

нять эти знания на практике. Таким образом, учащийся имеет возможность управлять процедурой получения и использования знаний, учиться алгоритмически мыслить.

Вместе с тем следует отметить, что компьютеризация обучения не отменяет и не заменяет процесса общения педагога с учащимся, а лишь помогает и тому и другому более эффективно участвовать в процессе обучения и уделять больше внимания элементам творчества, на которые раньше зачастую не хватало времени. Это становится возможным за счет более эффективной самостоятельной работы учащегося и более наглядного представления нового материала.

ЛИТЕРАТУРА

1. Верлань А. Ф. Информационные технологии в современной школе. — Каменец-Подольский, 1996.
2. Жалдак М. И. Система подготовки учителя к использованию информационных технологий в учебном процессе. — К., 1990.
3. Кан-Калик В. А. Грамматика общения. — М., 1995.
4. Караковский В. А. Воспитательная система школы: Педагогические идеи и опыт формирования. — М., 1992.
5. Коротяев Б. И. Педагогика как совокупность педагогических теорий. — М., 2001.
6. Краевский В. В. Что такое философия образования. — М., 1991.
7. Кудрявцев В. Т. Проблемное обучение: истоки, сущность, перспективы. — М.: — Знание, 1991.
8. Кулыгина Л. С. Активизация учения: Сущность и содержание. — М.: Педагогика. — 1994. — №1.
9. Левитан К. М. Основы педагогической деонтологии: Учебное пособие для высшей школы. — М., 1994.
10. Легенький Г. И. Цель и способы воспитания. — М., 1990.
11. Леклерк М., Дюбек Л., Бегин И. Оценка педагогического обеспечения в Канаде // Prospekt. — 1997.
12. Лернер И. Я. Процесс обучения и его закономерности. — М., 1980.
13. Лихачев Б. Т. Педагогика. — М., 1996.
14. Монахов В. М. Тенденция развития содержания общего среднего образования /Педагогика, 1992. — №1–2.
15. Педагогика / Под ред. П. И. Пидкасистого. — М., 1995.
16. Педагогика / Под ред. Ю. К. Бабанского. — М., 1998.
17. Педагогика и логика / Г. Щедровицкий, В. Розин, В. Алексеев, Н. Непомнящая — М., 1993.
18. Подласый И. П. Педагогика. В 2 кн. — М., 1999.
19. Пряникова В. Г., Равкин З. И. История образования и педагогической школы: Учебник-справочник. — М., 1995.
20. Ратанова Г. А. Способы активизации познавательной деятельности школьников // Вопросы психологии. — 1990. — №5.
21. Репкин В. В. Развивающее обучение как система: Пособие для учителей, студентов пед.институтов и работников народного образования. — Томск, 1993.
22. Смирнов А. Б. Matlab7. — СПб.: БХВ-Петербург, 2005.
23. Шевардин Н. И. Социальная психология в образовании. — М., 1995.
24. Шилова М. Н. Учителю о воспитанности школьников. — М., 1997.
25. [Интернет-ресурс] <http://www.vusnet.ru>

Інна ГРОД

КОНЦЕПЦІЯ КІЛЬКІСНОГО ОЦІНЮВАННЯ ОБ'ЄКТИВНОСТІ ПЕДАГОГІЧНОГО ТЕСТУВАННЯ ЗНАНЬ

У статті йдеться про часто згадувані недоліки «класичних» підходів до автоматизованого тестування, про аспекти, які забезпечують об'єктивність контролю знань і розрахунок рівня об'єктивності, про визначення природності того чи іншого методу оцінювання і ціль тестування, залежно від чого інтерпретуються результати тесту, про дві форми організації комп'ютерних тестів.

