

УДК 004.853:378.046

ПЕДАГОГІЧНА ДІАГНОСТИКА З ВИКОРИСТАННЯМ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Білоусова Л.І., Колгатін О.Г., Колгатіна Л.С.

Харківський національний педагогічний університет імені Г.С.Сковороди

Проаналізовано технологію автоматизованої педагогічної діагностики. Запропоновано стратегію тестування та алгоритм оцінювання. На базі цієї методики розроблено програмне забезпечення "Експерт 3.02". Запропоновано методику адміністрування бази тестових завдань, яку апробовано у Харківському національному педагогічному університеті.

The technology of the automated pedagogical diagnostics is analysed. The testing strategy and marking algorithm is proposed. The software "Expert 3.02", based on suggested methods, is designed. The methods of administrating of the database of the test tasks are proposed. The approbation of this methods in the educational process of the Kharkiv National Pedagogical University is carried out.

Сьогодні тестування як інструмент педагогічної діагностики привертає до себе особливу увагу освітян, це зумовлено, перш за все, впровадженням зовнішнього тестування, що орієнтує викладацький загал на систематичне застосування тестових технологій у навчальному процесі. Крім того розгортання експериментального тестування випускників шкіл з використанням автоматизованих систем зумовлює зацікавленість у вивченні їх можливостей. Сучасні автоматизовані системи, які можна кваліфікувати як експертні системи, здатні підтримувати розвинені алгоритми тестування та аналіз тестових результатів. Тестування з використанням комп'ютера дозволяє вчителю легко отримати як індивідуальні, так і інтегральні характеристики знань та вмінь групи студентів у цілому і використати цю інформацію для корекції власної методики навчання. Широке поле наукових досліджень пов'язано з алгоритмами обробки та інтерпретації результатів тестування.

Метою даної роботи є узагальнення здобутих авторами результатів у розробці методики педагогічної діагностики, яка спирається на застосування автоматизованої системи і відповідає таким вимогам:

- у процесі тестування реалізуються різні форми розумової діяльності студентів;
- автоматизована система педагогічної діагностики зберігає діагностичну здатність в широкому діапазоні підготовленості студентів;
- обробка тестових результатів забезпечує максимальну інформативність для учня з точки зору визначення рівня навчальних досягнень і для вчителя з точки зору аналізу якості та корекції процесу навчання.

Першим етапом організації педагогічної діагностики є побудова ідеалізованої педагогічної моделі, тобто відокремлення основних елементів знань та вмінь, а також установа рівня оволодіння ними.

Другий етап являє собою створення системи завдань, що покриває всі елементи знань та вмінь на всіх рівнях навчальних досягнень.

Слід особливо підкреслити, що в сучасній педагогіці тест розглядається саме як система завдань зростаючої складності [1]. Повна та високоякісна педагогічна діагностика має будуватися на основі системи завдань усіх рівнів, починаючи з репродуктивного до творчого. За аналогією з рівнями навчальних досягнень [2], ми пропонуємо такі рівні тестових завдань [3, 4]:

1 – початковий рівень – дуже прості завдання, які відповідають репродуктивному рівню діяльності студента, головним чином, розпізнавання.

2 – середній рівень – завдання, які орієнтовані на репродуктивну діяльність. Ці завдання покривають усі основні факти та прості вміння згідно з навчальною програмою.

3 – достатній рівень – завдання на застосування знань та вмінь для розв’язування проблем у стандартній ситуації.

4 – високий рівень – завдання, які потребують створення і виконання нового алгоритму, перенесення знань у нову, нестандартну ситуацію, тощо.

Після завершення підготовки системи завдань в тестовій формі починається **третій етап** створення системи педагогічної діагностики – **апробація** завдань та визначення їх характеристик, що є визначальними з точки зору застосування завдань у тесті.

Наша технологія дослідження статистичних характеристик завдань базується на рекомендаціях Стандарту [5] і враховує особливості автоматизованої педагогічної діагностики [6].

