

МОДЕЛІ І МЕТОДИ СТАТИСТИЧНОЇ ОБРОБКИ РЕЗУЛЬТАТІВ ДИСТАНЦІЙНОГО ДІАГНОСТУВАННЯ РІВНЯ ЗНАНЬ

Ю. В. Безгачнюк

*Івано-Франківський національний технічний університет;
Україна, 76019, м. Івано-Франківськ, вул. Карпатська, 15;
тел. +380 (342) 4-80-00; e-mail: yurkovskiy@au.nung.if.ua*

Розглядаються моделі і методи статистичної обробки результатів дистанційного діагностування рівня знань студентів, а також реалізація відповідного апарату в системі d-tester, яка використовується на кафедрі КТіСУ ІФНТУНГ впродовж п'яти років

***Ключові слова:** навчальний процес, дистанційне діагностування рівня знань студентів, статистична обробка результатів досліджень.*

Вступ

Історія педагогіки налічує достатню кількість форм і методів оцінки знань, вмінь та навичок тих, хто навчається (далі по тексту – студенти), частина яких прийшла до нас з глибокої старовини і з середніх століть. Серед них виділяються іспити, що стали традиційною формою контролю знань.

Іспит – це випробування, що проводиться відповідно до встановлених правил; іспит відображає ідею перевірки, випробування, при успішному результаті яких підтверджуються певні права особи. У кожній країні світу іспити є свідомством офіційного схвалення результатів закінчення навчального закладу, вони є незамінними для мотивації до серйозної і постійної роботи в процесі навчання [1].

В якості альтернативного методу педагогічних вимірювань слід розглядати використання тестів, при цьому ставиться питання про створення докімології (від грец. dokime) – науки про організацію та реалізацію контролю знань.

Сьогодні відносно тестів затвердилась норма рефлексії обов'язкової перевірки їх якості, найбільш суттєвої вимоги до даного методу педагогічного контролю.

Можна виділити три головні ознаки якості тестового контролю якості знань:

1. Якість – це визначеність.
2. Визначеність повинна бути тотожна з буттям, невід'ємною від існування об'єкта тестування.
3. Визначеність повинна бути зовнішньою, вона є границею, що виділяє певний об'єкт серед інших, породжує його своєрідність, специфіку, індивідуальність.

Якість тестів, з моменту їх появи, стала об'єктом серйозних наукових досліджень. Ще до початку ХХ століття з'явилися перші вказівки на необхідність підвищення точності тестових оцінок і їх адекватності. З часом якість тестів стала основним об'єктом загальної теорії тестів. У цій теорії виділяються два основні критерії якості тестів – надійність і валідність.

На додаток до двох цих відомих критеріїв якості тестів виникла необхідність у дослідженні третього критерію – ефективності тестів. Існує математико-статистичний апарат, який дозволяє кількісно оцінити рівень ефективності як тесту в цілому, так і кожного завдання окремо.

Ефективність тестового контролю є категорією, ширшою за надійність. Надійність є необхідною умовою забезпечення заданого рівня ефективності тесту. Ефективність функціонування складної системи тестування повинна оцінюватися фахівцем, що володіє системою в цілому, який знає її призначення і вимоги до неї.

За кордоном у ХХ столітті гостро відчувалася необхідність у створенні методів оцінки рівня складності кожного завдання тесту, які не залежать від складу оцінюваної групи. Зрозуміло, що якщо на певне завдання j відповідає частина добре підготовлених студентів, то частка правильних відповідей на це завдання (p_j) буде значно вищою, ніж у групі слабо підготовлених.

Багаторічні спроби пошуку відповідей на питання про визначення рівня складності завдань систем тестового контролю призвели до двох наступних результатів:

1. Рекомендації щодо збільшення вибірки випробовуваних до тих пір, поки ймовірність p_j не стане достатньо близькою до значення параметра складності на генеральній сукупності. Метод виявився тривіальним і марнотратним, його важко визнати ефективним для вирішення поставленого завдання.

