

РЕАЛІЗАЦІЯ МЕТОДУ АДАПТИВНОГО ТЕСТУВАННЯ У СИСТЕМАХ ДИСТАНЦІЙНОГО НАВЧАННЯ

Федорук П.І. pavlo@pu.if.ua

Прикарпатський національний університет імені Василя Стефаника

У статті розглянуто можливості реалізації методів адаптивного тестування, подано найпростіші та необхідні процедури статистичної обробки результатів тестування знань і методи оцінки якості тесту. Розглянуто підхід до конструювання тестів, представлений в сучасній теорії тестування на основі математичної теорії параметричної оцінки тестових завдань на базі однопараметричної моделі Раша.

Ключові слова: адаптивний тест, статистична обробка, логіти рівнів складності.

РЕАЛИЗАЦИЯ МЕТОДА АДАПТИВНОГО ТЕСТИРОВАНИЯ В СИСТЕМАХ ДИСТАНЦИОННОГО ОБРАЗОВАНИЯ

Федорук П.И.

Прикарпатский национальный университет имени Василя Стефаника

В статье рассмотрены возможности статистических методов анализа результатов тестирования, приведены простейшие и необходимые процедуры обработки результатов тестирования знаний и методы оценки качества теста. Рассмотрено подход к конструированию тестов, который представлен в современной теории тестов на основе математической теории параметрической оценки тестовых заданий на базе однопараметрической модели Раша.

Ключевые слова: адаптивный тест, статистическая обработка, логит уровня сложности.

REALIZATION OF METHOD OF ADAPTIVE TESTING IN SYSTEM OF DISTANCE EDUCATION

P.I. Fedoruk

Precarpathian National University by Vasyl Stefanyk

The article reveals the possibilities of static methods of analysis of test results, gives the simplest and the most necessary procedures of statistical processing of knowledge test results and methods of evaluation of test quality. The article describes the approach to test construction which is described in the modern test theory on the basis of mathematical theory of parametric evaluation of test tasks on the basis of oneparametric model of Rasch.

Key words : adaptive test, statistical processing , logit of level of evaluation.

ВСТУП. Одним із найоб'єктивніших методів контролю та оцінювання знань, умінь і навичок, який позбавлений таких традиційних недоліків інших методів контролю знань, як неоднорідність вимог, суб'єктивність екзаменаторів та невизначеність системи оцінок, є тестування [1]. Деталі, пов'язані з комп'ютерним контролем знань, є об'єктом неабиякого інтересу для викладачів вузів і творців засобів ре-

лізації такого контролю. Крім того, питання комп'ютерного контролю недостатньо широко освітлені в теоретичному плані, і інтерес до них, як правило, реалізовується в більшості випадків шляхом створення чергової програми комп'ютерного контролю із завчасно складеним набором контрольних завдань. Проте, в області комп'ютерного контролю не все так очевидно, як здається на перший погляд. Існують деякі важливі питання, які або висвітлені в малодоступній літературі, або пропрацьовані недостатньо [2].

Тому вивчення, розробка та вдосконалення нових форм і методів комп'ютерного контролю знань є важливою і актуальною проблемою.

2. Традиційні тести

Перші наукові праці із теорії тестів з'явилися на початку XX століття, на стику психології, соціології, педагогіки й інших так званих поведінкових наук (Behavioral Sciences)[3]. Психологи називають цю науку психометрикою (Psychometrika), а педагоги - педагогічним виміром (Educational measurement). Одним з перших займався конструюванням та впровадженням тестового контролю в американській школі всесвітньо відомий американський учитель Е. Торндайк (1874-1949). Тестування як об'єктивний контроль рівня освітньо-професійної підготовки фахівця впроваджував французький психолог А. Біне, який розробив у 1900-1915 роках тести для перевірки інтелекту. Праці інших вчителів - О. Стоуна, Б. Зекінгема, Т. Келлі, Ч. Спірмена - присвячені створенню системи базових тестових завдань. У радянській школі були невдалі спроби працювати за тестовою технологією у 30-х та 70-х роках, але розповсюдження цей вид контролю не отримав. Із впровадженням модульно-рейтингової системи, тестові завдання є базою для перевірки досягнутого рівня розвитку здібностей, знань та умінь.