Для проведення апробації звичайно підбирається група, яка є випадковою репрезентативною вибіркою із усієї сукупності студентів, для яких створюється тест. Студенти розв’язують завдання, які претендують на право увійти до тесту. Паралельно, незалежно від цього тестування встановлюється рівень навчальних досягнень студентів, що можна зробити на базі експертних оцінок, наприклад, традиційного екзамену. Апробацію слід організувати таким чином, щоб виключити суттєвої зміни кваліфікації студентів між розв’язуванням завдань тесту та незалежним експертним оцінюванням. Однак, неможливо організувати експертне оцінювання студентів так часто, як це потрібно для поновлення бази тестових завдань. Тому під час проведення поточної верифікації доцільно впорядковувати студентів за рівнем навчальних досягнень за допомогою тієї ж автоматизованої системи педагогічної діагностики, перевіряти кореляцію бала за окреме тестове завдання із загальною оцінкою за тестом. У такому випадку, дані апробації накопичуються неперервно, у тому числі при виконанні студентами самостійної роботи з використанням автоматизованої системи. Автоматизована система педагогічної діагностики дозволяє здійснити відбір результатів для поточної верифікації за датою та часом тестування, шифром студентської групи, варіантом тесту, часом обміркування відповіді [7, 4]. Валідність автоматизованої системи тестування у цілому, перевіряється через порівняння підсумкових тестових результатів студентів з результатами інших видів педагогічного контролю: опитування, екзамен, виконання практичних робіт тощо.

Після розподілу студентів за рівнями навчальних досягнень визначається відповідність рівня тестового завдання та його емпіричного індексу складності. Традиційно індекс складності обчислюється як частка правильних відповідей на завдання:

$$I_c = \frac{N_{np}}{N},$$

де N_{np} – кількість правильних відповідей, N – кількість випробуваних. Але для завдань з вибором правильних відповідей із запропонованих є ймовірність випадкового надання правильної відповіді навіть у випадку, коли студент не може розв’язати завдання. Тому при обчисленні міри складності (легкості) завдання доцільно враховувати тільки свідомо правильні відповіді. Для виключення правильних відповідей, які є наслідком випадкового вгадування, пропонується заповнити матрицю результатів тестування значеннями 1, 0 та

$\frac{-1}{k-1}$, де k – кількість варіантів відповіді, яка може бути різною для різних завдань [8]. Як

узагальнення такого підходу пропонуємо замість традиційного індексу складності обчислювати частку свідомо правильних відповідей за формулою [9]:

$$P_j = \frac{1}{N} \sum_{i=1}^N X_{i,j},$$

$$\text{де } X_{i,j} = \begin{cases} 1 & , \text{ правильна відповідь} \\ 0 & , \text{ відмова від відповіді} \\ \frac{-\alpha_j}{1-\alpha_j} & , \text{ неправильна відповідь} \end{cases}$$

де j – номер завдання; i – номер випробуваного; N – кількість випробуваних; α_j – ймовірність випадкового надання правильної відповіді для j -го завдання.

При будь-якій формі тестового завдання запропонована методика обчислення частки правильних відповідей забезпечує сутність цієї міри як імовірність того, що випадково обраний учень здатний виконати це завдання, а не вгадати відповідь. Надалі будемо використовувати саме частку свідомо правильних відповідей P_j як показник трудності (легкості) завдання.

За вимогами [2] учень із середнім рівнем навчальних досягнень (оцінка 4 за дванадцятибальною системою) “знає близько половини навчального матеріалу, здатний відтворити його відповідно до тексту підручника або пояснення вчителя, повторити за зразком певну операцію, дію”. Саме такий зміст мають завдання рівня 2 за обраною нами класифікацією. Таким чином, P_j для завдань рівня 2 не може бути меншим за 0,5 для таких студентів. Пропонуємо вважати придатним для завдань середнього рівня P_j у межах 0,5–0,9 на вибірці студентів із середнім рівнем навчальних досягнень. Слід зазначити, що такий діапазон частки правильних відповідей не є зручним з точки зору покращення статистичних параметрів тесту, але база завдань рівня 2 являє собою сукупність обов’язкових для вивчення фактів навчального матеріалу, тому автор завдань не може змінити їх трудність без зміни навчальної програми. Для завдань інших рівнів доцільно вибрати середнє значення $P_j = 0,5$ для студентів з оцінками відповідного рівня.

Для подальшого аналізу статистичних характеристик завдань необхідно визначитися щодо алгоритму оцінювання. Запропонований нами алгоритм (рис. 1) відповідає ідеології дванадцятибальної системи оцінювання [2]. Це спроба застосовувати абсолютну оцінку за ступенем опанування навчальним матеріалом, а не відносну оцінку, яка відображає ранжування студентів у групі. Такий підхід до оцінювання призводить до необхідності поновити погляд на методику підбору завдань до тесту за їх статистичними характеристиками.

