57.
17. *Камишин В. В.* Формування абсолютних шкал тестової кваліметрії знань вирішенням однокрокової задачі прийняття рішень з векторним показником ефективності / В. В. Камишин // Проблеми інформатизації та управління: Збірн. наук. праць. – Вип. 4 (40). – К.: НАУ-друк, 2012. – С. 24–29.
18. *Суходольский Г. В.* Основы математической статистики для психологов [Текст] / Г. В. Суходольский. – Л.: ЛГУ, 1972. – 430 с.
19. *Дорофеев В. Н.* Использование кваліметрії для оценивания деятельности студентов в техническом вузе [Текст]: Учеб.-метод. пособ. / В. Н. Дорофеев, С. Н. Петрушков, Л. В. Шевцов, О. А. Сухинина. – Алчевск: ДГМИ, 2002. – 108 с.
20. *Рева О. М.* Однокрокові методи рішення задач з векторним показником ефективності [Текст]: Метод. вказівки з курсу «Основи теорії прийняття рішень» / О. М. Рева. – Кіровоград: ДЛІАУ, 1996. – 23 с.
21. *Губанов А. А.* Введение в системный анализ [Текст]: Учеб. пособ. / А. А. Губанов, В. В. Захаров, А. В. Коваленко; Под ред. Л. А. Петросян. – Л.: ЛГУ, 1988. – 288 с.
22. *Анфилов В. С.* Системный анализ в управлении [Текст]: Учеб. пособ. / В. С. Анфилов, А. А. Емельянов, А. А. Кукушкин. – М.: Финансы и статистика, 2002. – 368 с.