

МІФ ЧЕТВЕРТИЙ

Лідія ВАЩЕНКО, кандидат педагогічних наук

Спирним, як на мене, є міф четвертий щодо використання статистичних методів у педагогічних дослідженнях.

Свою думку з цього приводу хочу висловити у вигляді тези, сформульованої ще в минулому столітті британським фізиком Вільямом Томсоном (лордом Кельвіном): «Якщо ви можете виміряти те, про що говорите, і виразити це в цифрах, значить ви щось про цей предмет знаєте. Але якщо ви не можете виразити це кількісно, ваші знання вкрай обмежені. Можливо, це початковий етап, але це не рівень справжнього наукового пізнання...». Водночас не можна не погодитись з автором статті про те, що зарозуміле подання статистичних даних щодо результатів дослідження (завжди правильних) ставлять під сумнів необхідність їх використання.

© Ващенко Л. С., 2016

На мою думку, в сучасній школі вкрай недостатньо уваги приділяється статистиці як науці про збирання, оброблення та аналіз кількісних даних, що характеризують масові явища та процеси. Елементи статистики використовуються під час вивчення шкільного курсу географії, економіки, загальної біології в старшій школі тощо, але цього замало для нормального життя в сучасному світі.

Результати міжнародного порівняльного дослідження TIMSS, свідчать про те, що у наших школярів виникли труднощі щодо аналізу статистичного розподілу даних, представлених графічним зображенням варіаційних рядів у вигляді діаграм, графіків, гістограм тощо. Школярі не змогли використати інформацію, подану у графіках, таблицях, діаграмах; виконати завдання, проілюстровані малюнками чи схемами, завдання, що їх включено до розділу «Робота з даними» тощо.

ЯКБИ МІФИ «ПРАЦЮВАЛИ»

Олександр КОЗЛЕНКО, науковий співробітник відділу біологічної, хімічної і фізичної освіти НАПН України

У цьому відгуку хотілося б зосередитися не на перших двох високорівневих міфах (про формування системи знань та педагогічні технології), а на двох прикладних міфах – про можливість визначення рівня складності завдань (принагідно подякувавши авторові за те, що він залишив за тестуванням статус «технології») і про статистику. Це дуже щільно пов'язані один з одним міфи, або, як сказав би представник однієї з міфічних професій, «закорельовані один на одній». Але про все за порядком.

У будь-якій людській діяльності перетинаються певним чином мистецтво, наука й технологія. Навіть художник, чия діяльність видається майже повністю мистецтвом, має попередньо пройти курс засвоєння техніки малюнка, використання різних типів фарб (технологія), виготовлення яких, а також способи збереження створеного зображення є плодами науки. Так само в медицині: фізіологія чи фармакологія – науки, що відповідають усім критеріям наукового способу пізнання (експеримент, статистична обробка результатів тощо), хірургія чи стоматологія – суто технологічні галузі, в той час як терапію

© Козленко О. Г., 2016

(і особливо діагностування захворювання) лікарі наполягають вважати мистецтвом (і жодні протоколи ВООЗ їм не указ). Втім саме медицині людство має бути вдячним за створення **доказової медицини**, або **медицини, заснованої на доказах**.

Термін *Evidence-Based Medicine* (медицина, заснована на доказах) уперше був запропонований групою канадських учених з університету Мак Майстер в 1990 р. без чіткого визначення, але швидко прижився і поширився в англomовному науковому середовищі. Медицина, заснована на доказах, – це сукупність методологічних підходів до проведення клінічних досліджень, оцінювання і застосування їхніх результатів; передбачає пошук, порівняння, узагальнення і поширення здобутих доказів для використання в інтересах хворих. Основою методології клінічних випробувань, за ідеологією доказової медицини, є стандартизація умов (дослідна і контрольна групи різняться тільки одним параметром – обраним лікарським препаратом), рандомізація (забезпечення випадковості вибірки пацієнтів, які беруть участь в експерименті, і відсутність відмінностей між контрольною та експеримен-

тальною групами) і подвійне осліплення досліджень (пацієнт не знає, отримує він досліджуваний препарат чи плацебо; лікар, що проводить дослідження, не знає, хто з пацієнтів отримує досліджуваний препарат, а хто – плацебо). У поєднанні з великою вибіркою (йдеться про тисячі і десятки тисяч пацієнтів) у такому дослідженні можуть бути отримані однозначні висновки про те, чи допомагає даний лікарський препарат краще, ніж плацебо, чи ні. За їх результатами лікар у будь-якій лікарні може запропонувати пацієнтові саме ті ліки, що лікують дану хворобу, а не якісь інші.

