

АДАПТИВНІ ТЕСТИ: ЗАГАЛЬНІ ПОЛОЖЕННЯ

Abstract: In the article the problems of test knowledge control in modern educational establishments are considered. Methods and instruments of classic theory apart big possibilities do not solve many tasks, that gives modern level of education development. Adaptive tests in the given case let us solve the given questions and may be used effectively for solving any tasks of learning process optimization – evaluation of effectiveness of pedagogical innovations and technologies, monitoring etc.

Key words: knowledge control, adaptive test, effectiveness of test.

Анотація: У статті розглянуто проблеми тестового контролю знань у сучасних навчальних системах. Методики та інструментарій класичної теорії тестів попри великі можливості даної технології не вирішують багатьох задач, які ставить сучасний рівень розвитку освіти. Адаптивні тести у даному випадку дозволяють вирішувати ці питання і можуть бути ефективно використані для вирішення будь-яких задач оптимізації навчального процесу – оцінки ефективності педагогічних інновацій і технологій, моніторингу і т.д.

Ключові слова: контроль знань, адаптивний тест, ефективність тесту.

Аннотация: В статье рассмотрены проблемы тестового контроля знаний в современных учебных системах. Методики и инструментальный классической теории тестов, несмотря на большие возможности данной технологии, не решают многих задач, которые ставит современный уровень развития образования. Адаптивные тесты в данном случае позволяют решать эти вопросы и могут быть эффективно использованы для решения любых задач оптимизации учебного процесса – оценки эффективности педагогических инноваций и технологий, мониторинга и т.д.

Ключевые слова: контроль знаний, адаптивный тест, эффективность теста.

1. Вступ

Інформатизація та комп'ютеризація навчального процесу призводить до зміни та трансформації сучасних навчальних систем. Такі зміни нероздільно пов'язані із постійним вдосконаленням різних складових частин таких систем, де одним із визначальних компонентів є контроль знань. У традиційних формах навчання управління процесом засвоєння знань, визначенням рівня, тобто виставлення оцінки, займається викладач – людина-інтелектуал, якій властиво помилятися, проте вона володіє інтуїцією, семантикою. Одним із найбільш об'єктивних 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]. Тест – це система завдань зростаючої складності та спеціальної форми, що дозволяє якісно оцінити структуру знань і ефективно визначити рівень знань учасників тестування. Автоматизована навчальна система повинна якомога адекватніше оцінити набуті знання студентів. Тобто вона повинна володіти засобами комп'ютерного контролю знань. Деталі, пов'язані з комп'ютерним контролем знань, є об'єктом неабиякого інтересу для викладачів вузів і творців засобів реалізації такого контролю. Крім того, питання комп'ютерного контролю недостатньо широко висвітлені в теоретичному плані, і інтерес до них, як правило, реалізується в більшості випадків шляхом створення чергової програми комп'ютерного контролю із завчасно складеним набором контрольних завдань. Проте, в області комп'ютерного контролю ще не все так очевидно, як здається з першого погляду. Існують деякі важливі питання, які або висвітлені в малодоступній літературі, або пропрацьовані недостатньо прозоро [2].

2. Традиційні тести

Перші наукові праці по 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 із перших займався конструюванням та впровадженням тестового контролю в американській школі всевітньо 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 на перевірку 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 меншу кількість завдань, але який може відобразити весь спектр задач.

Поняття "довжина тесту" введено на початку XX століття Ч. Спірманом і означає кількість завдань у тесті. Чим довший тест, тим більше в ньому завдань. Від кількості завдань деяким чином залежить точність виміру знань. Для тестування намагаються відібрати мінімально достатню кількість завдань, що дозволяє порівняно точно визначити рівень і структуру підготовленості. Інтерпретація результатів тестування ведеться переважно з опорою на середню арифметичну і на так звані процентні норми, що показують, скільки відсотків випробуваних мають тестовий результат гірший, ніж у будь-якого іншого випробуваного. Така інтерпретація тестових результатів називається нормативно-орієнтованою.

Навчальний тест визначається як система завдань визначеного змісту, зростаючої складності, специфічної форми, що дозволяє якісно й ефективно виміряти рівень і оцінити структуру підготовленості учнів. Навчальний тест по своїй суті є класичним представником традиційних тестів. Для кращого розуміння визначення навчального тесту розглянемо коротке тлумачення його основних термінів. Будь-який тест складається з тестових завдань, але останні представляють не просто сукупність довільно об'єднаних завдань, а утворюють систему.