Тестування як спосіб контролю знань має безсумнівними переваги, найважливішою з яких слід визнати універсальність, наповненість, результативність і невелику затрату праці. Та аналіз автоматизованого тестування вищих та інших навчальних закладів показує, що багато викладачів насторожено і навіть негативно ставляться до систем, що використовуються для такого тестування. Серед найбільш важливих недоліків «класичних» підходів до автоматизованого тестування, названих як причини такого негативного ставлення, можна вказати:

- неможливість врахування неповних (не зовсім правильних) відповідей учня, що приймаються до уваги педагогом при «живому» оцінюванні знань;
- неможливість автоматизації різних методик оцінювання, що широко застосовуються в педагогічній практиці і спрямовані на вирішення конкретних завдань (оцінка широти або глибини знань, врахування відносної важливості окремих тем чи розділів предмета, що вивчається, вибір складності тесту з врахуванням рівня підготовки і самооцінки того, хто тестується, стимуляція правильних відповідей);
- значну затрату праці для ручного формування необхідної множини тестових завдань;
- невідповідність інформаційної моделі, за якою перевіряються знання учня, еталонної моделі знань предметної області;

Остання проблема більш відома як проблема об'єктивності контролю знань. Педагогічний контроль вважається об'єктивним, коли його первинні показники, оцінка й інтерпретація не залежать від суб'єктивних суджень особи, що перевіряє знання. При визначенні рівня об'єктивності повинні братися до уваги як здібність і можливість учня демонструвати наявні у нього знання, так і спосіб формування моделі, стосовно якої будуються знання учня. З одного боку, на оцінку знань впливає суб'єктивна здатність учня формулювати відповіді на запропоновані конкретні запитання в формі, необхідній і достатній для їх обліку. З іншого боку, в багатьох випадках виставлена оцінка не може називатись достовірною через суб'єктивність моделі знань, з якою зрівнюються знання учня. З формальної точки зору оцінка функціонально залежить від ступеня відповідності моделі знань учня і деякої еталонної моделі. Тому в разі, коли еталонною моделлю є наперед неповна і через це суб'єктивна модель знань конкретного вчителя, підсумкова оцінка в принципі не може бути визнана об'єктивною.

Мета статті: показати переваги тестів як засобу контролю знань, їх універсальність, наповненість, результативність і невелику затрату праці.

Проблеми забезпечення об'єктивності контролю знань повинні розглядатися щонайменше в шести аспектах:

1. Повинен враховуватись особистісний фактор: хто здійснює контроль знань, одноосібно чи колегіально проводиться ця процедура.

2. Необхідно враховувати форму реєстрації учня. З юридичної точки зору, якщо результати контролю знань не зафіксовані на матеріальному носії, то при незгоді учня з виставленою оцінкою предмет спору юридично відсутній і не може бути прийнятий до розгляду як об'єкт конфлікту.

3. Об'єктивність контролю знань залежить від умов та засобів його проведення (чи відомі учневі категорії оцінювання знань і часові рамки процедури контролю, чи є в нього навички фіксації своїх відповідей).

4. Об'єктивність контролю знань залежить від формулювання контрольних завдань. Причинами, що викликають незгоду учнів з виставленою їм оцінкою, часто називаються неоднозначні формулювання, що мають різні трактування і через це допускають суб'єктивну інтерпретацію ступеня істинності отриманих відповідей.

5. Особлива увага повинна звертатись на характер контрольних завдань. Перевірка формальної пам'яті, тобто перевірка знань учня, конкретних фактів і точних формулювань, навіть найважливіших правил і теорем, не може замінити вміння застосувати ці знання на практиці, а також аналізувати та інтерпретувати факти.

6. Об'єктивність контролю знань залежить від сили шкали підсумкового оцінювання (інакше називається бальністю) і застосованих критеріїв оцінювання. Чим більшу кількість градацій (підінтервалів) має шкала, тим більш точно можна оцінити знання учня. Водночас відсутність однозначних, педагогічно та юридично бездоганих критеріїв співставлення конкретній відповіді учня того чи іншого підінтервалу оцінюючої шкали спонукає до суб'єктивності індивідуального оцінювання знань і зумовлює часту незгоду учня з виставленою йому оцінкою.

Наведений перелік фактів не можна вважати повним, адже його можна продовжити. Проте усвідомлений облік їх впливів на відповідних стадіях педагогічного контролю знань дозволяє підняти об'єктивність цього процесу, зробити виставлені оцінки більш обґрунтованими і зменшити кількість конфліктів між суб'єктами та об'єктами цього процесу.