На простому прикладі організація подвійного сліпого дослідження описана в чудовій популярній статті В. Башмакової і А. Доброчаєва «Чи пахне морквяний відвар?» у журналі «Квантик» [1]. У ній головному героєві, який терпіти не може моркву в супі, в ході експерименту належить дізнатися, чи справді він відчуває запах моркви.

Звісно, педагогіка – це значною мірою мистецтво, менше – технологія і ще менше – наука. Чи можлива взагалі «доказова педагогіка»? В описаному вище сенсі – швидше за все, ні. Почнемо з того, що організація дослідної та контрольної груп у педагогіці є істотно складнішою за біологічні та медичні експерименти. Наприклад, дослідник винайшов геніальну педагогічну технологію спільної експертної діяльності в групах під час вирішення складних завдань. Він хоче провести експеримент в 11 класах. Щоб довести її ефективність, досліднику мало залучити будь-яких інших учнів 11 класів, вони повинні бути ще й близькими за рівнем підготовки і викладання предмета (рівень стандарту, академічний або профільний рівень), з ними потрібно також організувати групову роботу (а що як чинить вплив не педтехнологія, а просто робота в групах?), але вирішувати ті самі завдання (а що як вся справа – в експериментальних завданнях?). Почули: «Кладц»? Це пастка зачинилася.

Квінтесенцією педагогічної специфіки є поняття **формуального експерименту**, без якого неможливе жодне педагогічне дослідження. «Для методу формуального експерименту характерно активне втручання дослідника в досліджувані ним психічні процеси. ...Проведення формуального експерименту передбачає проектування та моделювання змісту формованих психічних новоутворень, психолого-педагогічних засобів та шляхів їх формування. Під час дослідження шляхів реалізації такого проекту (моделі) в процесі навчально-пізнавальної роботи з дітьми разом з тим можна вивчати умови й закономірності походження, генезис відповідного психіч-

ного новоутворення» [2, 165]. Іншими словами, пропонується взяти участь в експерименті, методика проведення якого відбивається на результатах експерименту. Це *рандомізоване подвійне сліпе дослідження* строго навпаки.

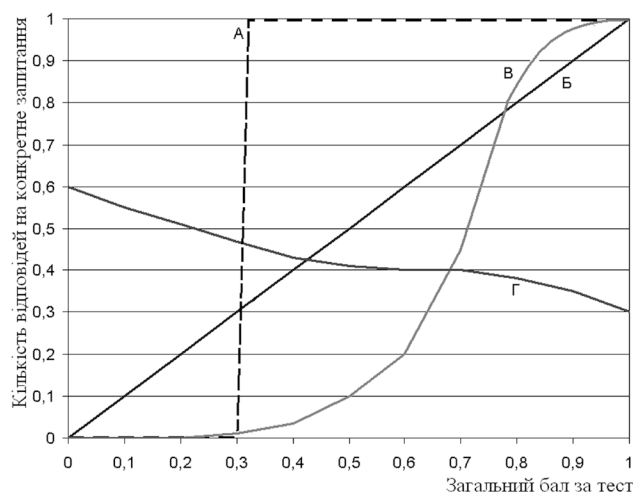
Чи означає сказане вище, що застосування статистичних методів у педагогіці неможливе? Аж ніяк ні. Якщо ставитися до них не як до «красивих гудзиків на костюмі» дисертанта, а саме як до робочого інструменту.