2. Другий результат, отриманий в 1943 р. Д. Н. Лавли, був спробою отримати показник складності завдання тестового контролю, безвідносно до рівня конкретної групи студентів [2]. Як виявилось, інваріантною властивістю володіють, так звані, характеристичні криві тестових завдань і тесту загалом. Побудувавши за емпіричними даними криві, де на осі абсцис відкладаються тестові бали, а вздовж осі ординат – ймовірність або відсоток правильних відповідей, можливо виявити, що для студентів групи слабкої підготовки “працює” нижня частина характеристичної кривої, а для студентів групи сильної підготовки – верхня. Таким чином, для оцінки вірогідності правильної відповіді студента залежно від рівня його підготовленості (q_i) почала використовуватися функція логістичного вигляду, з фіксованим рівнем складності завдання. Дану ідею пошуку оцінки ефективності складності завдання тесту пізніше розвинув і довів до комп'ютерної технології Ф. М. Лорд і його колеги.

Ефективний тест не може складатися з неефективних завдань. Тому необхідно поставити питання про ознаки, які відрізняють ефективне завдання тестового контролю. Ефективне завдання перевіряє важливий

елемент змісту учбової дисципліни, який нерідко називають ключовим для необхідної структури знань студентів. У тест, відповідно, включаються тільки такі завдання, які експерти визнають як ключові елементи навчальної дисципліни.

Спеціалістами в області тестового контролю вже давно досліджується один з показників якості завдання – так званий індекс диференціюючої (дискримінанта) здатності завдання (ДЗЗ). У новій теорії тестів (Item Response Theory) з поняттям ДЗЗ асоціюється значення параметра крутизни характеристичної кривої тестового завдання, що дозволяє істотно уточнити і збагатити поняття ДЗЗ. Таким чином, відкривається можливість введення в науковий обіг таких понять, як компоненти ефективності, локальна ефективність тесту, локальна ефективність завдання, порівняльна ефективність двох і більшої кількості завдань, з'являється можливість конструювання ефективного тесту за рахунок відбору ефективних завдань. Математичним аналогом значення інформативності, а отже, ефективності кожного завдання, на відповідній ділянці складності є поняття “Інформаційна функція”.

Це поняття і відповідний математичний апарат в науковий обіг для оцінки диференційованої точності педагогічних вимірювань введені Г. Рашем та згодом А. Бернбаумом [3]. Згідно даної теорії, тест не може бути ефективним на всьому діапазоні підготовленості студентів. Він може бути ефективним на одному діапазоні рівня знань студента і менш ефективним на іншому.

1. Теоретичні основи статистичної обробки результатів діагностування рівня знань студентів

Актуальним науковим і педагогічним завданням є проведення контролю знань студентів дистанційно з використанням комп'ютерної техніки і відповідного програмного забезпечення. Тестування є невід'ємною частиною дистанційного навчання, оскільки являє собою зворотній зв'язок в колі “викладач – студент”. При цьому в очному тестуванні викладачем студента виникають проблеми, пов'язані із суб'єктивністю оцінок викладачів, неможливістю протестувати одним викладачем значний потік студентів тощо. У зв'язку з цим в рамках розвитку інформаційних технологій існує актуальна проблема автоматизації процесу тестування – створення інформаційних систем тестування знань, які б не тільки забезпечували значну економію часу викладача і студента, але й дозволяли швидко і об'єктивно оцінити знання студента, а також дозволяли б студенту проводити самоконтроль своїх знань з метою усунення “прогалин” у них [4].

Для забезпечення якісного аналізу результатів проходження тестового контролю, а також визначення якості (ефективності) тестових завдань необхідно використовувати методи статистичної обробки результатів.

На етапі статистичної обробки та інтерпретації результатів дистанційної діагностики (ДД) слід застосовувати результати Item Response Theory, що перекладається на українську мову, як математич-

но-статистична теорія оцінки латентних параметрів завдань тесту та рівня підготовленості студентів. На цьому етапі уточнюється зміст змінної, яка цікавить дослідника, визначається шкала вимірювання, застосовується комп'ютерна обробка результатів тестування.