Один із 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. Інша сторона питання полягає в тому,

що від часу тестування істотно залежить якість результатів. Кожен тест має оптимальний час тестування, зменшення чи перевищення якого знижує якісні показники тесту. Оптимальний час тестування визначається емпірично, за показником дисперсії тестових даних. Якщо по осі абсцис відкласти час тестування, а по осі ординат – значення дисперсії тестових результатів, що отримується після кожної спроби контролю, то, з'єднавши точки, одержимо представлення зміни дисперсії; де максимум значення останньої вкаже на оптимум часу, необхідного для тестового контролю.

Якість тесту традиційно зводиться до визначення міри його надійності і валідності отриманих результатів. Якісним, як і об'єктивним, можна назвати тільки той метод виміру, який обґрунтований науково і здатний дати необхідні результати. Валідність означає відповідність тестових результатів меті, заради якої проводилося тестування. Валідність залежить від якості завдань, їхнього числа, від ступеня повноти і глибини охоплення змісту навчальної дисципліни у завданнях тесту. Крім того, валідність залежить також від балансу і розподілу завдань за складністю, від методу добору завдань у тест із загального банку завдань, інтерпретації тестових результатів, організації збору даних, від підбору вибіркової сукупності випробуваних. Специфічна форма тестових завдань говорить про те, що завдання тесту являють собою не питання і не задачі, а завдання, сформульовані у формі висловлювань, істинних чи помилкових, у залежності від відповідей. Визначений зміст означає використання в тесті тільки такого контрольного матеріалу, що відповідає змісту навчальної дисципліни; інше в навчальний тест не включається. Зміст тесту існує, зберігається і передається в одній із чотирьох основних форм завдань. Поза тестовими формами ні тест, ні його зміст не існують. 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 визначати 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 учня. Показник складності завдання розглядається як важливий системо- і одночасно структуроутворюючий фактор тесту.

Наступний критерій – це критерій логічної визначеності тестового завдання. Його можна сформулювати словами, близькими до формулювання Х.Каррі: “Завдання є визначеним, якщо на нього можна відповісти позитивно чи негативно і якщо існує ефективний процес для знаходження такої відповіді”. Відповідь на завдання навчального тесту являє собою коротке судження, пов'язане за змістом і формою зі змістом завдання. Кожному завданню ставляться у відповідність відповіді правильні і неправильні. Критерії правильності заздалегідь визначаються авторами тесту. Ймовірність правильної відповіді на будь-яке завдання залежить від співвідношення рівня знань

випробуваного і рівня складності завдання. При наявності порівняльної шкали цю ймовірність виражають значенням від нуля до одиниці або шкалами інших видів. Оцінка відповідей за ступенем їх правильності проводиться досить рідко, але при необхідності завдання створюються з такими відповідями, які можуть бути правильні в різному їх сприйнятті.

За допомогою тестування перевіряються знання, уміння, навички та уявлення. З погляду навчальних вимірів корисно ввести два основних показники якості знань: рівень і структура знань. Вони оцінюються за допомогою реєстрації оцінок як за знання, так і за незнання всіх необхідних компонентів матеріалу, що перевіряється. Для об'єктивізації цього процесу всі компоненти системи повинні бути однакові. Однаковими є і правила виставлення оцінок. Ці умови відкривають дорогу для об'єктивного порівняння індивідуальних структур знання і незнання. Рівень знань виявляється при аналізі відповідей на всі завдання тесту. Чим більше правильних відповідей, тим вищий індивідуальний тестовий бал. Звичайно в системах дистанційної освіти цей тестовий бал асоціюється з поняттям "рівень знань" і проходить процедуру уточнення на основі тієї чи іншої моделі виміру [4]. Той самий рівень знань може бути отриманий за рахунок відповідей на різні завдання. Наприклад, у тесті з тридцяти завдань студент одержав десять балів. Ці бали скоріш за все отримані за рахунок правильних відповідей на перші десять, порівняно легких завдань. Властиву для такого випадку послідовність одиниць, а потім нулів можна назвати правильним профілем знань учня. Якщо тестовий бал нижчий від необхідного рівня (вимірюваного критерію), то виявлені при цьому знання, уміння, навички і представлення вказують на докритеріальний рівень підготовленості. Структура знань оцінюється на основі послідовності правильних і неправильних відповідей на завдання зростаючої складності. Формою представлення індивідуальної структури знання і незнання є профіль знань, що представляється послідовністю одиниць і нулів, одержуваних кожним студентом.