Ідеальні характеристики завдань за часткою правильних відповідей, з точки зору обраного алгоритму оцінювання, наведено у таблиці 1. Безумовно, характеристики кожного реального завдання відрізняються від ідеальних, але у середньому за всіма завданнями тесту частка правильних відповідей має належати інтервалам, які наведено у таблиці 1. У випадку, коли оцінка виставляється за результатами того ж самого тесту, інакше і не може бути, оскільки зазначені частки свідомо правильних відповідей є межовими точками в алгоритмі оцінювання (рис. 1). Нове завдання, яке додається до тесту має відповідати цим умовам, інакше зміниться загальна трудність тесту.

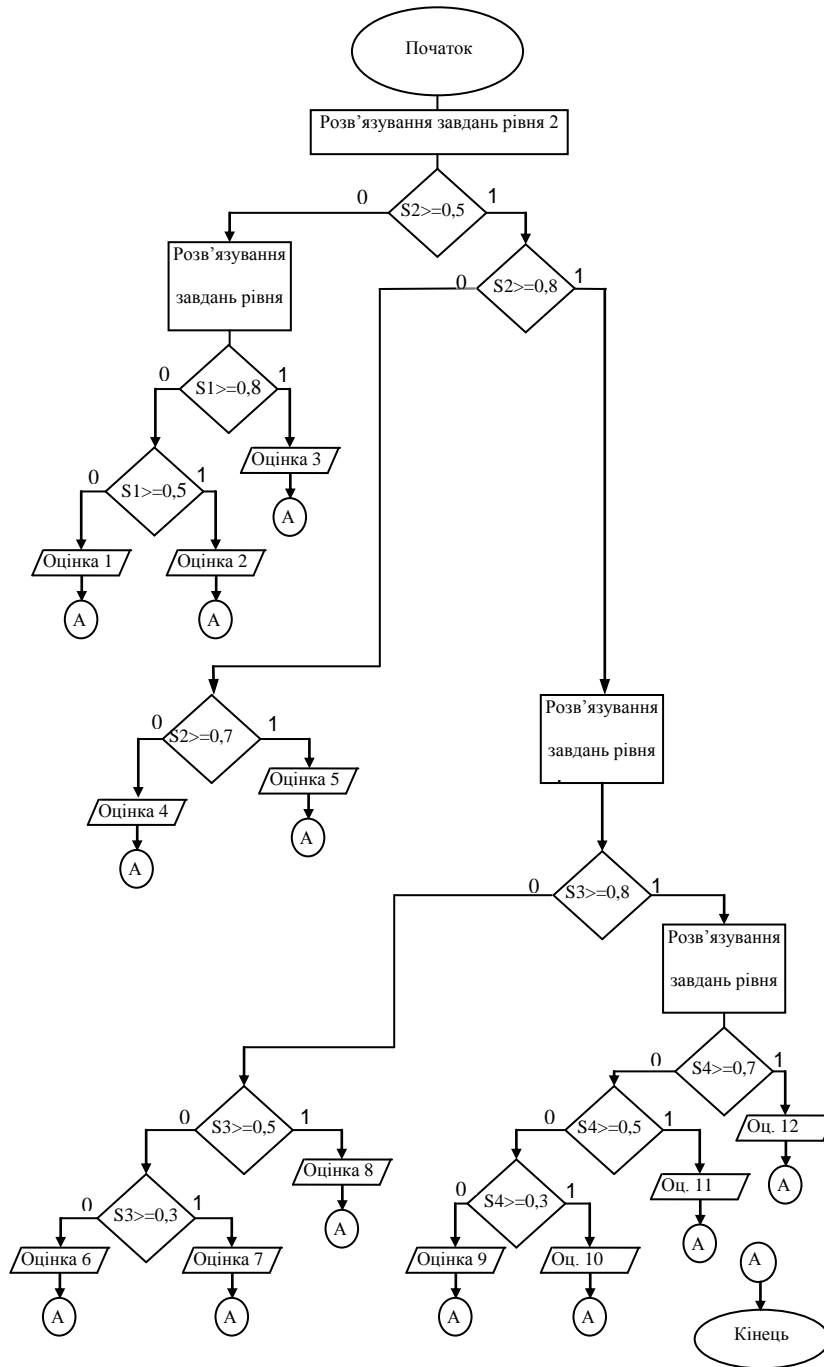


Рис. 1. Алгоритм оцінювання за результатами тестування.

Таблиця 1.

Ідеальна частка свідомо правильних відповідей для студентів з різною оцінкою.