Латентними називаються позитивні та негативні якості особистості, які неможливо безпосередньо виміряти – рівень підготовки студентів, знання учбової дисципліни, здатність сприймати інформацію, інтелектуальний розвиток особистості та багато інших чинників. Спроби виміряти дані якісні показники на рівні буденної свідомості закінчуються словесними або чисельними оцінками, що містять значні похибки.

На відміну від класичної теорії тестів для IRT характерно спрямування до фундаментального теоретичного підходу і одночасно до коректного вирішення низки практичних задач педагогічного тестування. В практичному плані це спрямування пов'язано з деякими труднощами, які, до речі, не завжди усвідомлюються відомими тестологами – створювачами сучасної теорії тестування. Необхідно застосовувати доволі складний математико-статистичний апарат, використовувати потужну комп'ютерну техніку, необхідно розробляти спеціалізоване програмне забезпечення.

Очевидно, що логіка повинна бути така: латентні параметри, вірніше взаємодія двох множин їх значень, породжує спостережувані результати виконання тесту. Елементи першої множини – це значення латентного параметра, який визначає рівні знань N студентів Θ_i , де $i = 1, 2, \dots, N$. Друга множина утворюють значення латентного параметра B_j , $j = 1, 2, \dots, n$, рівні складності n завдань тесту. Однак на практиці формулюється зворотня задача: за відповідями студентів на завдання тесту оцінити значення латентних параметрів Θ і B . Для її вирішення необхідно відповісти щонайменше на два питання.

Перший пов'язаний з вибором співвідношення між латентними параметрами Θ і B . Датський математик Георг Раш (George Rasch) запропонував ввести це співвідношення у вигляді різниці $\Theta - B$, вважаючи, що параметри Θ і B оцінюються за однією і тією ж шкалою.

Значення параметра Θ_i можна розглядати як положення i -го студента, а значення B_j – як положення j -го завдання на одній і тій же осі змінних Θ , B . У такому разі ідея введення різниці параметрів одержує цікаву геометричну інтерпретацію. Абсолютна величина різниці $|\Theta_i - B_j|$ – це відстань, на якій знаходиться студент з рівнем знань Θ_i від завдання складністю B_j . Якщо ця різниця велика за модулем і негативна, то завдання даремне для вимірювання рівня знань i -го студента. Студент, напевно, не зможе виконати його правильно. З іншого боку, великі позитивні значення цієї різниці теж не ефективні ані для процесу контролю, ані для навчання i -го студента. Завдання такої складності давно ним засвоєне. З погляду підходу, пропонуваного в IRT, такі завдання неефективні для оцінювання даного значення Θ .

Відповідь на друге питання, яке є центральним в IRT, пов'язане із вибором математичної моделі. Слідуючи за основним припущенням IRT, можна стверджувати, що існує деяка математична модель взаємозв'язку між емпіричними результатами тестування і значеннями латентних змінних Θ і B .

Під час вибору моделі слід враховувати, що в реальних умовах на спостережувані результати здійснюють вплив як випадкові, так і не випадкові чинники. Не зважаючи на всю “випадковість” окремих результатів тестування, виявляється відносна інваріантність значень латентних змінних від конкретного випробування або від ряду випробувань. Наприклад, певна стійкість частот появ значень змінних $\Theta_1, \Theta_2, \dots, \Theta_N$ спостерігається при багатократному тестуванні групи N студентів паралельними тестами. Ця стійкість є підставою для використання поняття вірогідності події як міри можливості його появи. За таку подію зазвичай вибирається правильна відповідь i -го студента на j -е завдання тесту. Умовну ймовірність правильного виконання студентами завдань тесту виражають за допомогою тієї або іншої математичної моделі.