Автор цілком слушно зазначає, що «в переважній більшості вчителів думки різняться не лише стосовно «присвоєння» завданню балів, а й визначення належності завдань до певного рівня, особливо якщо завдання розміщується на межі рівнів, тобто на 3 – 4, 6 – 7, 9 – 10 балів». Це не дивно – складність завдання є інтегральною характеристикою, що, за визначенням О. І. Башмакова та І. О. Башмакова [3, 310], «пропорціональна обсягу знань та інтелектуальним зусиллям, що потрібні для його розв'язування». Автори наводять такий перелік чинників, що впливають на складність комп'ютерного навчально-тренувального завдання (НТЗ):

- «наближеність результату (відповіді) або даних, що потрібні для його отримання, до змісту навчального матеріалу»;
- повнота і характер представлення в основі завдання інформації, необхідної для її виконання (подібні відомості виражаються у вигляді рекомендацій, указівок, порад, посилань, навідних запитань тощо);
- обсяг інформації, необхідної для виконання НТЗ, її розподіл за змістом курсу, а також складність відтворення її по пам'яті;
- складність методу розв'язування (кількість етапів, характер переходів між ними, кількість винятків із правил тощо);
- склад і характер відносин, що пов'язують положення навчального матеріалу, які перевіряються, з тими, що їх треба врахувати під час виконання НТЗ;
- форма подання основи завдання і результату (з погляду їх наочності);
- характер діалогу, включаючи взаємодію з моделями.

Але існує математичний апарат, призначений саме для цього: як серед певної кількості завдань виокремити такі, що краще диференціюють учнів за рівнями (різняються за дискримінантною здатністю). Цей апарат був розроблений ще в 60 – 80-ті роки ХХ ст. під назвою теорії педагогічних вимірювань (у першоджерелі – *Item Response Theory, IRT*); її активно використовують такі провідні тестологи, як В. С. Аванесов [4, 5], Д. А. Новіков [6], В. С. Кім [7] та інші.

Розглянемо мал. 1. Графік показує залежність кількості балів, набраних за конкретне завдання (вісь y) від загальної оцінки за тест (вісь x); обидві величини виражено в частках від одиниці. Ідеальне завдання – це таке, що уможливило б чітке розмежування тих, кому під силу впоратися з ним, і тих, хто не в змозі (А): всі учні з балом за весь тест менш ніж 0,3 завдання не виконали, всі з балом 0,3 і більше – виконали. Вочевидь, що так у реальних тестах не буває: хтось знав, але випадково відзначив не ту відповідь, хтось вгадав. Варіант Б, коли ймовірність відповіді на запитання поступово збільшується, здається реальнішою, але таке запитання є некоректним з погляду мети педагогічного вимірювання: адже на виході потрібно отримати *відмітку*, що розділяє учнів на класи, а запитання Б для цього непридатне. Запитання В вирізняється кращою дискримінантною здатністю: частка учнів із загальним балом за тест менш ніж 0,6, які відповіли на нього правильно, не перевищила 20 %, а серед «сильних» учнів, які впоралися з тестом – на рівні 0,8 – 1,0 – за 90 %. Навпаки, запитання Г складено так, що сильні учні відповідають на нього гірше, ніж слабкі (ті, що набрали менший загальний бал); таке завдання треба відбракувати (але знайти його можна тільки за допомогою *IRT*-статистики та після опрацювання даних з певної вибірки).



Мал. 1. Дискримінантна здатність завдань за *IRT*-статистикою

Що цікаво, за допомогою цього методу можна не лише з'ясувати, які із завдань мають більшу дискримінантну здатність і є цінними для подальшого використання в педагогічній практиці, а й скоригувати «сирі» бали, що їх отримали учні за тест, і отримати рейтингові бали, які точніше віддзеркалюють рівень підготовленості учнів, аніж проста арифметична сума «сирих» балів.

Звісно, пересічного вчителя можуть налякати формули для розрахунків дискримінантної здатності, валідності тестових завдань, кореляцій

завдань між собою або обчислення надійності тесту в цілому:

$$r_{xy} = \frac{\sum_{i=1}^n [x_i - \bar{x}](y_i - \bar{y})}{(n-1)\sigma_x\sigma_y} \quad r_i = \frac{r}{r-1} \left(\frac{\sigma_x^2 - \sum pq}{\sigma_x^2} \right)$$

Формули обчислення надійності тестів:

Пірсона і К'юдера – Річардсона

Але за наявності *Microsoft Excel* (з пакетом аналізу даних) обчислення всіх цих показників, після того як один раз створено або запозичено шаблон, розпочинається і закінчується вставленням таблиці з правильними (1) і неправильними (0) відповідями кожного учня на кожне із завдань тесту. Всі показники процесор електронних таблиць обчислить сам. Крім того, для цієї роботи непотрібна велика за обсягом вибірка: найсуттєвіші тенденції можна наочно продемонструвати навіть на малих вибірках.