Традиційний тест являє собою стандартизований метод оцінки рівня знань і структури підготовленості людини. При проведенні такого тестування всі відповідають на одні і ті ж завдання, протягом однакового часу, в однакових умовах і з однаковими правилами оцінювання відповідей. Головна мета застосування традиційних тестів - визначити співвідношення порядку, що встановлюється за рівнем знань між тими, хто проходить тестування. І на цій основі визначити місце (чи рейтинг) кожного. Тут виникає одне з головних питань теорії тестів - питання підбору оптимального за деякими критеріями тесту. Кожен тест може відрізнитися від інших тестів кількістю завдань, їх складністю, а також багатьма іншими характеристиками. З прагматичної точки зору краще робити тест, що має порівняно менше число завдань, але який може відобразити весь спектр задач.

Один з актуальних напрямків сучасної організації тестового контролю - це *індивідуалізація контролю*, що приводить до значної економії часу тестування. Контроль ведеться за допомогою заздалегідь відкаліброваних за рівнем складності завдань. Якість тесту традиційно зводиться до визначення міри його *надійності* і *валідності* отриманих результатів. Якисним, як і об'єктивним, можна назвати тільки той

метод виміру, який обґрунтований науково і здатний дати необхідні результати. Теоретично виправданим критерієм упорядкування змісту є критерій складності завдань. Позанавчальний зміст (наприклад, перевірка рівня інтелектуального розвитку) у навчальний тест не включається. Це предмет психологічного виміру.

Складність завдань можна визначити двома способами:

1. На основі оцінки передбачуваного числа і характеру розумових операцій, необхідних для вдалого виконання завдань і оцінки підготовки студента;

2. На основі емпіричної перевірки завдань, з підрахунком частки неправильних відповідей на запитання тестів різної складності.

У класичній теорії тестів багато років розглядалися тільки емпіричні показники складності. У сучасних теоріях навчальних тестів, які використовуються в дистанційному навчанні, більше уваги почали приділяти характеру розумової діяльності у процесі виконання тестових завдань різних форм і психологічного типу студента. Показник складності завдання розглядається як важливий системо- і, одночасно, структуроутворюючий фактор тесту.

3. Адаптивні тести

Доцільність адаптивного контролю впливає з необхідності раціоналізації традиційного тестування. Кожен викладач розуміє, що добре підготовленому студенту немає необхідності давати легкі завдання, оскільки такі матеріали не володіють помітним потенціалом розвитку. Аналогічно, через високу імовірність неправильного рішення немає рації давати важкі завдання слабкому учню. Відомо, що важкі і дуже важкі завдання знижують навчальну мотивацію багатьох учнів. Потрібно знайти порівняння в одній шкалі для міри складності завдань і міри рівня знань. Ця міра була знайдена в теорії педагогічних вимірів. Данський математик Г. Раш назвав цю міру словом «логіт» [4]. Після появи комп'ютерів ця міра лягла в основу методики адаптивного контролю знань, де використовуються способи регулювання складності і числа пропонованих завдань, залежно від відповіді учнів. При успішній відповіді комп'ютер підбирає більш важке наступне завдання, при неуспішній - легше. Зрозуміло, цей алгоритм вимагає попереднього випробування всіх завдань, визначення їх міри складності, а також створення банку завдань і спеціальної програми. Використання завдань, що відповідають рівню підготовленості, істотно підвищує точність вимірів і мінімізує час індивідуального тестування до 5-10 хвилин. Адаптивне тестування

дозволяє забезпечити комп'ютерну видачу завдань на оптимальному, приблизно 50%-му рівні імовірності правильної відповіді для кожного учня.

Модель Раша. Успіх учасника тестування при розв'язанні деякого тестового завдання залежить від двох факторів: складності завдання і рівня підготовки учасника. Ймовірність того, що деякий учасник вірно виконає конкретне завдання, є функцією щонайменше двох аргументів - рівня підготовки учасника тестування S та рівня складності даного завдання t :

$$P = P(S, t).$$

Таку функцію називають функцією успіху. Якщо вигляд функції успіху відомий, то за результатами випробувань можна методами математичної статистики з певною точністю оцінити аргументи цієї функції, в -ому числі і рівень складності завдань [5].