Враховуючи наведені інтуїтивні посилання і зупиняючись на методі експертних оцінок як ефективному інструменті врахування ступеня узгоджуваності думок експертів, сформуємо ав-

томатичний базис для розрахунку рівня об'єктивності педагогічного контролю знань. За аналогією з методом виведення формул для розрахунку ентропії інформаційного об'єкта сформулюємо умови, яким повинен відповідати показник рівня об'єктивності Q , і запропонуємо формулу, що задовільняє ці умови.

1. Значення Q мають розміщуватися в півінтервалі $[0; 1)$, причому $Q = 0$ буде означати повну суб'єктивність процедури контролю знань, $Q \approx 1$ — максимальну об'єктивність цієї процедури.

2. Якщо еталонна модель знань формується єдиним експертом, то $Q = 0$ незалежно від рівня його кваліфікації.

3. Якщо думки експертів щодо процедури контролю знань виявляються абсолютно різними (цьому відповідає значення коефіцієнта конкордації $W = 0$), то $Q = 0$.

4. Якщо $W \neq 0$ і кількість експертів $M > 1$, то значення Q повинно монотонно зростати при зростанні значення M , рівня кваліфікації кожного експерта (що характеризується коефіцієнтом γ_p , $\gamma_p = 1, M$) і коефіцієнта конкордації W .

5. Якщо рівень кваліфікації кожного експерта $\gamma_p \rightarrow 1$, а кількість експертів $M \rightarrow \infty$, то при $W \rightarrow 1$ значення Q також повинно прямувати до 1.

Перерахованим умовам відповідає формула

$$Q = W \frac{M-1}{M^2} \sum_{p=1}^M \gamma_p.$$

яку і пропонується використовувати як розрахунковий інструмент для оцінювання об'єктивності педагогічного контролю знань.

У наш час існує можливість перекласти роботу з накопичення, зберігання і аналізу результатів тесту «на плечі» комп'ютера. Робиться це за допомогою спеціальних програм, що називаються тестуючими оболонками (ТО). Розробка таких програм ведеться вже не один рік, тому існує чимало програм, які нібито можуть застосовуватись для тестування. Але не застосовуються! Навіть більше, одне із запитань, що задаються найчастіше на курсах користувачів — вчителів-предметників: «За допомогою чого проводити тестування?», а на курсах вчителів інформатики: «Як написати свою ТО?» Ця ситуація виникає внаслідок невідповідності вимог, що висуваються до ТО діагностами, і можливостей ТО, передбачених розробником.

Перша вимога, що пред'являється діагностом до ТО, стосується обсягу тестів, які обслуговуються. Тести, що обслуговуються ТО, мають складатися з необмеженої кількості запитань, кожне з яких повинно мати необмежене число варіантів відповідей, і будь-яка кількість варіантів відповіді може бути вірною. Звичайно, важко уявити собі тест із декількох сотень запитань або з десятками варіантів відповідей, але рішення про оптимальну кількість запитань має прийматися діагностом на основі методик розробки тесту, а не розробником ТО на онові технічних міркувань.

По-друге, очевидно, що програмка повинна володіти зручними візуальними засобами для створення, редагування тексту тестів, пошуку потрібного тесту. Діагност не повинен спеціально навчатись роботі з програмою, відповідно, інтерфейс як цих засобів, так і всієї програмки має бути розрахований на користувача, а не на програміста.

Кількість тестів, створених програмою, повинна бути необмеженою, причому всі вони мають зберігатися в одному файлі, бажано разом із програмою. Ця третя вимога трохи незвична для розробника ТО; в будь-якому випадку практично не зустрічається ТО з таким підходом до зберігання створених тестів. Найбільш розповсюджений нині варіант — це розміщення тесту в одному файлі, найчастіше у вигляді тексту із складною внутрішньою розміткою. Діагност має справу з необмеженою кількістю файлів, кожен із яких може розміщуватись в будь-якому місці, і для пошуку потрібного файла потрібно використовувати в принципі не призначені для цього засоби операційної системи. При створенні нового тесту діагност зустрічається зі складною внутрішньою розміткою, затрачуючи свої сили не на розробку свого тесту, а на оформлення його електронного варіанта. Той, хто тестується, шукає потрібний тест, не вибираючи його із списку, а використовуючи вікно пошуку файла, що дуже незручно. І на кінець можна отримати доступ до оригіналу тесту, наприклад, за допомогою програмки «Блокнот», і виявити вірні відповіді, аналізуючи текст програми. Така ситуація призводить до появи недовіри в діагностів до існуючих ТО.