Чи можуть результати розрахунків статистичних показників тестів сказати тому, хто їх склав і використав, щось таке, чого він сам не може передбачити? Так, але й не зовсім так. Якщо ми перевіряємо розподіл завдань і взаємозв'язки між відповідями на пари (кореляції) і бачимо, що два завдання закорельовано одне на одне майже стовідсотково – чи не побачили б цього, якщо уважно переглядалися самі завдання? Дискримінантна здатність, обчислена за допомогою *IRT*-статистики, «змушує» уважно подивитися на самі завдання: а що в них погано (або навпаки – добре) відпрацьовано? Яких помилок припустилися під час складання завдань? Чого не врахували, на що не звернули уваги?

Наведемо такий показовий приклад. У ході дослідження навчально-виховного процесу з біології, що їх проводить відділ біологічної, хімічної та фізичної освіти Інституту педагогіки НАПН України, особливе значення мають створення й апробування авторських експериментальних тестових матеріалів (тих самих, в яких рівень складності «легко піддати критиці»). Зазвичай такі добірки завдань складаються в кількох варіантах (щонайменше двох), і, за замовчуванням, вважається, що варіанти є однаковими не тільки за структурою і формою завдань, а й за складністю. Ми перевірили цю гіпотезу для тесту з біології для 11 класу (добірка завдань до тем «Закономірності спадковості», «Закономірності мінливості» та «Генотип як цілісна система») за допомогою статистичних методів *IRT*.

Обчислювали два показники: рівень складності завдання та дискримінантну здатність завдання. Дослідження підтвердило рівноцінність дискримінантної здатності більшості завдань різних варіантів (або високий рівень показника для обох варіантів, або низький – також для обох варіантів), що свідчило про збалансованість варіантів

завдань. Показовим був випадок нерівноцінності одного із завдань (першого завдання для обох варіантів). За формою завдання близькі: необхідно за фотографією обрати номер хромосоми, за якою спостерігається моносомія (I варіант) або трисомія (II варіант). Статистична обробка результатів виявила значну різницю в дискримінантній здатності (висока для I варіанта і відсутня для II варіанта). Пояснити таку різницю можна лише виходячи з неочевидності для учнів того факту, що наявність у каріотипі чоловіка однієї X-хромосоми та однієї Y-хромосоми не є ознаками моносомії. Таким чином, статистична обробка дає змогу дослідникові (або вчителю) краще зрозуміти складні для учнів елементи змісту за допомогою відносно простого обчислювання.

Ще одна цитата зі статті, що її можна використати як епіграф для подальшого обговорення міфу про різнорівневі завдання: «Неприпустимо таврувати учня, підкреслювати його низькі досягнення, пропонуючи найлегшу контрольну роботу».

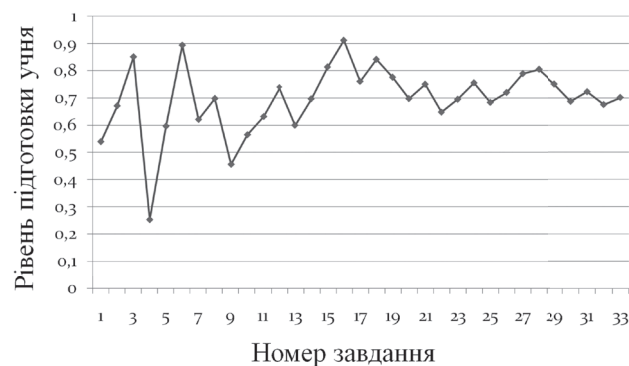
З цим складно не погодитися. Але це не така проблема, що неможливо розв'язати. І зробити це можна як на рівні самого вчителя, так і на рівні дослідників, які створюють тестові бази та програми для тестування.