основна логістична модель Раша полягає в тому, що ймовірність правильної відповіді першим учасником (більш підготовленим) на перше завдання має співпадати із ймовірністю правильного виконання другим учасником (менш підготовленим) другого завдання (менш складного).

Із цього випливає, що

аргументи S і t тісно пов'язані між собою, неможливо визначити один із них, не визначивши другий;

ймовірність успіху залежить не від кожного аргументу S і t окремо, а від їх відношення:

$$P = P(S, t) = P_1(\xi), \quad \xi = \frac{S}{t}$$

Параметри S і t називають латентними (не спостережуваними) параметрами [6], оскільки вони описують деякі приховані характеристики учасників тестування та тестових завдань.

Функція успіху запишеться у вигляді однорідної функції нульового порядку:

$$P = P(S, t) = \frac{S}{S + t} = \frac{S/t}{S/t + 1} = \frac{\xi}{\xi + 1}$$

Ця найпростіша модель ймовірності успіху була запропонована в 1957 році математиком із Данії, Джорджем Рашем [7]. Вона вперше дала можливість об'єктивно визначати співвідношення між учасниками тестування і тестовими завданнями довільних рівнів підготовки та складності.

4. Реалізація адаптивного тестування на основі моделі Раша

За допомогою моделі Раша можна на основі тестувань визначити рівні складностей тестових завдань. Як наслідок, питання, які були задіяні в тесті можна розмістити на одній шкалі. Але для шкалування всієї

сукупності питань із бази тестів не можна просто застосувати модель Раша. Адже для одного тесту здійснюється лише вибірка окремої кількості питань. Необхідною умовою для застосування моделі Раша є те, що всі вибрані завдання повинні бути запропоновані кожному із студентів. В результаті ми отримуємо щільну матрицю результатів. Для шкалування наступної (відмінної від попередньої) вибірки питань на ту ж саму шкалу отримати логіти рівнів складності недостатньо, тому що при наступному тестуванні рівні знань студентів можуть відрізнятись і навіть склад групи дещо може різнитись від складу протестованих у попередньому тесті. Тому ми запропонували метод калібрування на основі трьох опорних завдань із попереднього тесту. Таким чином, у результаті скінченної кількості поточних тестувань можна звести на одну шкалу всі питання з бази. Кількість таких експериментальних тестувань залежить від вибірок.

Процес застосування даного методу зображено на схемі (рис. 1).

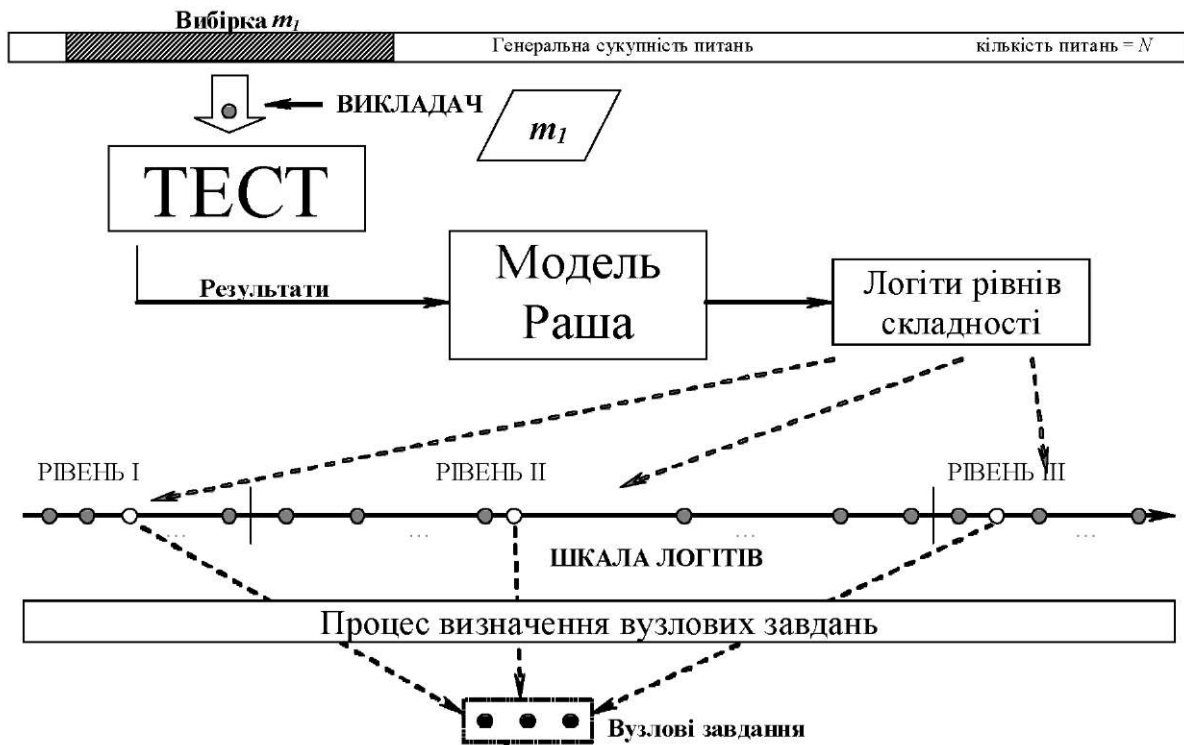
Таким чином було сформульовано постановку задачі для практичної реалізації методу визначення міри складності тестового завдання на основі однопараметричної моделі Раша. Для реалізації цього методу нами було створено програмний модуль до існуючої системи дистанційного навчання за допомогою засобів РНР, який виконує контроль над подачею тестових питань, опрацьовує результати тестувань за допомогою методів, описаних вище, ставить у відповідність ідентифікатору питання значення логіта рівня складності. Це дозволяє розмістити питання на одній метричній шкалі.