Профіль знань являє собою упорядкований набір оцінок (вектор-рядок) у матриці тестових результатів. Якщо випробуваний відповідає правильно на перші, порівняно легкі завдання, можна говорити про правильну структуру знань. Профіль називається правильним, якщо в рядку балів у випробуваного всі нулі йдуть за одиницями. Якщо ж виявляється протилежна картина, (випробуваний правильно відповідає на важкі завдання і неправильно - на легкі), то це суперечить логіці тесту і тому така структура знань може бути названа інвертованою. Вона зустрічається рідко і найчастіше через порушення вимоги розташовувати завдання в міру зростаючої складності. За умови, що тест зроблений правильно, кожен профіль свідчить про структуру знань. Цю структуру можна назвати елементарною (оскільки є ще факторні структури, які виявляються за допомогою методів факторного аналізу).

Роль структури знань багаторазово підкреслювалася видатним педагогом А. Дістервегом, а також психологом Д. Брунером. Останній вважає, що "виклад структури знань, оволодіння цією структурою, а не просте засвоєння фактів і технічних прийомів є центральним моментом". Розробники систем дистанційного навчання повинні прагнути, у першу чергу, до формування правильних індивідуальних структур знань, у яких не було б прогалин (розривів у знаннях), і на цій основі підвищувати рівень підготовки. Рівень знань у значній мірі залежить від особистих зусиль і здібностей тих, хто навчається за допомогою таких систем, у той час як структура знань помітно

залежить від правильної організації навчальної системи, індивідуалізації навчання, рівня поданого матеріалу, об'єктивності контролю – загалом, від усього того, чого зазвичай не вистачає.

Найкращий підхід при засвоєнні знань учнем полягає в використанні дедуктивно-індуктивного підходу, що відповідає принципу змішаного екстремуму [5], який вирішує протиріччя між засвоєнням від простого до складного або від загального до часткового.

Класи і види класичних тестів

Традиційний тест являє собою єдність щонайменше трьох систем [6]:

1) змістовної системи знань, яка описується мовою навчальної дисципліни, що перевіряється;

2) формальної системи завдань зростаючої складності;

3) статистичних характеристик завдань і результатів.

Традиційний навчальний тест потрібно розглядати з двох сторін: як метод педагогічного виміру і як результат застосування тесту. Обидва ці критерії характеризують тест із різних сторін, тому що тест треба розуміти одночасно і як метод, і як результат навчального виміру. Одне доповнює інше. Тест як метод неможливий без результатів, що підтверджують якість його самого і якість оцінок виміру випробуваних різного рівня підготовленості.

Головна ідея традиційного тесту – мінімальним числом завдань за короткий час, швидко, якісно і з найменшими витратами порівняти знання як можна більшої кількості учнів. Власне цим і відображається ідея ефективності навчальної діяльності в області контролю знань у системах дистанційного навчання. До традиційних тестів відносяться тести гомогенні і гетерогенні.

Гомогенний тест являє собою систему завдань зростаючої складності, специфічної форми і визначеного змісту – систему, створювану з метою об'єктивного, якісного й ефективного методу оцінки структури і виміру рівня підготовленості учнів по одній навчальній дисципліні. Легко бачити, що у своїй основі визначення гомогенного тесту збігається з визначенням традиційного тесту. Гомогенні тести поширені більше інших. Вони створюються для контролю знань по одній навчальній дисципліні чи по одному розділу такої, наприклад, об'ємної навчальної дисципліни, як фізика. У гомогенному педагогічному тесті не допускається використання завдань, що виявляють інші властивості. Наявність останніх порушує вимога дисциплінарної чистоти навчального тесту. Адже кожен тест вимірює щось заздалегідь визначене.

Гетерогенний тест являє собою систему завдань зростаючої складності, специфічної форми і визначеного змісту – систему, створену з метою об'єктивного, якісного й ефективного методу оцінки структури і виміру рівня підготовленості по декількох навчальних дисциплінах. Нерідко такі тести містять і психологічні завдання для оцінки рівня інтелектуального розвитку. Звичайно гетерогенні тести використовуються для комплексної оцінки випускника шкіл, оцінки особистості при прийомі на роботу і для добору найбільш підготовлених абітурієнтів при прийомі у вузи. Оскільки кожен гетерогенний тест складається з гомогенних тестів, інтерпретація результатів тестування ведеться по відповідях на завдання кожного тесту (тут вони називаються шкалами) і, крім того, за допомогою різних методів агрегації балів робляться спроби дати загальну оцінку підготовленості.