Як було сказано вище, проблему легко подолати шляхом зберігання тестів в одному файлі-базі даних. Сучасні системи ксрування базами даних пропонують сильні засоби для створення зручного інтерфейса обробки і захисту даних, що зберігаються. Як це здійснити технічно? Перший і найбільш вагомий з них — створення ТО як програми. Дії, що виконуються ТО, — це зберігання, ввід, обробка даних. Другий, на жаль, практично непереборний стереотип для розробників ТО — створення для тесту окремого документа (в базі даних — таблиці). Для бази даних розміщення кожного тесту в окремій таблиці призводить до невизначеної кількості таблиць у базі даних, що недопустимо, хоча б з технічних причин. Правильне рішення — розміщення всіх тестів в скінченому числі зв'язаних таблиць.

Інформацію, що знаходиться в тестах, можна розділити на три категорії:

- 1) назва тесту і вся інформація, зв'язана з тестом загалом;
- 2) запитання тесту;
- 3) варіанти відповідей на запитання.

При цьому інформацію для кожної із названих категорій слід зберігати в окремій таблиці. Таким чином, ми зможемо всю інформацію про всі тести зберігати в трьох таблицях, де в першій буде зберігатись назва тесту і час, який відводиться на його проведення, друга міститиме запитання тестів і поле, що зв'язуватиме її з першою таблицею, а в третій будуть знаходитися варіанти відповідей, прапорець правильності відповіді і поле для зв'язку з другою таблицею. Крім вищеперерахованих полів, усі таблиці бази повинні мати ключове поле, що служитиме для ідентифікації запису таблиці і організації міжтабличних зв'язків. Усі значення цього поля унікальні.

Така структура зберігає дані тестів і задовольняє першу і третю вимоги. На «совісті» розробника лежить і виконання другої вимоги — грамотна візуалізація тестів, що зберігаються та редагуються. Створена структура даних зберігає інформацію тестів в випадковому порядку, але при роботі з тестом інформація може і повинна впорядкуватися програмно. Досягається це шляхом розробки форм для вводу.

Важливо зауважити також ще одну вимогу до ТО. Результати проходження тесту тими, хто тестується, та інформація про тих, хто тестується, має зберігатись у цій базі з усіма подробицями. ТО повинна мати засоби для перегляду й аналізу відповіді як однієї людини, що тестується, так і всієї групи (класу), а також засоби для різних вимірів, наприклад, визначення важкості та стандарту тесту. Причому, всі засоби роботи з тестами повинні володіти простим зрозумілим інтерфейсом. Також групове проходження тесту розуміється як застосування мережених технологій та одночасної роботи з тестом багатьох користувачів. При цьому база даних має бути захищена як від «дурня», так і від розумного користувача. Це можна досягнути шляхом грамотного розподілу доступу до бази даних та організації автоматизованих робочих місць з різним рівнем доступу.

Програми дистанційних курсів передбачають виконання на різних етапах навчання комп'ютерних тестів різного виду, які з точки зору функцій, що реалізуються, можна поділити на дві групи: контролюючі і навчаючі. Контролюючі тести реалізують функції самоконтролю і тренінгу і застосовуються після завершення дистанційним учнем вивчення певного розділу навчального курсу.

Можливі дві основні форми організації комп'ютерних тестів: тести з вибірконими відповідями і з відкритими відповідями.

У тестах з вибірконими відповідями тому, хто навчається, пропонується вибрати відповідь на поставлене запитання з представлених варіантів. При цьому забезпечується простий діалог з тим, хто тестується, оскільки йому достатньо натиснути клавішу з номером вибраної відповіді. Недоліком питання з вибірконими відповідями є наявність прихованої підказки на запитання: вибрати відповідь значно легше, ніж шукати її самостійно.