Алгоритм функціонування такого модуля можна подати в вигляді блок-схеми, яка зображена на рисунку 2.

На перших стадіях функціонування дистанційного курсу формується перший тест для поточного тестування. Для тестового контролю в даному курсі викладач формує категорії завдань, із яких будуть вибиратися завдання для тестувань. Для першого тесту викладач задає кількість випадково відібраних завдань з цих категорій (рисунок 3). На рисунку видно, що для тесту відібрано два завдання із категорії 1 і набір з десяти випадкових завдань з цієї ж категорії.

Після цього формується нова категорія „звжених” завдань, в котру ці завдання автоматично переносяться з наявних вже категорій. Відбувається організоване тестування. Після закінчення ми отримуємо щільну первинну матрицю результатів. Вона опрацьовується за допомогою моделі Раша. В результаті ми отримуємо масив логітів рівнів складності цих завдань.

Етап I



Етап II

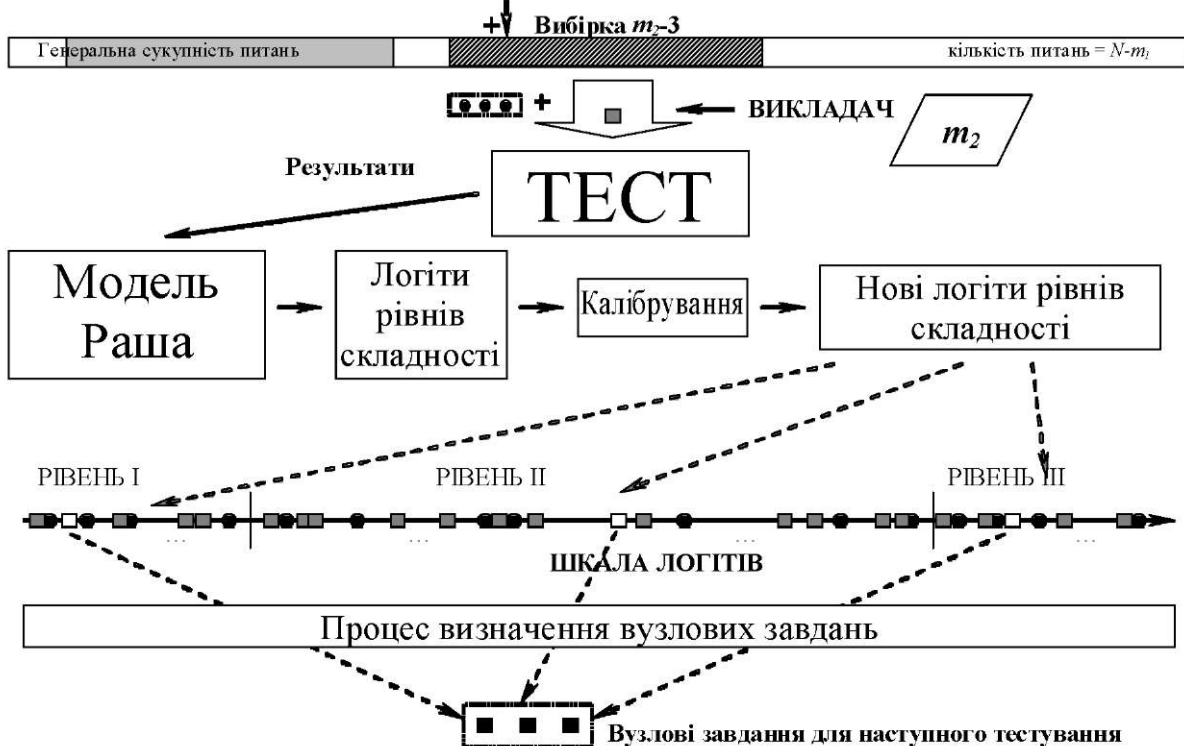


Рис. 1. Процес шкалування тестових завдань генеральної вибірки (I-II етап)