Рівень завдання	Оцінка											
	1	2	3	4	5	6	7	8	9	10	11	12
Початковий	<0,5	0,5÷0,8	>0,8	1	1	1	1	1	1	1	1	1
Середній	0	0	<0,5	0,5÷0,7	0,7÷0,8	>0,8	1	1	1	1	1	1
Достатній	0	0	0	0	0	<0,3	0,3÷0,5	0,5÷0,8	>0,8	1	1	1
Високий	0	0	0	0	0	0	0	0	<0,3	0,3÷0,5	0,5÷0,7	>0,7

Таким чином, для завдань середнього рівня на вибірці студентів із середнім рівнем навчальних досягнень маємо такий алгоритм аналізу результатів апробації:

1. Перевірити здатність завдання розділяти студентів за їх навчальними досягненнями. Звернемо увагу на те, що індекс диференціюючої здатності [5] можна обчислити як різницю між індексами складності для груп студентів з різними оцінками [9]. Якщо завдання не забезпечує монотонного зростання частки свідомо правильних відповідей під час зростання оцінки студентів, або різниця між найбільшим та найменшим значеннями частки правильних відповідей не перевищує 0,25, то таке завдання не є тестовим, його слід переробити або відкинути (рис. 2).
2. Якщо завдання не має роздільної здатності для студентів з оцінками від 3 до 6, але добре розділяє студентів з оцінками від 1 до 3, то таке завдання слід перевести до початкового рівня (рис. 3).
3. Якщо завдання не має роздільної здатності для студентів з оцінками від 3 до 6, але добре розділяє студентів з більшими оцінками, то потрібно проаналізувати зміст завдання. Можливо при його виконанні студенти змушені використовувати більш складні види розумової діяльності, ніж передбачено автором, тоді це завдання можна перевести до достатнього або високого рівнів. Слід, також, проаналізувати процес навчання, виявити, чи не є зростання трудності завдання результатом нехтування деякими розділами програми.

Наприклад, під час викладання дисципліни “Основи інформатики та застосування ЕОМ в психології” студентам пропонується таблиця двійкових кодів для чисел від 1 до 15. У зв’язку з цим автори тесту віднесли завдання, що представлено на рис. 4, до репродуктивного рівня. Але, як показали результати тестування, під час практичного знаходження двійкових кодів студенти не користуються цією таблицею, а застосовують відомий алгоритм переведення чисел до двійкової системи числення. Таким чином виявилось, що для даного контингенту випробуваних завдань (рис. 4) відповідає достатньому рівню [7].

4. Завдання має добру роздільну здатність для студентів з оцінками від 3 до 6, але виявилось занадто трудним, частка правильних відповідей значно нижче ніж зазначено у таблиці 1. Можливо завдання виявилось комбінованим, та потребує відтворення одночасно декількох незалежних навчальних фактів. У такому випадку доцільно перевірити знання кожного факту за допомогою окремого завдання. Не виключено, що зростання трудності завдання пов’язано з огріхами у процесі викладання, тоді слід проаналізувати та вдосконалити методику викладання.
5. До тесту відбираються завдання середнього рівня (рис. 5, 6) таким чином, щоб середня частка правильних відповідей студентів з оцінками 4 та 6 на усі завдання тесту дорівнювала $0,5 \div 0,7$ та $0,8 \div 1$ відповідно.

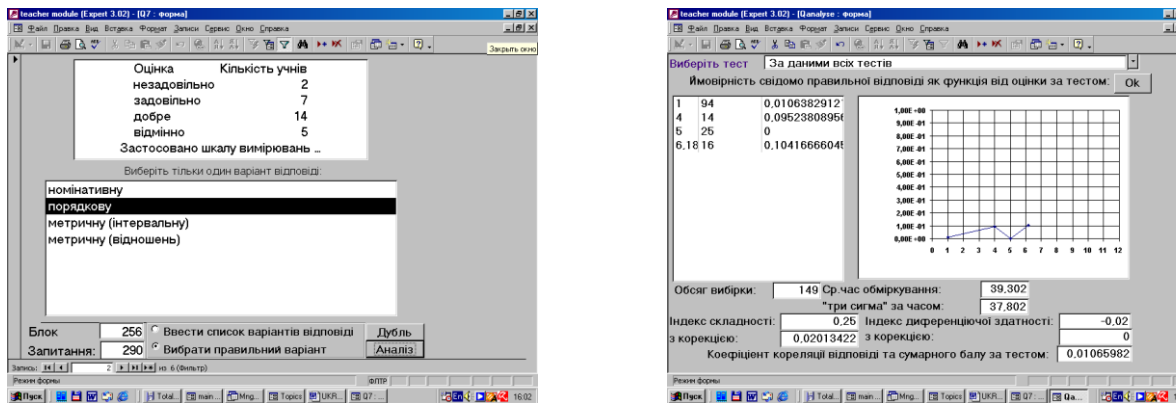


Рис. 2. Невдале завдання.