Тести з відкритими запитаннями передбачають, що той, хто тестується, буде вводити відповіді на запитання в текстову стрічку. Такий ввід потребує доброго знання клавіатури і вміння набирати текст і вносити в нього виправлення. Швидкість виконання такого тесту залежить від ступеня розвитку навичок роботи на клавіатурі.

Основними вимогами, що були поставлені до контролюючих програм, є такі:

- необхідність адекватного аналізу відповіді — різні помилки, розпізнавання правильної відповіді в будь-яких із еквівалентних форм його представлення;
- підтримка широкого спектра різноманітних методів оцінювання знань;

- фіксація результатів контролю, обслуговування та зберігання, роздрукування та статистичний аналіз.

Нині широко використовуються два методологічних підходи до конструювання педагогічних тестів: класична теорія тестування і латентно-структурний підхід.

Відповідно до класичної теорії тестування, тест відрізняється від збірки завдань системоутворюючими властивостями, які виявляються методами кореляційного і факторного аналізу. У цьому випадку кінцевий результат тестування прямо залежить від вибірки студентів загалом і класичні методи не завжди дають можливість адекватно оцінювати отримуванні результати тестування на початкових його етапах, коли вимоги до аудиторії можуть бути завищені або занижені.

Основними вимогами до забезпечення ефективності тесту є змістовність, надійність, узгодженість, дискримінативність, адекватність, обґрунтований вибір шкали оцінювання результатів тесту.

Тест вважається змістовним, якщо дозволяє оцінити те, для визначення чого призначений, ступінь засвоєння навчального матеріалу. За своїм змістовним наповненням тести оцінюють професійну підготовку і призначені для психологічного консультування. Валідність тесту забезпечується авторами навчального матеріалу.

Тест надійний, якщо він дає одні і ті ж показники для кожного учня.

Тест називається несуперечливим, якщо він є внутрішньо узгодженим.

Дискримінативність тесту характеризується можливістю окремих завдань і тесту загалом диференціюванням учнів щодо максимального, мінімального результатів тесту. Наприклад, тестове завдання, на яке дають правильну відповідь всі учні чи не дає відповіді ніхто, не мають ніякої практичної цінності.

Адекватність тесту означає відповідність чи невідповідність отриманих результатів реальним результатам учня.

У наш час існують три способи формування адаптивного тесту.

Перший — пірамідальний, коли без попередньої оцінки кожному учневі даються завдання середньої важкості і далі залежно від відповіді формуються наступні завдання, шкала трудності якої нижча або вища в два рази. При другому способі на початку використовується будь-який рівень складності, далі підбирається необхідний рівень для даного учня. При третьому, сертифікаційному способі завдання беруться з банку знань, де вони розділені за рівнями складності. При правильній відповіді наступне завдання береться з вищого рівня, при неправильній — з нижчого. Сучасна педагогічна наука визначає кількісну міру трудності завдань — логіт трудності — $\ln(q_j/p_j)$, де q_j — частка неправильних відповідей, p_j — частка правильних відповідей на завдання j з множини учнів, і міру знань — логіт знань — як $-\ln(p_i/q_i)$, де p_i — частка правильних відповідей на завдання, q_i — частка неправильних відповідей i -го учня на одні і ті ж завдання. Проте існують і альтернативні методи формування адаптивних тестів, наприклад, що базуються на результатах попереднього тестування учнів.

Одним із найважливіших питань, які виникають при використанні контролюючих програм, є те, як оцінювати отримані результати. Для визначення природності того чи іншого методу оцінювання дуже важливо сформулювати ціль тестування і залежно від цілі інтерпретувати результати тесту. Як правило, можна визначити такі цілі:

- диференціація учнів на декілька груп залежно від їх рівня знань;
- порівняння рівня знань з еталоном і визначення тих, хто відповідає цьому еталону;
- диференціація учнів на декілька груп на основі відповідності їх рівня знань еталону.

У випадку використання еталонів дуже важливо забезпечити валідність таких тестів і сформулювати еталони знань на основі обробки великого статистичного матеріалу. Залежно від цілі тестування необхідно використовувати як класичну теорію тестування для підготовки тестів і оцінки знань, так і можливості IRT.