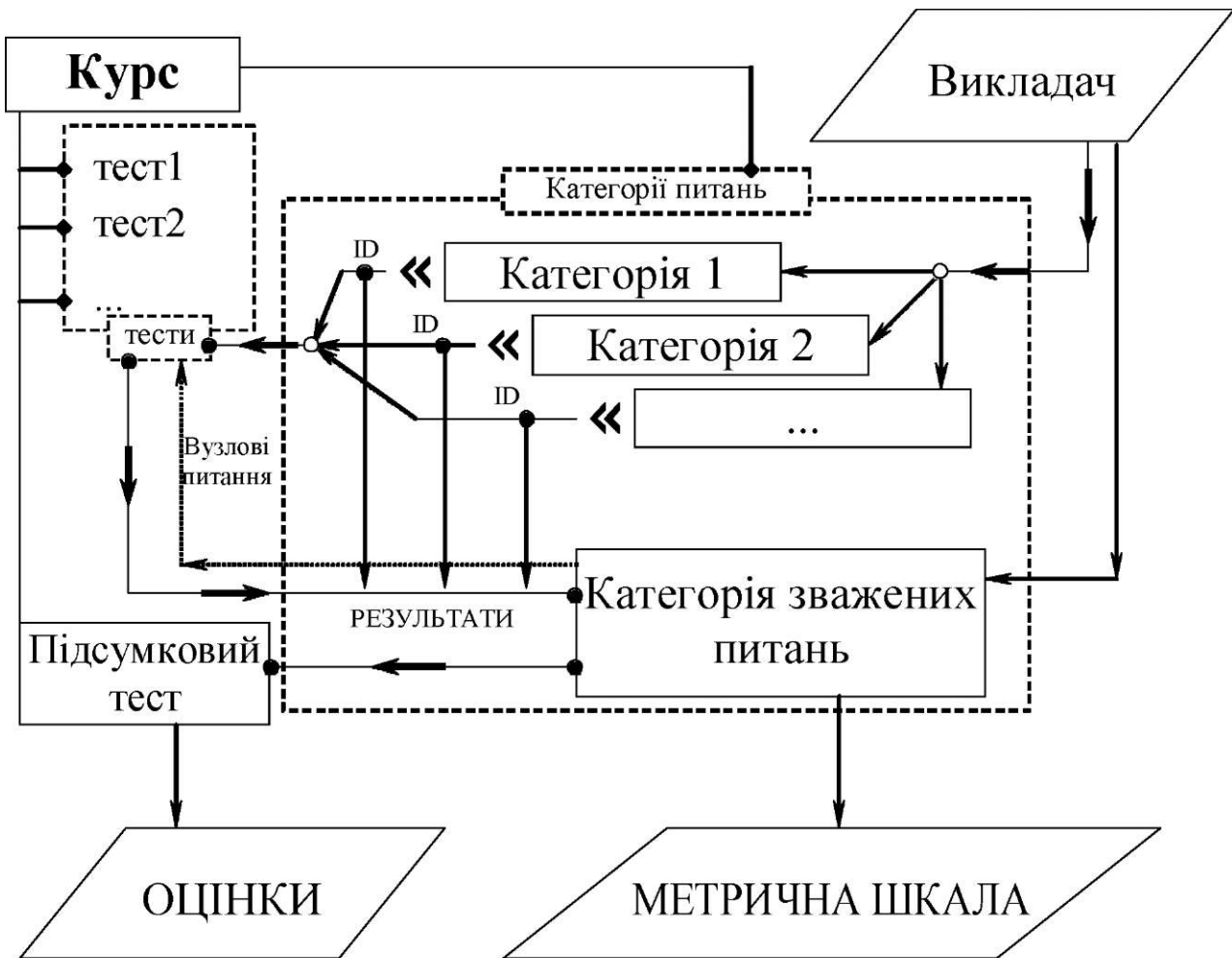


Рис.2. Блок-схема програми

При створенні наступних тестів викладач вибирає завдання із групи „незважених”, що залишилися в наявних категоріях. Таким чином, протягом поточних тестувань вичерпується вся генеральна сукупність тестових завдань, передбачених для даного курсу. Для того, щоб звести на одну шкалу логіти рівнів складності завдань, задіяних в другому тестуванні, перед тестуванням відбувається процес визначення вузлових завдань. В даному експерименті це значення логітів, що набувають від’ємного, близького до нуля, додатного значень відповідно. Фактично, тепер ці завдання відповідають трьом рівням складності. За допомогою результатів відповідей на ці ж питання в другому тестуванні ми можемо визначити поправку для визначення логітів, які розмістяться на потрібній нам метричній шкалі.

Процес вибору і переміщення завдань між категоріями аналогічний до попереднього. В експерименті нами було передбачено, що протягом поточних тестувань будуть зважені всі завдання з курсу і при

підсумковому тестуванні можуть бути враховані ваги питань. Таким чином, ми вирішуємо проблему неадекватності визначення рівня знань студента. Маючи логіти, які відповідають завданням на метричній шкалі, такий метод надає можливість оцінити знання студента, враховуючи складність завдань, на які він відповідав.

ВИСНОВОК. Створення описаного вище програмного модуля є вагомим кроком у розвитку систем адаптивного комп'ютерного контролю знань. Результати роботи програм, що базуються на математичній моделі Раша, дозволяють адекватно визначити міри складностей тестових завдань, незалежно від рівнів знань студентів, що брали участь в тестуваннях. Це суттєво впливає на ефективність визначення адекватної оцінки знань. Встановлення рівнів складностей тестових завдань і можливість звести логіти на одну метричну шкалу дозволяє розбити базу тестових завдань на рівні складності. Це дуже важливий крок для адаптації тестового контролю знань в систе-