Нагадаємо, що традиційний тест являє собою метод діагностичної оцінки учня, у якому всі відповідають на одні й ті ж завдання, в однаковий час, в однакових умовах і з однаковою оцінкою. При такій орієнтації задачі визначення точного обсягу і структури освоєного навчального матеріалу відступають на задній план. У тест відбирають таку мінімально достатню кількість завдань, яка дозволяє порівняно точно визначити не "хто що знає", а "хто знає більше". Тут висновок добудовується рейтингом: завдання \Rightarrow відповіді \Rightarrow висновки про знання \Rightarrow рейтинг, що розуміється як висновок про місце чи ранг випробуваного.

Інтегрованим можна назвати тест, який складається із системи завдань, що відповідають вимогам інтегрованого змісту, тестової форми, які збільшують складність завдань, націлених на узагальнену підсумкову діагностику підготовленості. Діагностика проводиться за допомогою таких завдань, правильні відповіді на які вимагають інтегрованих (узагальнених, явно взаємозалежних) знань в області двох і більшого числа навчальних дисциплін. Створення таких тестів дається тільки тим викладачам, що володіють знаннями ряду навчальних дисциплін, розуміють важливу роль міжпредметних зв'язків у навчанні, здатні створювати завдання, правильні відповіді на які потребують знань різних дисциплін і умінь застосовувати такі знання. Перевага інтегрованих тестів перед гетерогенними полягає в більшій змістовній інформативності кожного завдання й у меншому числі самих завдань. Методика створення інтегрованих тестів подібна до методики створення традиційних тестів, за винятком роботи з визначенням змісту завдань. Для створення змістовних інтегрованих тестів використання експертних методів є обов'язковим. Це пов'язано з тим, що тільки експерти можуть визначити адекватність змісту завдань цілям тесту. Але, насамперед, самим експертам важливо буде визначитися з метою утворення і вивчення тих чи інших освітніх програм, а потім і домовитися між собою по принципових питаннях, залишивши для експертизи лише варіації в розумінні ступеня значимості окремих елементів у загальній структурі підготовленості.

3. Адаптивні тести

Доцільність адаптивного контролю впливає з необхідності раціоналізації традиційного тестування. Кожен викладач розуміє, що добре підготовленому учню немає необхідності давати легкі завдання, тому що занадто висока імовірність правильного рішення. До того ж легкі 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 було знайти порівняння в одній шкалі для міри складності завдань і міри рівня знань. Ця міра була знайдена в теорії педагогічних вимірів. Датський математик Г. Раш назвав цю міру словом "логіт" [7]. Після появи комп'ютерів вона лягла в основу методики адаптивного контролю знань, де використовуються способи регулювання складності і кількості пропонованих завдань, у залежності від відповіді учнів. При успішній відповіді наступне завдання комп'ютер підбирає більш важким, при неуспішному - легшим. 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.

Зараз у світі відомо три варіанти адаптивного тестування. Перший називається пірамідальним тестуванням. При відсутності попередніх оцінок всім дається завдання середньої складності і вже потім, у залежності від відповіді, кожному дається завдання легше чи важче. Другий варіант – flexilevel-контроль починається з рівня складності, який обирає саме той, хто проходить тестування, з поступовим наближенням до реального рівня знань. Третій варіант – stradaptive (від англ. stratified adaptive), коли тестування проводиться за допомогою банку завдань, розділених за рівнями складності. При правильній відповіді наступне завдання береться з верхнього рівня, при неправильному – з нижнього. Таким чином, адаптивний тест являє собою варіант автоматизованої системи тестування з заздалегідь відомими параметрами складності і здатністю кожного завдання, що диференціює. Ця система створена у вигляді комп'ютерного банку завдань, упорядкованих відповідно до характеристик, які нас цікавлять.