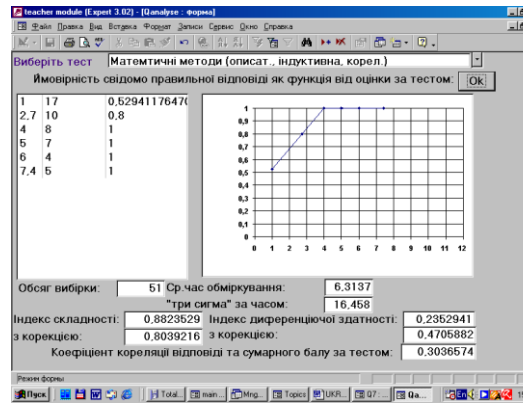
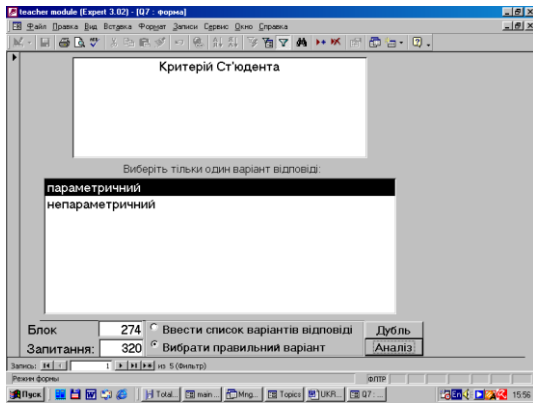


Рис. 3. Завдання відповідає рівню 1.

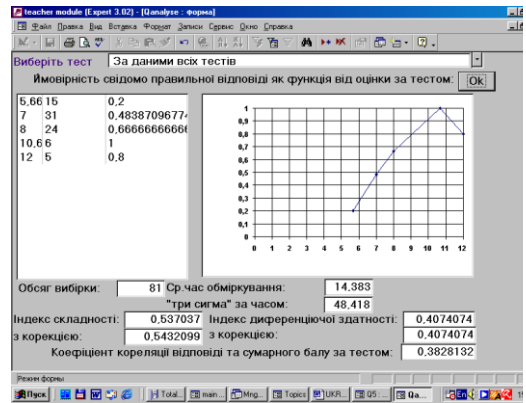
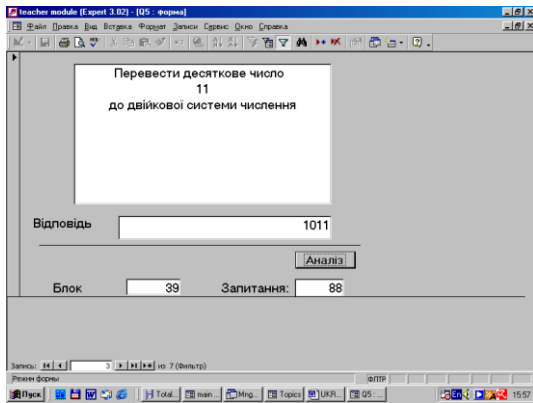


Рис. 4. Завдання відповідає рівню 3.

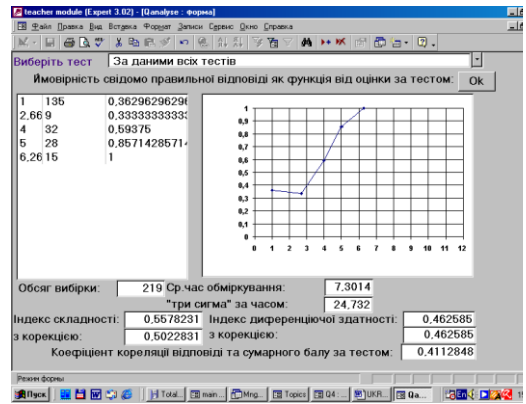
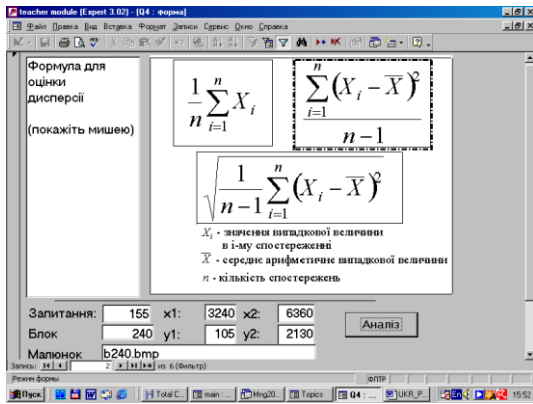


Рис. 5. Завдання відповідає рівню 2.