Можна розглядати умовну ймовірність правильного виконання i -м студентом з рівнем знань Θ_i різних за складністю завдань тесту, вважаючи Θ_i параметром i -го студента, а β — незалежною змінною. В цьому випадку умовна ймовірність буде функцією латентної змінної β :

$$P_i \{x_{ij} = 1 | \Theta_i\} = f(\Theta_i - \beta), i = 1, 2, \dots, N. \quad (1)$$

Аналогічно вводиться умовна ймовірність правильного виконання j -го завдання, складністю β_j будь-якими студентами групи. Тут незалежною змінною є Θ , а β_j — параметр, який визначає складність j -го завдання тесту:

$$P_j \{x_{ij} = 1 | \beta_j\} = f(\Theta - \beta_j), j = 1, 2, \dots, n, \quad (2)$$

де

$$x_{ij} = \begin{cases} 1, & \text{якщо } \Theta_i \geq \beta_j \\ 0, & \text{якщо } \Theta_i < \beta_j \end{cases}$$

N — число студентів; n — кількість завдань в тесті.

Однопараметрична модель Г. Раша

Однопараметрична модель, яку часто називають простою логістичною моделлю, є однією із сімейства логістичних кривих, описаних Г. Рашем в статті "On general laws and the meaning of measurement in psychology". Аналітичне задання моделі представлено формулами (3) і (4).

$$P_j(\theta) = \{1 + \exp[-1.7(\theta - \beta_j)]\}^{-1}; \quad (3)$$

$$P_i(\beta) = \{1 + \exp[-1.7(\theta_i - \beta)]\}^{-1}. \quad (4)$$

Підхід Г. Раша до побудови моделі вимірювання та інтерпретації результатів тестування (*Rasch Measurement*) є апіорним, теоретичним, спрямованим на створення такої моделі та одиниці педагогічного

вимірювання, за допомогою яких можна за однією шкалою порівнювати рівень знань студента із рівнем складності завдання.

Практичні переваги даного досягнення в області педагогічних вимірювань останніх років важко переоцінити.

Перша перевага полягає у стандартизованому характері отриманої одиниці вимірювання рівня знань студентів. Як і будь-яка стандартизована одиниця вона є результатом перетворення вихідних даних, що надає можливість об'єктивно порівнювати досягнення різних студентів за різними дисциплінами.

Значення рівня знань студента (латентна змінна) θ_i та рівень складності j -того завдання тесту, що вимірюється на латентному континуумі знань, β_j на першому етапі створення ефективного набору тестових завдань можуть бути апроксимовані з матриць тестових результатів, які згідно з моделлю Г. Раша зведені програмним шляхом у таблиці, приклад структури яких наведений на рис. 1.

Предмет: Математичні методи в АСУ
Тест: Іспит ММАСУ
Група: АУ-02-1

№ Студент	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Y(i)	p(i)	q(i)	p(i)/q(i)	ln(p(i)/q(i))
1 Бабюк Володимир	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	5	0.33	0.67	0.49	-0.71
2 Безгачнюк Юрій	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	12	0.8	0.2	4	1.39
3 Бурій Ігор	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	3	0.2	0.8	0.25	-1.39
4 Васків Ігор	1	0	1	1	0	0	0	0	0	0	0	1	0	0	1	5	0.33	0.67	0.49	-0.71
5 Воевода Владислав	1	1	0	0	0	1	0	0	0	1	1	1	0	1	0	7	0.47	0.53	0.89	-0.12
Rj	5	2	2	3	2	3	0	1	0	2	2	4	1	2	3	32				
Wj	0	3	3	2	3	2	5	4	5	3	3	1	4	3	2					
p(j)	1	0.4	0.4	0.6	0.4	0.6	0	0.2	0	0.4	0.4	0.8	0.2	0.4	0.6	6.4				
q(j)	0	0.6	0.6	0.4	0.6	0.4	1	0.8	1	0.6	0.6	0.2	0.8	0.6	0.4					
p(j)q(j)	0	0.24	0.24	0.24	0.24	0.24	0	0.16	0	0.24	0.24	0.16	0.16	0.24	0.24					
q(j)/p(j)	0.01	1.5	1.5	0.67	1.5	0.67	0.01	4	0.01	1.5	1.5	0.25	4	1.5	0.67					
ln(q(j)/p(j))	-4.61	0.41	0.41	-0.4	0.41	-0.4	-4.61	1.39	-4.61	0.41	0.41	-1.39	1.39	0.41	-0.4					