Одним з варіантів учительського рівня може бути така форма контролю знань, як «Рейтингова самостійна робота, що самоіндивідуалізується» [8]. Під час підготовки до неї учитель добирає 7 – 10 відкритих завдань, запитань чи задач з теми таким чином, щоб щонайменше декілька завдань вимагали нестандартного підходу до розв'язування і давали змогу виявити себе учням з «правонапівкульним» дивергентним мисленням. Кожному із завдань присвоюється свій «максимальний бал». У ході самої роботи вчитель пропонує учням завдання (написані на дошці, роздруковані, виведені на екрані через проектор) без уточнення, які з них дають більший максимальний бал, а які – менший, але орієнтує учнів не «зацікловатися» на окремих завданнях й не «ковзати» по них, уникати поверхових відповідей. Учні **самостійно** обирають і виконують завдання із числа запропонованих. Для отримання високої оцінки не обов'язково треба виконати всі завдання, оскільки високоякісне виконання вже трьох-чотирьох з них дасть змогу набрати потрібний рейтинговий бал. При перевірці відповідь учня на кожне завдання оцінюють певною кількістю балів, аж до максимального. Підсумковий рейтинговий бал складається із суми «сирих» балів і додаткових балів за якість, пропорційних відсотку від максимальної кількості балів за ті питання, на які надано відповіді. За отриманими рейтинговими балами здійснюється ранжування учнів, з якого можна перерахувати рейтингові бали в 12-бальну шкалу.

Інший підхід пов'язаний з використанням так званих адаптивних тестів, обговорення яких ми почнемо з відомої інтернет-гри «Акінатор». Вона реалізована 12-ма різними мовами (української, на жаль, немає, але є польська – <http://pl.akinator.com/> та російська – <http://ru.akinator.com/> версії). У цій грі ви загадуєте персонажа (будь-якого: реального або героя книжки чи фільму, людину або тварину і т. п.), а джин з комп'ютера ставить вам 20 запитань, на які ви відповідаєте «Так» або «Ні» (з варіантами «Можливо, частково» або «Не зовсім»). Після серії запитань джин угадує (або не вгадує), кого ви задумали. Зробіть паузу, зайдіть в Інтернет і пограйте з джином, а після цього повертайтеся до статті.

Тепер, коли ви пересвідчилися у можливостях адаптивних тестів (покладіть руку на серце чи інший орган, що відповідає за педагогічну адекватність, і скажіть, що складніше: вгадати персонажа серед майже нескінченної множини кандидатів чи визначити рівень знань учнів з конкретної обмеженої теми?), продовжимо.

Адаптивні тести знань (хоча аналогічно можуть вимірюватися й інші показники разом з рівнем сформованості певної предметної компетентності) побудовано так: комп'ютерна тестова програма пропонує учневі серію завдань з бази. Кожному із завдань відповідає певний рівень складності, визначений експериментально на великій вибірці. Завдання пропонуються таким чином, щоб наступне було складнішим за попереднє, правильно виконаним, але простішим за неправильно виконане (мал. 2). Так само, як в «Акінаторі», перші запитання слугують для грубого визначення рівня підготовки (в «Акінаторі» – визначенню основних властивостей персонажа: людина це чи тварина, та взагалі чи жива це істота, реальна або вигадана), а подальші запитання уможливають коригування точності вимірювання.



Мал. 2. Графік типової сесії САТ – комп'ютерного адаптивного тестування. Кадр із презентації Іване Міндадзе «Комп'ютерне адаптивне тестування в Грузії» на V Міжнародній науково-методичній конференції «Освітні вимірювання-2015. Реформування зовнішнього незалежного оцінювання: методологія, модель, основні складові», 30 вересня – 2 жовтня 2015 р., м. Одеса

Зроблений вище висновок про недосяжність «доказової педагогіки», можливо, вже скоро виявиться неправильним. Використання методів обробки великих масивів даних (*BigData*) під час методологічно коректно організованих масових досліджень, можливо, дає змогу з'ясувати, від яких чинників залежить успішність колишнього учня, рівень його добробуту і рівень задоволеності своїм життям. Можливо, відповіді виявляться простішими, ніж ми думаємо. Ф. Ніцше писав: «Одвічні запитання ходять по вулицях». Одвічні відповіді також.