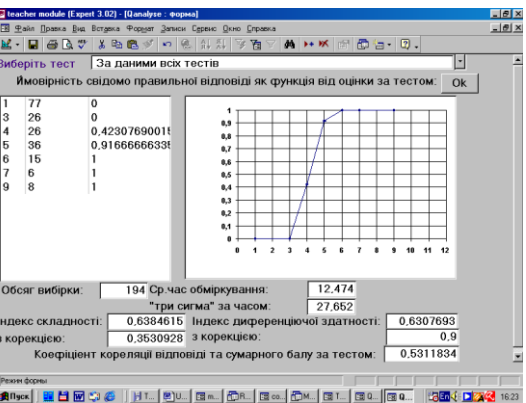
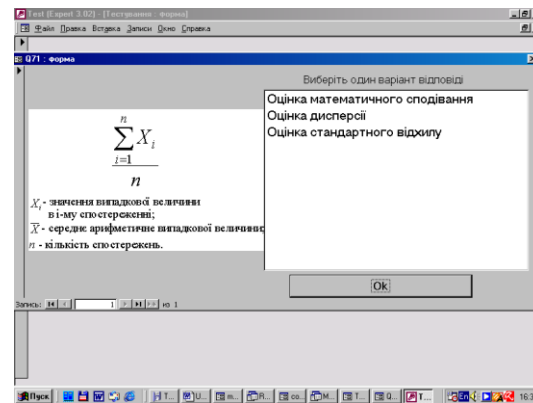


Рис. 6. Завдання відповідає рівню 2.

Для завдань рівнів 1, 3, 4 найкращим буде діапазон значень частки правильних відповідей 0,5–0,6, припустимим 0,3–0,7 на відповідній за рівнем вибірці студентів [5]. Аналіз цих завдань виконується аналогічно рівню 2.

Четвертий етап побудови системи педагогічної діагностики передбачає **побудову тесту** на базі тестових завдань з урахуванням їх статистичних характеристик. У сучасній практиці тестування відомо декілька стратегій. Найбільшого розповсюдження набула стратегія, яка передбачає подання завдань у порядку зростання труднощі, всім студентам пропонуються усі завдання тесту, але обмежується час на виконання роботи [1]. Можливо використання паралельних варіантів тесту, щоб знизити ймовірність попереднього ознайомлення студентів із завданнями. Така стратегія найкращим чином відповідає умовам бланкового тестування. Автоматизація процесу тестування з використанням комп'ютера дозволяє удосконалити стратегію, організувати процес так, щоб визначити рівень навчальних досягнень студента та запропонувати йому найбільшу частку завдань, які відповідають рівню його підготовки. У цьому напрямку розробляються різноманітні стратегії адаптивного тестування, серед яких поступове збільшення труднощі завдань з неперервним статистичним аналізом достовірності переваги правильних відповідей над неправильними [10] та навіть використання алгоритмів бінарного пошуку. Особливість нашої задачі полягає у тому, що потрібно визначити не тільки досягнутий студентом рівень, але також проаналізувати структуру його навчальних досягнень, зробити тестування якомога інформативнішим. Тому ми пропонуємо комбіновану стратегію тестування з векторною обробкою тестових результатів: окреме обчислення тестового бала для завдань кожного рівня та роздільна обробка результатів виконання тестових завдань відповідно елементам знань та вмінь.

Вибір рівня завдань для початку тестування є важливим питанням адаптивної стратегії. Тестування звичайно починають з найпростіших завдань. Такий підхід дозволяє знизити психологічний дискомфорт та створює атмосферу змагання, почуття росту відповідно із зростанням труднощі завдань. Приймаючи ці міркування до уваги, ми пропонуємо починати тестування саме із завдань рівня 2, які є найпростішими для випробуваного що зорієнтований на позитивну оцінку.