Експорт даних у форматі Winstep(Ministep)

Рис. 1. Матриця тестових результатів (за теорією Г. Раша)

У таблицю зведені наступні статистичні дані:

p_i – кількість вірних відповідей студента на завдання всього тесту; q_i – кількість хибних відповідей того самого студента на завдання даного тесту; p_i / q_i – потенціал i -того студента; $\ln(p_i / q_i)$ – логіт знань (за Г. Рашем); p_j – кількість студентів, що вірно відповіли на j -те завдання даного тесту; q_j – кількість студентів, які не справились з поставленим завданням; $\ln(q_j / p_j)$ – логіт складності j -того завдання даного тесту.

Логарифмічні оцінки рівня знань студентів та рівня складності тестових завдань дали можливість порівнювати їх між собою, що обумовило значний вплив на розвиток закордонної педагогічної теорії і практики. Вперше з'явився інструмент для безпосереднього зіставлення будь-якого завдання тесту з будь-яким випробовуваним, і на основі цього зі-

ставлення обчислити імовірність вірної відповіді.

2. Переваги використання параметричної теорії Г. Раша

1. На основі теорії Г. Раша і програмного алгоритму його реалізації ЕОМ самостійно добирає студентові завдання в системах адаптивного навчання та діагностування знань. Якщо імовірність вірної відповіді є низькою, то обирається завдання меншої складності, у протилежному випадку – навпаки. Загальний принцип підбору завдань відповідає 50% імовірності отримання вірної відповіді.

2. Перевага даного методу вимірювання рівня знань заснована на специфічних особливостях моделі Г. Раша. Отримані з її допомогою оцінки рівня знань студентів, через відносну незалежність від конкретного підбору того або іншого завдання, набувають характеру достатньо об'єктивних результатів тестування, що також позитивно відбивається на якості оцінок, що використовуються під час педагогічної діагностики.

3. Ця ідея перетворилась у нове для науки положення про специфічну об'єктивність, як основу отримання вірних оцінок рівня знань суб'єктів, що не залежать ні від конкретного набору тестових завдань, ні від підбору групи студентів та самого викладача. Не випадково у західній літературі вимірювання за Г. Рашем ще називають *model based measurement*.

4. Перевага даної теорії пов'язана із можливостями отримання шкали інтервалів. Rasch Measurement володіє всіма достатніми ознаками фундаментальної теорії, яка має досить просту аксіоматику, що зводиться до простих тверджень: властивість суб'єкта діагностики знань, що цікавлять викладача, існує у латентному стані, воно є нестійким, і тому, вимірюється з деякою похибкою. Імовірність вірної відповіді випробовуваного залежить від співвідношення рівня його підготовки і від рівня складності завдань.

5. Теорія Г. Раша виявилась несуперечливою, ефективною та має підтвердження у тисячах практичних застосувань. Висунутий нею принцип *reparability of estimates* дозволяє звільнитися від нестійких статистик на вибірках завдань і вибірках студентів для отримання необхідних параметрів діагностування.

6. Перевагою моделі Г. Раша є порівняльна стійкість розрахованих значень рівня знань та складності тестових завдань. Це дає підстави стверджувати, що однопараметричні моделі вдало оцінюють якості особистості, що цікавлять педагога, і є недоступними для безпосереднього вимірювання.

Висновки

Можна говорити про можливість використання методології Rasch Measurement в практиці дистанційного діагностування рівня знань та проектуванні систем адаптивного контролю рівня знань, вмінь та навичок студентів, яка базується на філософії об'єктивного пізнання, має власний метод фундаментального вимірювання із власною одиницею та властивостями шкали інтервалів. Дана методологія відповідає всім ви-

могам, що висувалися видатними вченими до психологічних і педагогічних вимірювань:

1. Лінійність, що припускає можливість застосування арифметичних властивостей та операцій.
2. Взаємна незалежність параметрів завдань та суб'єктів випробування.
3. Відносна простота методу вимірювання параметрів діагностування рівня знань.
4. Одновимірність вимірюваного об'єкта.
5. Монотонність відображення властивостей у числовій шкалі вимірювання.

