

**Персоналізація контролю знань студентів у імітаційній моделі тестування**

Однією з основних форм процесу навчання є контроль знань, і як одна з його складових, що широко використовується нині, форма контролю знань – тестування.

У традиційній системі тестування не завжди існує можливість студентові продемонструвати свої істинні знання – тобто розуміння тієї або іншої теми чи дисципліни в цілому, через те, що процес тестування викликає труднощі і протиріччя, пов'язані з однозначністю оцінювання знань випробовуваних. У багатьох системах тестового контролю принцип опрацювання результатів зводиться до одинарної вибірки (дихотомічної системи оцінювання тестових завдань) [4, 55], у інших випадках – це політомічна система оцінювання, що допускає кілька категорій відповіді на завдання, кожна з яких оцінюється по-різному [4, 56]. Очевидним недоліком цих систем є рівнозначність питань у базі даних. Ці недоліки були компенсовані в системі контролю з розподілом завдань на різні рівні [4, 56], відсутність яких може привести до недостатньої об'єктивності оцінювання знань. Так, якщо сильному студентові дістаються тільки складні завдання, а слабкому – тільки легкі, то в результаті оцінювання у обох студентів буде констатований однаковий рівень знань, що не відповідає дійсності. І навпаки, якщо студентам з однаковим рівнем знань дістануться нерівнозначні питання, то перевірка знань виявить у них різний рівень підготовки, що так само не буде вірогідним.

Методика розподілу тестових завдань на різні рівні, не дивлячись на уявну ефективність, має свої недоліки: при визначенні «ваги» питання (відносної значущості) з боку викладача накладається певний суб'єктивізм у процес оцінювання знань студентів, оскільки не кожне легке завдання для викладача є так само легким і для студента. Крім того з точки зору педагогічної теорії не буває питань простих і складних. Простим питання завжди буде для того, хто знає на нього відповідь, і складним – для того, хто відповідь не знає. Таким чином, визначаючи «рівні» питань, викладач фактично розставляє їх відповідно до власного уявлення про їх складність, тобто виходячи із власного рівня компетентності.

Проблеми тестового контролю, якості тестових завдань, у тому числі критерії значущості, складності і труднощі тестових завдань розглядають у своїх роботах А.Н. Майоров, В.В. Семенець, Г.О. Мірських, І.А. Морєв, І.А. Белоус, І.В. Куцевич, М.Б. Челишкова, М.Ф. Бондаренко, Н.В. Белоус, Н.Ф. Єфремова, Ю.Ф. Зінковський [3-9].

Однак, питання, пов'язані з персоналізацією тестового контролю, не знайшли в них вичерпного рішення.

Більшість авторів, як і розробники тестових завдань, розглядають нерівнозначність як складність або важкість тестових завдань [7], і пропонують визначати їх рівні встановлення трьома способами, хоча ці способи не позбавлені певних недоліків [1]:

1) На основі оцінки передбачуваного числа і характеру розумових операцій, необхідних для вдалого виконання завдань і оцінки підготовки студента. Цей спосіб у більшості випадків має теоретичне значення, оскільки його практичне застосування обмежене порівняно вузькою галуззю знань, контролюючи яку, можна з високою вірогідністю встановити кількість виконуваних розумових операцій і зв'язати їх специфіку з мірою складності контрольованого навчального матеріалу.

2) На основі емпіричної перевірки завдань, із підрахунком частини неправильних відповідей на питання тестів різної складності. Недоліком цього способу є неможливість створення однакових умов тестування для усіх студентів, а також неможливість застосування для контрольованого матеріалу, який вивчається обмеженою кількістю студентів.

3) Експертним оцінюванням, що відбувається шляхом попарного порівняння прототипів тестових завдань [2]. Проблемним при реалізації методу є призначення експертів, здатних коректно оцінити складність завдань.

Спираючись на принципи А.І. Морєва [8, 125-126], можна сформулювати ключові критерії, яким повинна відповідати сукупність завдань, що входять у тест, серед них: складність, значущість, трудомісткість та інші.

М.Б. Челишкова [9], розглядаючи статистичні характеристики тестових завдань у класичній і сучасній теоріях тестів, виділяє найважливішу – це важкість тестових завдань, і визначає її по-різному, залежно від вибраної теорії методів опрацювання емпіричних результатів виконання тесту.