Критеріально-орієнтовані тести

Якщо головною задачею є прагнення з'ясувати, які елементи змісту навчальної дисципліни засвоєні, то це випадок предметно-педагогічного підходу до інтерпретації результатів тестування. При цьому визначається: що з головної сукупності завдань випробуваний знає і чого не знає. Інтерпретація результатів ведеться педагогами, мовою навчальної дисципліни. Висновок вибудовується уздовж логічного ланцюжка: зміст навчальної дисципліни \Rightarrow головна сукупність завдання для виміру знань \Rightarrow тест як вибірка завдань з цієї сукупності, відповіді \Rightarrow імовірнісний висновок про знання навчальної дисципліни. При орієнтації на такі тести потрібні велика кількість завдань і досить повне визначення змісту досліджуваної дисципліни. Інтерпретація результатів ведеться педагогами. Другий вид тестів пов'язаний з орієнтацією на такі конкретні цілі і задачі, як, наприклад, перевірка рівня засвоєння порівняно короткого переліку необхідних знань, умінь і навичок, що виступають як заданий чи стандарт критерію засвоєння. Наприклад, для атестації випускників дистанційних курсів важливо мати такі завдання, що дозволяють робити висновок про мінімально припустиму їх компетентність (Minimum Competency Tests). При перевірці мінімально припустимого рівня знань зміст завдань носить принципово полегшений характер. Оскільки такі завдання повинні виконувати усі випускники, допущені системою до атестації, тут важко говорити про тести як метод об'єктивного й ефективного виміру різного рівня підготовленості, у строгому контексті поняття "тест". У такому випадку говориться про тести з критеріально-орієнтованою інтерпретацією результатів. Висновок вибудовується уздовж логічного ланцюжка:

завдання \Rightarrow відповіді \Rightarrow висновки про відповідність заданому критерію.

Під критеріально-орієнтованою інтерпретацією мається на увазі порівняння змісту атестаційних матеріалів з результатами тестування і висновок, що із заданого стандарту і на якому рівні реально засвоєно. При критеріально-орієнтованій інтерпретації потрібна трохи менша кількість завдань, за допомогою яких визначається, що випробуваний знає і чого не знає з заданого стандарту. Інакше кажучи, тут відповіді оцінюються не до всієї області (Domain) необхідних знань, а тільки до області, обмеженої конкретним стандартом чи рівнем (критерієм) знань.

Концепція добору змісту тестових завдань

Зміст тесту можна визначити як оптимальне відображення навчального матеріалу в системі тестових завдань [8]. Слова "оптимальне відображення" припускають необхідність добору такого контрольного матеріалу, відповіді на який з високою імовірністю (більше 95 %) свідчили б про рівень підготовленості кожного учня. Оптимальне відображення припускає також можливість вибору придатної форми і необхідного числа завдань. Виділяють чотири форми:

- 1) завдання з вибором одного чи декількох правильних відповідей з числа запропонованих;
- 2) завдання відкритої форми, де відповідь дописується, у відведеному для цього місця;
- 3) завдання на встановлення відповідності;
- 4) завдання на встановлення правильної послідовності дій.

Кількість завдань традиційного тесту, звичайно, буває не менше тридцяти. Банк завдань адаптивного тесту містить більшу кількість завдань. При тестуванні перевіряється не весь програмний матеріал, а лише та частина навчальної програми, що входить у зміст. Це знання, що підлягають обов'язковій перевірці в кожній системі дистанційного навчання. Деякі елементи знань, що перевіряються (переважно з окремих тем), використовуються тільки в поточному контролі. Інші елементи, що охоплюють знання декількох тем, використовуються в рубіжному контролі, наприклад, наприкінці навчальної чверті. І, нарешті, у підсумковому контролі використовуються завдання, правильні відповіді на які вимагають знання багатьох, а й іноді і всіх тем, вивчених протягом навчального року.

Складність тесту і його вплив на оцінку знань

Якщо педагогічний тест визначити коротко як систему завдань зростаючої складності, то стане зрозуміло, що складність завдань є найважливішим тестоутворюючим показником [9]. Перша вимога до тестових завдань: у тесті завдання повинні розрізнятися за рівнем складності, що впливає з даного раніше визначення тесту і розглянутого принципу. Завдання мають шанс стати тестовими лише після емпіричної перевірки міри їх складності. Показник складності тесту і тестових завдань є змістовним і формальним одночасно. Змістовним показником, тому що в ефективному тесті складність може залежати як від змісту, так і рівня підготовленості самих випробуваних, у той час як у неефективному тесті на результати починають помітно впливати форма завдань (особливо, якщо вона не адекватна змісту), неефективна організація тестування і т.д. Формальна складова сторони показника складності виникає при розгляді тестування як процесу протистояння запропонованим завданням. Одержуваний при цьому результат корисно розглядати як результат такої боротьби. При спрощеному тлумаченні кожного випадку протистояння з черговим завданням звичайно розглядаються тільки два результати: перемога при правильному рішенні завдання, де отримується один бал, чи поразка, за що дається нуль балів. Оцінка результату такої боротьби залежить від співвідношення рівня знання до рівня складності завдання, від обраної одиниці виміру знань і від заздалегідь прийнятого правила: що вважати "перемогою", а що ні. Традиційною мірою складності кожного завдання довгі роки була частка правильних відповідей у групах, де проводиться тестування. Цю міру позначимо символом p_j , де індекс j указує на номер завдання (1,