Слід наголосити на особливості процесу і культури теорії Rasch Measurement. Процес складається із етапів, культура включає у себе філософські, теоретичні і метрологічні основи вимірювання. В центр даної культури можна покласти такі головні ідеї:

1. Модель Model-based measurement, що можна трактувати як принципово інший підхід до аналізу статистичних даних діагностування рівня знань студентів. Якщо, зазвичай, намагаються знайти математичну модель для найкращого опису даних, то у теорії Г. Раша, навпаки, дані повинні відповідати моделі вимірювання. Багато західних спеціалістів надавали цій вимозі особистісно-психологічне тлумачення неадекватності позиції автора. В той же час, тут має місце інша дослідницька методологія та інша, більш просунута культура вимірювання.

2. Ідея залежності імовірності вірної відповіді студентів, яка виражається відповідною функцією імовірності вірної відповіді студента на завдання, у залежності від різниці двох значень – рівня підготовленості до тестування і міри складності завдання.

3. Одновимірність, що відображається у намаганні вимірювати лише одну важливу властивість особистості (латентність). В існуючих на сьогоднішній день тестових системах намагаються вимірювати рівень знань тільки за одною учбовою дисципліною.

4. Ідея локальної незалежності результатів, яка формулюється як аксіома для випробовуваних однакового рівня підготовленості: імовірність вірної відповіді на одне завдання не повинна залежати від імовірності вірної відповіді на будь-яке інше завдання.

5. Параметри завдань і випробовуваних не повинні бути взаємозалежними.

Модель педагогічних вимірювань Rasch Measurement має відомі переваги побудови лінійної шкали інтервалів вимірювання, із достатньою статистикою, а також ефективні математичні обґрунтування. Тому можна стверджувати, що Rasch Measurement є не просто спрощеною однопараметричною моделлю Item Response Theory, як це вважається у Західних наукових колах, а зовсім іншою культурою вимірювання рівня знань студентів та їх діагностики.

Література

1. Анастаси А. Психологическое тестирование. Кн.1 / А. Анастаси. – М.: Педагогика, 1982. – 320 с.
2. Аванесов В.С. Основы теории педагогических заданий / В.С. Аванесов // Педагогические Измерения. – 2006. – №2. – С. 26-62
3. Чельшкова М. Б. Разработка педагогических тестов на основе современных математических моделей: уч. пособие / М.Б. Чельшкова. – М.: Исследовательский центр проблем качества подготовки специалистов, 1995. – 32 с.
4. Безгачнюк Ю.В. Структура системи дистанційного діагностування рівня знань студентів / Ю.В. Безгачнюк, Л.М. Заміховський, Л.О. Сав'юк // Наукові вісті Івано-Франківського ІМЕ “Галицька академія”. –2005. – №2(8). – С.40-49.

*Стаття поступила в редакційну колегію 09.11.2009 р.
Рекомендовано до друку д.т.н., професором Заміховським Л.М.*

**MODELS AND METHODS OF STATISTICAL TREATMENT
OF RESULTS OF THE CONTROLLED FROM DISTANCE
DIAGNOSTICS OF LEVEL OF KNOWLEDGES****YO. V. BEZGACHNYOC**

*Ivano-Frankivs'k National Technical University of Oil and Gas;
76019, Ivano-Frankivs'k, Carpats'ka street,15;
phl. +380 (342) 4-80-00; e-mail: yurkovskiy@au.nung.if.ua*

Models and methods of statistical treatment of results of the controlled from distance diagnostics of level of knowledges of students are considered in the article, and also realization of the proper vehicle in the system of d-tester, which is used on a department CT and SU IFNTOUNG during five years

Key words: *educational process, controlled from distance diagnostics of level of knowledges of students, statistical treatment of results of researches.*