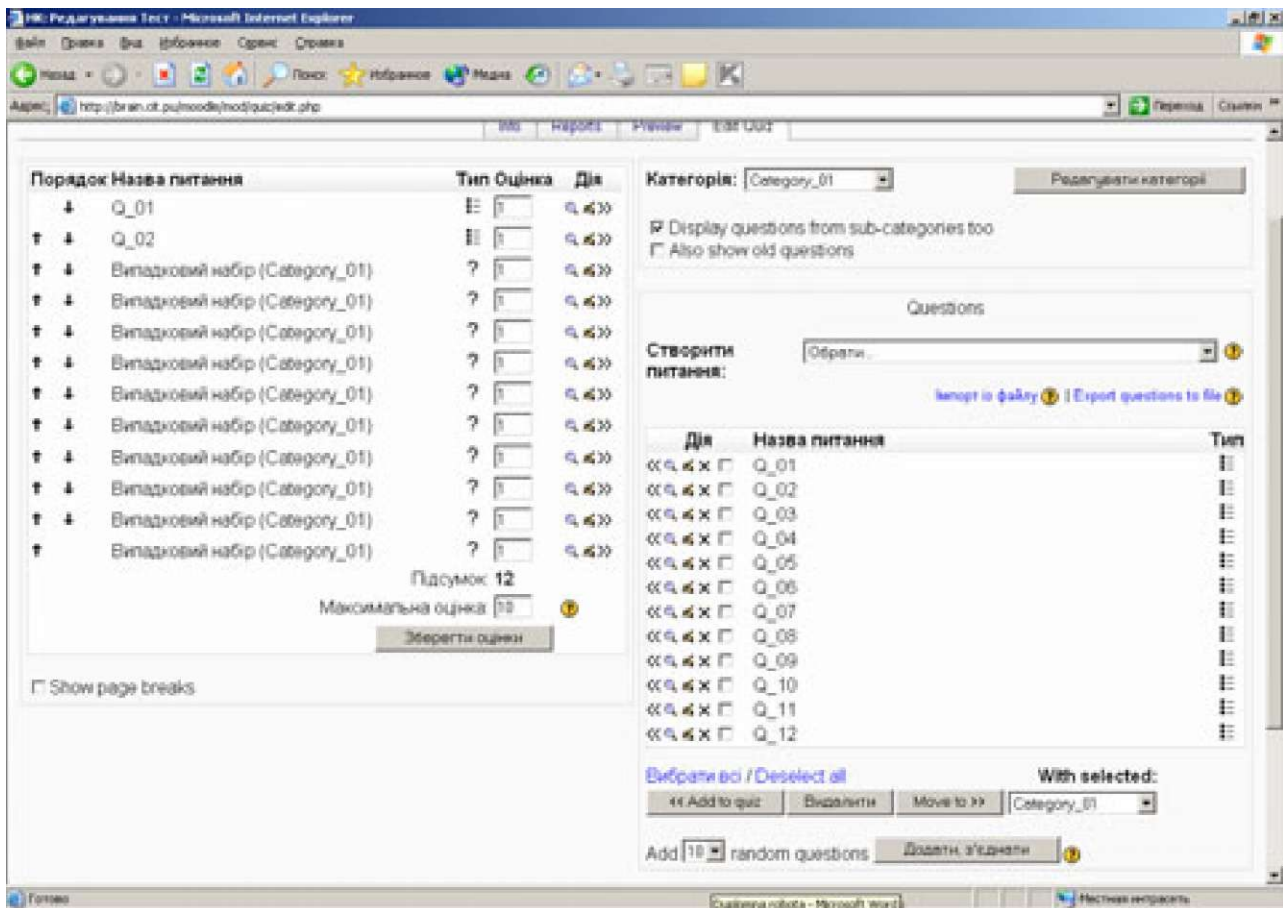


Рис.3 Інтерфейс викладача

мах дистанційного навчання. Адже адаптивний тест включає регулювання подачі різних за складністю завдань під час процесу тестування. Особливістю методу, що використовується в даному програмному модулі є те, що зведення логітів рівнів складності тестових завдань відбувається на основі результатів поточних тестувань. Тобто протягом проходження курсу студент перевіряє свої знання, а ми, в свою чергу, використовуємо результати таких перевірок для визначення мір складності тестових завдань.

Важливим елементом вдосконалення вищезгаданих процесів є можливість уточнення логітів рівнів складності тестових завдань. Тобто, розміщення числових значень на метричній шкалі буде змінюватись і покращуватись протягом наступних поточних тестувань і навіть в інших групах студентів. Цим і визначається одна з особливостей даного програмного модуля, який базується на обробці первинних результатів за допомогою математичної моделі Раша.

Література

1. Аванесов В.С. Теория и методика педагогических измерений (материалы публикаций). - М.: Подготовлено ЦТ и МКО УГТУ-УПИ, 2005. - 98 с.
2. Кривицкий Б.Х. К вопросу о компьютерных программах учебного контроля знаний // Международный журнал "Образовательные технологии и общество". - Т. 7, № 2. - 2004. - С. 158-169.
3. Аванесов В.С. Научные проблемы тестового контроля знаний—М., 1994. — 135 с.
4. Trevor G. Bond, Christine M. Fox. Applying the Rasch Model. Fundamental Measurement in the Human Sciences. 2001. -280p.
5. Чельшкова М.Б. Разработка педагогических тестов на основе современных математических моделей: Уч. пособие. - М: Исследовательский центр проблем качества подготовки специалистов, 1995. - 32 с.
6. Lord F. M. Application of Item Response Theory to Practical Testing Problems. Hillsdale N-J. Lawrence Erlbaum Ass., Publ., 1980. - 266 p.
7. Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. With a Foreword and Afteword by B.D. Wright. The Univ. of Chicago Press. Chicago & London, 1980. - 199 p.