2 і т.д.). Наприклад, якщо правильні відповіді на третє завдання тесту оцінювати одним балом, а неправильні - нулем, то значення показника p_3 можна знайти з елементарного відношення:

$$p_3 = R_3 / N,$$

де R_3 означає число правильних відповідей на дане завдання, а N – загальне число випробуваних у групі. Загальна формула розрахунку частки правильних відповідей на будь-яке завдання (j) має відповідний вигляд:

$$p_j = R_j / N.$$

Показник p_j довго використовувався як міра складності у класичній теорії тестів. Пізніше була усвідомлена суттєва неточність, що міститься в ній: збільшення значення p_j вказує не на зростання складності, а, навпаки, на зростання легкості, якщо можна використовувати таке слово. Тому в останні роки з показником складності завдань стали асоціювати протилежну статистику – частку неправильних відповідей (q_j). Ця частка обчислюється з відношення числа неправильних відповідей (W_j – від англ. слова Wrong – неправильний) до числа випробуваних (N):

$$q_j = W_j / N.$$

Відповідним чином приймається, що $p_j + q_j = 1$.

У класичній теорії тестів багато років розглядалися тільки емпіричні показники складності. У нових варіантах психологічних і педагогічних теорій тестів більше уваги стали приділяти характеру розумової діяльності учнів у процесі виконання тестових завдань різних форм. Зміст тесту не може бути тільки легким, середнім чи важким. Тут повною мірою виявляється відома думка про залежність результатів застосовуваного методу. Легкі завдання тесту створюють тільки видимість наявності знань в учнів, тому що ними перевіряються мінімальні знання. Спотворює результати тестування і підбір свідомо важких завдань, у результаті чого в більшості студентів виявляються занижені бали. Орієнтація на важкі завдання нерідко розглядається як засіб посилення мотивації до навчання. Однак цей засіб діє неоднозначно. Одних важкі завдання можуть підштовхнути до навчання, інших – навпаки. Подібна орієнтація спотворює результати й у підсумку, знижує якість педагогічного виміру. Якщо тест побудований строго із завдань зі зростаючою складністю, то цим відкривається шлях до створення однієї із самих цікавих шкал виміру – шкали Л. Гутмана. Розповсюджена донедавна рекомендація включати в тест більше завдань середньої складності виправдана з погляду визначення надійності виміру по формулах класичної теорії тестів. Існуючі в цій теорії методи оцінки надійності тесту дають зниження надійності при включенні в тест помітно легких і важких завдань. У той же час захоплення завданнями однієї лише середньої складності приводить до серйозної деформації змісту тесту: останній втрачає здатність відображати зміст досліджуваної дисципліни, у якій завжди є легкий і важкий матеріал. Таким чином, у погоні за теоретично високою надійністю губиться змістовна валідність тесту. Прагнення ж підняти валідність тесту нерідко супроводжується зниженням його надійності. У науковій літературі це називається парадоксом Ф. Лорда, який описав його з точки зору статистичної теорії тестів.

4. Знання як об'єкт для виміру

Класифікація знань

При тестуванні виявляються як знання, так і незнання, що досягаються за допомогою системи завдань, націлених на виявлення знань. У рамках інформаційної моделі навчання знанням визначається як інформація про ту чи іншу предметну область, зафіксовану в реченнях звичайних чи спеціалізованих мов. Тому можна сказати, що оцінка якості знань залежить не тільки від змісту навчальної дисципліни, але і від якості фіксації у змісті тесту таких елементів, які дозволили б оцінити міцність, повноту, глибину, оперативність, гнучкість, конкретність і узагальненість, системність і систематичність, а також і інші відомі в літературі характеристики знань. Для рішення задач виміру знань корисне ще одне визначення знання: це об'єктивно і суб'єктивно достатнє визнання істинності судження. Знання, що виявляється у відповіді на конкретне завдання тесту, можна розглядати як логічне судження. В.С. Bloom і R.M. Gagne запропонували таку класифікацію знань [10]:

- 1) знання назв, імен;
- 2) знання змісту назв і імен;
- 3) фактуальні знання;
- 4) знання визначень;
- 5) порівняльні знання;
- 6) класифікаційні знання;
- 7) знання протилежностей, протиріч, синонімічних і антонімічних об'єктів;
- 8) асоціативні знання;
- 9) причинні знання, знання причинно-наслідкових відношень, знання основ, принципів класифікації;
- 10) процесуальні, алгоритмічні, процедурні знання;
- 11) технологічні знання;
- 12) узагальнені, системні знання;
- 13) імовірнісні знання;
- 14) абстрактні знання;
- 15) структурні знання;
- 16) методологічні знання.

Перелічені види знань не утворюють поки що повної класифікаційної системи і тому допускають можливість помітного розширення представленої номенклатури, заміни одних видів знань іншими, об'єднання їх у різні групи.

Оцінювання, оцінка і вимір знань

У світовій педагогічній літературі використовується близьке до виміру поняття "evaluation", що можна перекласти як "оцінювання", маючи на увазі не тільки кінцевий результат, але і процес формування оцінки. Відповідно виділяються два основних види оцінювання: формуюче і підсумкове. Мета першого – за допомогою оцінювання впливати на поточний процес навчання за рахунок встановлення зворотного зв'язку від студента до викладача. Мета другого – одержати підсумкові

результати навчання. Оцінка, як поняття, має два основних значення: як судження про цінність (чи значимість) і як приблизна характеристика деякої величини. Оцінки поділяються на дві групи: абсолютні і відносні. У формулюванні відносних оцінок використовуються терміни "добре", "погано" і т.п. Оцінки викладачів вузів ближче до абсолютних оцінок, розділених на чотири рівні (градації), відмінно, добре, задовільно і незадовільно.

Вимір знань і його об'єктивність

Для виміру знань вимагаються емпіричні показники. Як такі, зараз виступають будь-які методи контролю знань, хоча інтуїтивно зрозуміло, що вони неоднаково представляють знання. Априорі можна стверджувати, що тест це робить краще через більшу його обґрунтованість. В основу обґрунтування якості тесту, крім згаданої вище інтерпретації одержуваних результатів, покладені концепція і модель виміру, а також дані статистичної обробки й інтерпретації результатів перевірки знань. Зазвичай інтерпретація результатів тестування не представляється важкою справою. Отримана сума балів приймається за відповідний рівень знань, хоча кожний визнає, що знання і бали, отримані за показниками знань, різні явища. Адже можна придумати тисячі різних показників і усі вони будуть з різною кількістю балів вказувати на рівень знань. Головна теза інтерпретації – не хто, що знає, а хто за даним набором завдань виявився вище чи нижче. Хоча при такій інтерпретації завжди залишається місце для критики тесту з позиції неповноти його змісту, у практиці цей підхід завжди виявлявся вирішальним через можливість скоротити число завдань і прийняти рішення на основі тесту з меншою кількістю завдань. Власне кажучи, це форма редукції, успіх якої залежить від якості розробки тесту і від обґрунтування валідності скороченого тесту з нормативно орієнтованою інтерпретацією.

Поняття ефективності тесту

Ефективним можна назвати тест, який краще, ніж інші тести, вимірює рівень знань за певним критерієм, з меншою кількістю завдань, якісніше, швидше, дешевше, і все це, по можливості, у комплексі. З поняттям "ефективність" близьке до нього за змістом поняття "оптимальність". Останнє трактується якнайкраще з можливих варіантів, з погляду задоволення декільком критеріям ефективності, узятим по черзі чи разом. Тому ефективний тест не може складатися з неефективних завдань. У такому випадку природно порушити питання про ознаки, що відрізняють ефективне завдання від неефективного. З погляду на зміст ефективне завдання перевіряє важливий елемент змісту навчальної дисципліни, що нерідко називають ключовим для необхідної структури знань випробуваних. У тест відповідно включаються тільки такі завдання, які експерти визнають як ключові елементи досліджуваної навчальної дисципліни. У визначенні ефективності тесту звертається увага на два ключових елементи - це кількість завдань тесту і рівень підготовленості тих, хто проходить цей тест. Якщо з якого-небудь тесту з великою кількістю завдань зробити оптимальний вибір меншого числа, то може утворитися система, що не поступається за своїми властивостями тесту з порівняно великою кількістю завдань. Тест із меншою кількістю завдань у такому випадку можна називати порівняно більш ефективним.