ЛІТЕРАТУРА

1. Башмакова В., Доброчаев А. Пахнет ли морковный отвар? // Квантик. – 2014. – № 9. – С. 2 – 7; режим доступу до електронної версії: <http://kvantik.com/files/2014-09.pdf>
2. Давыдов В. В. Проблемы развивающего обучения : Опыт теорет. и эксперимент. психолог. исслед. /

В. В. Давыдов. – М. : Педагогика, 1986. – 240 с. – (Труды дейст. чл. и чл.-кор. АПН СССР).

3. Башмаков А. И., Башмаков И. А. Разработка компьютерных учебников и компьютерных обучающих систем, – М. : Филинь, 2002, – 616 с.

4. Аванесов В. С. Научные проблемы тестового контроля знаний. – М., 1994. – 135 с. Сайт Аванесова В. С. – <http://www.testolog.narod.ru>.

5. Аванесов В. С. Тесты: теория и методика их разработки // Приложение к газете «Первое сентября». – 2001. – № 32.

6. Новиков Д. А. Статистические методы в педагогических исследованиях : типовые случаи. – М. : МЗ-Пресс, 2004. – 67 с.

7. Ким В. С. Анализ результатов тестирования в процессе *Rasch measurement* // Пед. измерения. – 2005. – № 4. – С. 39 – 45.

8. Козленко А. Г. Рейтинговая самоиндивидуализирующаяся самостоятельная работа. – Режим доступа: <http://www.kozlenkoa.narod.ru/rating1.htm>.

ВАЛІДНІСТЬ, ТЕХНОЛОГІЯ, КОМПЕТЕНТНІСТЬ

Володимир СІПІЙ, молодший науковий співробітник, Інститут педагогіки НАПН України

Система освіти України в останні десятиліття перебуває у стані неперервного реформування. То що ж ми бачимо в шкільній практиці: «педагогічні міфи» чи все ж неповною мірою реалізовані реформи системи освіти? Науково обґрунтовані підходи до реформи освіти розробляє МОН України у співпраці з НАПН України. Але основною дієвою особою, яка впроваджує Державні стандарти, концепції, навчальні програми, є вчитель. Усі «педагогічні міфи» сформувались унаслідок браку інформації й практичної допомоги вчителів з боку держави.

Надалі розглянемо деякі педагогічні міфи, що запропоновано для обговорення в статті Л. Величко.

Про валідність педагогічних досліджень. Погоджуючись з тим, що недооцінено роль якісного аналізу результатів дослідження. Але проблема полягає не в самій лише математичній обробці здобутих даних. Розроблені методики навчання того чи іншого предмета апробовуються на обмеженій кількості навчальних закладів, причому тих, де вчитель готовий до інновацій. Дисертант проводить констатувальний, форму-

вальний експерименти, враховуючи специфіку навчального закладу, контингент учнів, проводиться науковий супровід діяльності педагога, що й зумовлює наявність позитивних результатів. Запровадження ж цієї самої методики без належного супроводу з боку методичних служб та допомоги в подоланні ускладнень під час упровадження методики, нехтування психолого-педагогічними особливостями учнів, невідповідність учителя до роботи за цією методикою матиме негативні результати від її впровадження.

Про технологізацію освіти. Пріоритетом розвитку освіти є впровадження сучасних інформаційних технологій, що забезпечують доступ до мережі високоякісних баз даних, розширюють можливості комунікації учителя й учня. Передусім слід зазначити, що парк комп'ютерної техніки більшості шкіл є застарілим й ДСНІП обмежують час роботи учня з технікою. А мультимедійні засоби навчання не можуть замінити традиційних ТЗН й зрештою крейди з дошкою, учнівського зошита з кульковою ручкою.

Що ж таке «технологічний підхід у навчанні»? Хибно вважати, що використання ІКТ для навчання це вже й є технологізацією освіти. Педагогічна технологія – спеціальний набір форм,