Із зазначеного слідує, що хоча наступні критерії якості тестових завдань: значущість, складність і важкість [8] націлені на один і той самий процес – тестовий контроль знань, і розкривають різні його сторони, однак вони не забезпечують персоналізацію контролю знань студентів.

На сьогодні основним видом комп'ютеризованого контролю сформованості знань студентів є тестування, можливості якого обмежені введенням завдань, варіантами відповіді (за вибором), вказуванням часу і шкалою оцінювання. Рівень сформованості знань визначається порівнянням процентної частки правильних відповідей із вибраною шкалою оцінювання. Тести проектуються з використанням загальної бази тестових завдань і при відборі завдань немає можливості повною мірою врахувати специфіку окремих спеціальностей. Завдання, що включаються в тести, недостатньо диференційовані, не мають вагових коефіцієнтів складності і значущості, тому можуть не нести певної значущості конкретно для цієї спеціальності, і студенту можуть дістатися ті самі питання, що й на попередньому тестуванні, складність завдання не вказана і не регламентована.

Для диференціації контролю знань студентів напряму підготовки «Початкове навчання»: «Інформатика», «Англійська мова» і «Дошкільне навчання» проводилося дослідне тестування з використанням програмного комплексу SSUQuestionnaire (<http://www.test.sumdu.edu.ua>), теоретичною основою якого є імітаційна модель тестування [2], що дає можливість отримати результати контролю, зіставлені за критеріями складності і значущості.

На підготовчому етапі, виходячи з цілей навчання, конкретизувалися програмні вимоги спеціальностей. Для включення в тест добиралися завдання, які відповідали задачам навчання дисципліни в цілому і змісту окремих модулів і тем. При цьому розробник, обираючи прототип тестового завдання, мав встановити міру необхідності, актуальності включення в тест завдань для перевірки ключових знань з конкретної дисципліни, повноту відображення в тестовому матеріалі її змісту. А це ототожнюється з таким поняттям, як значущість. Тобто при попередньому доборі прототипів тестових завдань викладач може (навіть апріорі) керуватися тільки значущістю цього матеріалу в темі, розділі чи дисципліні в цілому, його належністю до основних або додаткових тем дисципліни.

Значущість навчального матеріалу є одним з основних критеріїв визначення результативності оцінювання якості навчання і однією з найважливіших характеристик, обов'язкових для врахування у процесі тестування. Вірогідність встановленого рівня значущості у значній мірі залежить від правильності планування змісту; від суб'єктивності мислення розробника (обґрунтованого співвідношення матеріалу дисципліни з кількістю годин, виділених для неї, правильного співвідношення кількості балів з дисципліни з конкретним питанням теми і так далі); від необхідності створення банків тестових завдань із деяких дисциплін, які можуть включати одні і ті самі питання, але при цьому знаходити різне віддзеркалення в окремих групах студентів, тобто від правильного встановлення рівнів нерівнозначності прототипів тестових завдань.

Наступний етап – створення прототипів тестових завдань, основою якого є визначення складності тестових завдань і значущості тем контрольованої дисципліни.

Правильно сконструйований тест не може складатися тільки з легких або тільки з важких завдань. Він включає завдання різної складності. Але при цьому має бути встановлена *відносна значущість* відображеного в завданні елемента теми, модуля, дисципліни, яка пропорційна *відносній кількості праці*, витраченій студентом на виконання цього тестового завдання, і яка, у свою чергу, визначається через усереднене відношення часу його виконання.

Визначення *значущості* на етапі експертного аналізу дисципліни, думка групи експертів повинна вказувати на значущість змісту теми, що рецензується, відносно інших тем дисципліни. Відповідно до цього аналізуються не окремі тестові завдання кожної теми, а завдання усіх тем дисципліни в цілому, оскільки для досягнення цілей навчання важливо виділити найбільш значні.

Наприклад, якщо тестуються дві спеціальності «Початкове навчання. Інформатика» і «Початкове навчання. Англійська мова», то тема «Перекладачі Lingvo, Pragma» буде значиміша для підгрупи «англійців», ніж для «інформатиків». Це означає, що при плануванні цієї теми викладач групі «англійців» повинен виділити більше годин для її опрацювання, а для «інформатиків» – може винести зазначену тему з основного складу в додатковий матеріал і запропонувати її для розгляду студентам самостійно. Відповідно завдання цієї теми для «англійців» входять у діапазон значимих (будуть мати індекс значущості), а для «інформатиків» – ні.