Ефективність тестів може залежати і від форми. Порушення тестової форми завжди приводить до гіршого вираження змісту і до гіршого розуміння змісту завдання випробуваними. Ефективність тесту залежить також і від принципу підбору завдань. Якщо підбирати завдання для виміру на всьому діапазоні зміни складності, то знижується точність виміру на окремій ділянці. І навпаки, якщо прагнемо точно вимірити знання випробуваних, наприклад, середнього рівня підготовленості, то це буде вимагати більше завдань саме даного рівня складності. Тому тест не може бути ефективним на всьому діапазоні підготовленості випробуваних. Він може бути більш ефективним на одному рівні знань і менше – на іншому. Саме цей зміст вкладається в поняття диференційної ефективності тесту.

4. Висновок

Виходячи із вищесказаного, можна зробити висновок, що практично всі змінні, які використовуються для аналізу освітнього процесу, є латентними (прихованими), тобто безпосередньо не вимірюваними. Найбільш часто використовується в освітньому процесі латентна змінна – рівень знань тих, хто навчається. При використанні традиційної або класичної системи тестування латентна змінна – рівень знань – вимірюється як частка правильних відповідей. Перевага цієї системи вимірювання в тому, що вона проста, наочна і легко доступна для розуміння. Проте у цій системі є істотні недоліки. Так, оцінка рівня знань залежить від набору тестових завдань. Результати ж об'єктивних вимірювань не повинні залежати від вимірювального інструменту, що використовується (в даному випадку, набору тестових завдань), а також результати вимірювання при використанні класичної системи тестування є істотно нелінійними. Система тестування, а точніше система вимірювання латентних змінних, що використовується, на основі адаптивних тестів принципово відрізняється від найвідомішої класичної або традиційної системи тестування по багатьох аспектах. Обмеження, які з'являються при використанні методів класичної теорії тестів, при цьому багато в чому знімаються. Перш за все це можливість на відміну від традиційної системи тестування одержати більш об'єктивні оцінки. Так, при використанні традиційної системи тестування (вимірювання) рівень знань визначається часткою правильних відповідей. Проте така оцінка залежить від складності завдань у тесті. Якщо тест складається з легких завдань, частка правильних відповідей буде високою, якщо тест складається з важких завдань, то – низькою, тобто оцінка рівня знань при використанні традиційної системи вимірювань не може вважатися об'єктивною. Принциповою ж відмінністю системи тестування на основі адаптивних тестів є те, що оцінка рівня знань студентів не залежить від складності тесту, тобто є об'єктивною. Це означає, що такі оцінки рівня знань студентів можуть бути ефективно використані для вирішення будь-яких задач оптимізації навчального процесу — оцінки ефективності педагогічних інновацій технологій, моніторингу і т.д. Можливості використання вищезазначених технологій будуть розглянуті в наступних публікаціях.

СПИСОК ЛІТЕРАТУРИ

1. Аванесов В.С. Теория и методика педагогических измерений (материалы публикаций). – М.: 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". – 2004. – Т. 7, № 2. – С. 158–169.

3. Аванесов В.С. Научные проблемы тестового контроля знаний. – М., 1994. – 135 с.
4. Федорук П.І. Технологія розробки навчального модуля в адаптивній системі дистанційного навчання та контролю знань // Математичні машини і системи. – 2005. – № 3. – С.155–165.
5. Benjamin D. Wright & Mark H. Stone Best Test Design. – 2001. – 223 p.
6. Теслер Г.С. Новая кибернетика. – Киев: Логос, 2004. – 404 с.
7. Bond Trevor G., Christine M. Fox Christine M. Applying the Rasch Model. Fundamental Measurement in the Human Sciences. – 2001. – 280 p.
8. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.
9. Lord F.M. Application of Item Response Theory to Practical Testing Problems. Hillsdale N-J. Lawrence Erlbaum Ass., Publ., 1980. – 266 p.
10. Дюк В.А. Компьютерная психодиагностика. – Спб.: Братство, 1994. – 360 с.

Стаття надійшла до редакції 01.06.2007