Є додатковий аргумент щодо вибору рівня 2 як стартового рівня для тестування. Тестові завдання рівня 2 відбивають обов'язкові факти з теми, що вивчається. Вони складають основу для аналізу структури знань та вмінь за елементами навчального матеріалу. Ці завдання не можуть бути виключені з процедури тестування. Було б недоцільним починати тест із завдань рівня 3, тому що продуктивні, особливо, творчі завдання ґрунтуються на досить широкому спектрі знань та не завжди дозволяють визначити, які саме елементи навчальної програми не засвоєні студентом. Щодо завдань рівня 1, то вони орієнтовані на студентів з незадовільною підготовкою, немає необхідності пропонувати їх усім випробуваним. Нашу стратегію тестування та алгоритм оцінювання подано на рисунку 1.

Наведемо деякі пояснення до рис. 1. Тестування починається із завдань рівня 2. Студент розв'язує обов'язковий мінімум завдань рівня 2, автоматизована система обчислює S_2 – його тестовий бал на рівні 2 та оцінює похибку тестового бала. Якщо точність є достатньою, автоматизована система виставляє загальну оцінку або змінює (підвищує або знижує) рівень завдань, які пропонуються студенту; інакше студенту пропонуються додаткові завдання рівня 2 до досягнення заданої точності. Слід підкреслити, що досягнута точність залежить не тільки від кількості завдань, але й від індивідуального тестового бала [9]. Потрібна точність обумовлюється різницею між емпіричним тестовим балом та значеннями, які є ключовими для прийняття рішення про зміну рівня завдань або вибір оцінки.

Тестування на рівнях 3, 4 та 1 проводиться аналогічно з окремим обчисленням тестових балів S_3 , S_4 та S_1 відповідно.

Такий алгоритм тестування підвищує відповідність між рівнем завдань, які пропонуються студенту та рівнем його підготовленості.

Комп'ютерну підтримку запропонованої технології забезпечує інформаційна система "Експерт 3.02", що побудована нами на основі розподіленої бази даних в середовищі Microsoft Access [7, 3]. Важливою перевагою системи "Експерт" є, на нашу думку, модульний принцип її побудови, що дозволяє автору завдань (можливо за допомогою програміста) створювати нові варіанти подання завдань. На сьогодні реалізовано 12 таких варіантів, наприклад, завдання з вибором однієї або декількох правильних відповідей можуть подаватися з графічними зображеннями (рис. 8) або без них, можливо вибирати відповідь мишею на рисунку (рис. 4), тощо. Основною базою даних інформаційної системи є база завдань у тестовій формі. Завдання розташовуються за темами для зручності перегляду. До кожної теми автор тесту визначає елементи навчального матеріалу, які підлягають перевірці, та до кожного з них створює коментар для студента, який не засвоїв відповідного матеріалу. Для перевірки знань та вмінь з кожного елементу навчального матеріалу створюється декілька різнорівневих блоків завдань. Блок складається з завдань, які є паралельними за формою, змістом та трудністю. Автор вказує до кожного блоку рівень завдань (1-4), ваговий коефіцієнт для врахування важливості даного питання при підсумковому оцінюванні в гетерогенному тесті, максимальний час експозиції завдання. Автор створює тест як систему завдань зростаючої трудності та вказує послідовність подання блоків. Під час здійснення тестування автоматизована система випадково вибирає та пропонує студенту одне або декілька завдань з кожного блоку. Кількість блоків у тесті та потрібна кількість завдань у кожному блоці визначається потрібною якістю діагностики. Система гарантує, що студент отримає не менш одного завдання з кожного блоку рівня 2, для забезпечення заданої точності автоматично вибирається декілька завдань. Завдання блоків рівнів 1, 3 та 4 пропонуються студенту за алгоритмом, який подано на рис. 1.

За результатами тестування формується база даних, яка містить інформацію про відповіді кожного студента на кожне завдання. Ця база включає такі основні поля: код завдання, правильність відповіді, рівень завдання, ймовірність випадкового надання правильної відповіді, час, що витрачено на розв'язування завдання. Зберігається також додаткова сервісна інформація, наприклад, час та дата тестування, оцінка тощо.

За результатами тестування студент отримує діагностичні дані за рівнями та елементами навчального матеріалу і діаграму, що відбиває структуру навчальних досягнень (рис. 7).

Автор отримує стандартний статистичний аналіз та геометричний образ кожного завдання, який відбиває трудність цього завдання для студентів з різною загальною оцінкою (рис. 2-6). Автор має можливість з використанням вбудованих засобів Microsoft Access передати матрицю результатів до табличного процесора для подальшого їх аналізу. Такий аналіз є важливим елементом верифікації тесту.