При цьому є теми, які загально значущі для всього потоку студентів, – це, наприклад, «Робота з проектором і електронною дошкою», дисципліни «Обчислювальна техніка і технічні засоби навчання» на пряму підготовки «Початкове навчання», знання з яких знадобляться як упродовж навчання у Вузі, так і в майбутній професійній діяльності.

Етап встановлення *рівнів значущості і складності* виконується шляхом узагальнення думки групи експертів і прийняттям її як кількісної міри значущості і складності завдань – *індексів складності і значущості* [1].

Очевидно, що при встановленні повноти знань під час добору тестових завдань слід врахувати їх кількість і складність. Неправильна кількість тестових завдань, недостатня або

надмірна їх складність може привести до випадкових результатів тестування і спотворення цілісної картини (повноти) навчального матеріалу. Тому доцільно дотримуватися пропорції як між змістом матеріалу і кількістю тестових завдань для його перевірки, так і між завданнями різної складності. Слід також враховувати, що кожне завдання буде значиме для усієї групи студентів однаково, але по-своєму складне для розробника тесту і кожного студента окремо.

За результатами створення прототипів тестових завдань формується загальна база, яка після експертної оцінки набирає персоналізованого вигляду для підготовки студентів кожної спеціалізації.

З однієї і тієї самої бази тестових завдань можна сформувати різні тести для дисциплін одного напрямку підготовки, але різної спеціалізації. При цьому завдання повинні добиратися з урахуванням цілей навчання конкретної дисципліни – кожен набір завдань виражатиме значущість контрольованої дисципліни. Для об'єктивності і персоналізації процесу перевірки знань треба досягти зіставлення цих наборів за значимістю і складністю, тобто, щоб числові значення значущості окремих тестів можна було б співвіднести (урівноважити) один з одним.

Критерієм персоналізації виступають індекси складно-значущості і кількість тестових завдань (у кожному випадку свої), які можна встановити завдяки використанню імітаційної моделі тестування. Увесь набір завдань, пропонованих студентам, може містити завдання як загальні для усього потоку, так й індивідуальні за значимістю для окремої спеціальності. Отже завдання в загальній базі можуть бути поділені на загальнозначущі і спеціалізовано-значущі.

У традиційному тестуванні, якщо набрана кількість балів лише трохи (умовно) відрізняється від норми (порогового числа), тобто знаходиться відносно близько від порогового числа, викладач або ж «дарує» бракуючий бал студентові, або ж «жорстко» оцінює і знижує його. Інколи під час підсумкового контролю за браком кількох балів (наприклад, з 58 до 60), викладач пропонує студентові пройти тест ще раз, не змінюючи завдання, або ж готує заздалегідь новий і ще не апробований.

Щоб уникнути суб'єктивізму з боку викладача і персоналізувати перевірку знань, пропонується, відповідно до імітаційної моделі, проводити тестування у кілька етапів, у ході чого буде виявлений фактичний рівень знань кожного студента. При цьому наявність другого етапу може бути застосована не для усієї групи, а лише для деяких студентів, у яких є сумнів в оцінці рівня їх знань за результатами 1 етапу.

Отже, після призначення індексів складно-значущості готовий тест із заданими рівнями *складності і значущості* пропонується студентам для перевірки персональних знань, яку доцільно проводити у два етапи.

Спершу (1 етап тестування) – проводиться основна сесія контролю, задіявши критерії складно-значущості завдань. Після проведення основної сесії з'являються дані про те, якою мірою кожен студент володіє матеріалом. Продовження тестування пов'язане тільки з тим, що виявлені під час основної сесії контролю навчальні досягнення неможливо зіставити зі шкалою оцінок, і тоді в імітаційній моделі формується нова база тестових завдань шляхом відбору з найбільш проблемних для студентів тем.

Настає додаткова сесія (2 етап), під час якої здійснюється уточнення дійсних учбових досягнень студентів, стосовно яких виникли сумніви. У імітаційній моделі тестування реалізований підхід традиційного (усного) іспиту, коли екзаменатор сумнівається у знаннях питань з білета і ставить питання із значущих (основних) тем курсу, що вивчається. При цьому якщо студент робить у відповідях помилки на додаткові питання, то у викладача спектр питань розширюється і їх стає тим більше, чим гірше студент відповідає на попередні.