При використанні класичної теорії тестування для адекватного виставлення оцінок розроблені три методи: за середнім відхиленням, з розпізнаванням за правилом Байеса і пороговий.

Метод оцінки за середнім відхиленням. Суть цього методу полягає в наступному: проводимо тестові випробування для N учнів із M завдань і будемо бінарну тестову матрицю; обраховуємо результат по кожному завданні x_j , знаходимо математичне очікування x_0 і середнє від-

хилення як корінь із дисперсії σ_x^2 ; викидаємо завдання, для яких результат досліджений $x_j > x_0 + \sqrt{\sigma_x^2}$ і $x_j < x_0 - \sqrt{\sigma_x^2}$, як занадто важкі або занадто легкі і що мають низьку диференційну можливість.

Обчислюємо кінцевий бал кожного учня u_i , значення математичного очікування кінцевого балу u_0 і середнє відхилення σ_x^2 .

Оцінюємо отримані результати, порівнюючи їх з середнім відхиленням і математичним відхиленням.

Розпізнавання за правилом Байєса. У більшості відомих з літератури методів виставлення оцінок розуміється, що розподіл результатів тестування підпорядковується нормальному закону. Накопичений статистичний матеріал дозволяє стверджувати, що на вибірках об'ємом 100–200 осіб, нормальний розподіл відсутній і задачу оцінки потрібно вирішувати як класичну задачу класифікації, наприклад за правилом Байєса. В цьому випадку виділяються чотири класи результатів, відповідно з оцінками від 2 до 5. По суті задача зводиться до відшукування трьох локальних мінімумів, і виділення класів будується від них.

Пороговий метод. При використанні цього методу оцінювання вимоги до рівня знань передбачають розбиття множини запитань на три чи більше множин:

- X1 — множина завдань, які повинен виконувати кожен учень в обов'язковому порядку, так званий поріг;
- X2 — множина завдань, які можуть бути виконані частково на цій множині і проходить диференціація учнів після виконання завдань з множини X1;
- X3 — множина запитань, які можуть бути не виконані взагалі, наприклад, підвищеної складності.

Ні в кого не викликає сумніву наявність об'єктивних причин для прискореного розвитку засобів і методів автоматизованого навчання і контролю знань. Найбільш важливими з цих причин є зростання інтересу до дистанційного навчання, становлення системи державного тестування випускників середніх шкіл, тенденція до збільшення і наповненості процесу поточного контролю знань на всіх рівнях навчання, необхідність значного зниження затрат праці на його організацію, проведення і підсумування результатів. Підвищення об'єктивності, результативності й ефективності проміжного та підсумкового оцінювання знань розглядається як одна з цілей реформи системи навчання, що проводиться в нашій країні.

Висновки. Тести повинні задовольняти вимогам: вони мають складатися з величезної кількості питань, кожне з яких повинно мати певні варіанти відповідей; програма має володіти зручними візуальними засобами для створення, редагування тексту тестів, пошуку потрібного тесту; результати проходження тесту тими, хто тестується, та інформація про тих, хто тестується, повинна зберігатися в базі з усіма подробицями; тестова оболонка повинна мати засоби для перегляду і аналізу відповіді як однієї людини, що тестується, так і всієї групи, а також засоби для різних вимірів.

ЛІТЕРАТУРА

1. Моисеев В. Б., Усманов В. В., Таранцева К. Р., Пятирублевый Л. Г. Статистический подход к принятию решений по результатам тестирования для тестов открытой формы // Дистанционное образование. — 2001. — №1.
2. Кузовлева К. Т. Конструирование педагогических тестов на основе современных математических моделей // www.informika.ru
3. Семенов В. В. и др. Развитие компьютерных технологий в дистанционном обучении // Новые информационные технологии в образовании: Обзор Информ. — М.: НИИВО, 1999.
4. Гусева А. И., Шпурик А. П. Дистанционная система тренинга и адаптивного тестирования ВИОТ-2 // Труды X международного научно-технического семинара «Современные технологии в задачах управления, автоматизации и обработки информации». — Алушта, 2001