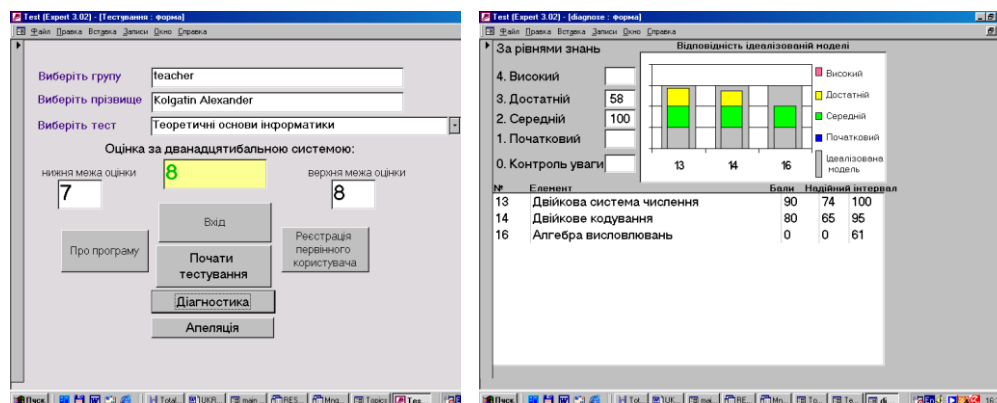


Рис. 7.

Після завершення підготовки та верифікації тесту можна вважати, що система педагогічної діагностики готова до експлуатації. **Етап практичного застосування** системи є

поєднанням процедури тестування і статистичної обробки отриманих результатів, яка включає інтерпретацію результатів, формування висновків стосовно кваліфікації випробуваного. У процесі експлуатації експертної системи педагогічної діагностики може виникнути потреба в удосконаленні бази тестових завдань, методики інтерпретації результатів або алгоритму формування висновків. Цілком істотно, це вимагатиме повернення до попередніх етапів роботи з системою.

Висновки:

1. Запропоновано розвинений алгоритм тестування та оцінювання, який враховує особливості застосування комп'ютерних технологій при тестуванні.
2. Запропоновано та впроваджено у навчальний процес Харківського національного педагогічного університету ім. Г. С. Сковороди методику адміністрування бази тестових завдань.
3. Розроблено автоматизовану систему педагогічної діагностики "Експерт 3.02".

ЛІТЕРАТУРА

1. Аванесов В.С. Композиция тестових завдань. – М.: Центр тестирования, 2002. – 240 с.
2. Критерії оцінювання навчальних досягнень студентів у системі загальної середньої освіти // Освіта України. – 2001. – № 6.
3. Білоусова Л.І., Колгатін О.Г. Педагогічне тестування та аналіз його результатів // Педагогіка та психологія: Збірник наук. праць. – Харків: ОВС, 2002. – Вип. 22. – С. 50-54.
4. Білоусова Л.І., Колгатін О.Г., Методика обробки та інтерпретації результатів педагогічної діагностики // Комп'ютер у школі та сім'ї. – №8, 2003. – С.28-31.
5. Засоби діагностики рівня освітньо-професійної підготовки. Тести об'єктивного контролю рівня освітньо-професійної підготовки. – Наказ МОН України № 285 від 31 липня 1998 р.
6. Білоусова Л.І., Колгатін О.Г., Колгатіна Л.С. Тестологічний аналіз у системі "Експерт" // Комп'ютер у школі та сім'ї. – №7, 2003. – С.41-43.
7. Колгатін О.Г., Колгатіна Л.С. Забезпечення валідності та надійності комп'ютерного тестування з інформатики // Актуальні проблеми та перспективи розвитку фінансово-кредитної системи України: Збірник наукових статей. – Харків: Фінарт, 2002. – С. 347-348.
8. Кромер В.В. О некоторых вопросах тестовых технологий // Тез. докл. Второй Всеросс. конфер. "Развитие системы тестирования в России", г. Москва 23-24 ноября 2000 г. – Ч. 4. – М.: Прометей, 2000. – С. 59-61.
9. Колгатін О.Г. Статистичний аналіз тесту з різними за формою завданнями // Засоби навчальної та науково-дослідної роботи. – ХДПУ ім Г.С.Сковороди. – Харків: ХДПУ, 2003. – Вип. 20. – С.50-54.
10. Шмелев А.Г., Бельцер А.И., Ларионов А.Г., Серебряков А.Г. Адаптивное тестирование знаний в системе "Телетестинг" // Тезисы докл. Всероссийской научно-практической конференции "Информационные технологии в образовании". – Москва, 2000. – <http://www.teletesting.ru/de/st109.html>.