Отже, для персоналізації результатів тестування в додатковій сесії кожному учаснику пропонуються завдання саме з тих тем, які особисто він засвоїв не повною мірою, спираючись при цьому на результати оцінювання, отримані в основній сесії. При цьому завдання мають бути з розряду значущих, завдяки чому перевіряється рівень володіння матеріалом в цілому – розуміння матеріалу з важливих розділів цієї дисципліни. Додаткова сесія персоналізуватиметься таким чином, що ймовірність включення в тест завдання з будь-якої теми становиться тим вищою, чим гірше студент впорався із завданнями до неї на першому етапі.

Отже, завдяки визначенню нерівнозначності (значущості) тестових завдань на етапі теоретичного оцінювання їх прототипів, встановленню складно-значущості тестового матеріалу на етапі експертної оцінки і проведенню контролю в два етапи, тест буде налаштований на визначення персональних знань студента конкретної спеціальності. Причому, завдяки закладеному в імітаційній моделі математичному обґрунтуванню рішень, що приймаються, виникає можливість мінімізувати суб'єктивізм під час проведення контрольних заходів, спрямованих на персоналізовану перевірку знань студентів.

Аналізуючи вище зазначене, можна зробити висновок, що під час двохетапного тестування в імітаційній моделі дається можливість перевіряти не лише знання окремих завдань тесту, але й виконувати перевірку розуміння матеріалу, що вивчається, з окремих тем, модуля, дисципліни в цілому. Тобто здійснювати персональну перевірку знань тих тем, які є значимими для конкретної дисципліни і викликали індивідуальні труднощі у самого студента.

Якщо за результатами першого етапу тестування (основної сесії контролю) не можна винести рішення (зробити висновок) про повноту знань конкретного студента, то пропонується провести додаткову сесію, під час якої кожен студент отримує свій набір завдань із проблемних тем особисто для нього. При цьому нова база тестових завдань добирається відповідно до цілей навчання дисципліни і складається зі складно-значущих завдань персонально для студента контрольованої спеціальності.

Завдяки використанню імітаційної моделі значущість можна пов'язати з сукупною складністю тестових завдань і якщо студент виконує тест заданої складно-значущості, то можна зробити висновок, що він досяг рівня розуміння дисципліни в цілому, хоча деякі завдання можуть бути не виконані.

### Література

1. Алексеев О. М. Визначення складності тестових завдань за методом попарних порівнянь / О. М. Алексеев // Педагогічний дискурс: збірник наукових праць. – Хмельницький, ХДПА, 2010. – Випуск 8. – С. 6-9.

2. Алексеев О.М. Імітаційна модель тестового контролю знань і умінь /О.М. Алексеев, Г.В. Алексеева //Комп'ютерно-орієнтовані системи навчання: збірник наукових праць. – К.: НПУ ім. М.П. Драгоманова, 2009. – Випуск 7 (14). – С. 65-71.

3. Бондаренко М.Ф., Семенец В.В., Белоус Н.В., Куцевич И.В., Белоус И.А. Оценивание тестовых заданий разных типов и определение их уровня сложности. //Искусственный интеллект – 2009. – №4. – С. 322-329.

4. Бондаренко М.Ф., Семенец В.В., Белоус Н.В., Куцевич И.В., Белоус И.А. Технология оценивания тестов в зависимости от типа и уровня сложности тестовых заданий на основе интегрированной модели // International Book Series "Information Science and Computing". – Sofia: Human Aspects of Artificial Intelligence. – 2009. –No:12. – С. 55-62.

5. Ефремова Н.Ф. Тестовый контроль в образовании: учебное пособие для студентов, получающих образование по педагогическим направлениям и специальностям /Н.Ф. Ефремова. – М.: Логос, 2007. – 368 с.

6. Зіньковський Ю.Ф., Мірських Г.О. Методика оцінювання рівнів складності навчальних тестів // Вісник Національного технічного університету України "КПІ" Серія – Радіотехніка. Радіоапаратобудування. –2010. – №41. – С.157-163.

7. Майоров А.Н. Теория и практика создания тестов для системы образования: Как выбирать, создавать и использовать тесты для целей образования /А.Н.Майоров. – М.: Народное образование, 2000. – 352 с.

8. Морев И. А. Образовательные информационные технологии. Часть 2. Педагогические измерения: Учебное пособие. – Владивосток: Изд-во Дальневост. ун-та, 2004. – 174 с.

9. Чельшкова М.Б. Теории и практике конструирования педагогических тестов: учебное пособие /М.Б.Чельшкова. – М.: Логос, 2002. – 432 с.