

УДК 303.21

СЕРГІЙ ДЕМБІЦЬКИЙ,

кандидат соціологічних наук, науковий співробітник відділу методології і методів соціології Інституту соціології НАН України

З приводу проблематичності статистичної перевірки теоретичних конструктів

Анотація

У статті описано результати методичних експериментів, присвячених проблемі статистичної перевірки теоретичних конструктів. Продемонстровано, що задовільні результати статистичної перевірки не є достатньою умовою валідазації шкал. Як приклад використовується комплексна шкала, котра дає змогу визначити тип студента залежно від докладених ним у процесі навчання зусиль (байдужий, декларативно зацікавлений, відповідальний).

Ключові слова: *шкалування, теоретична валідація, університетська освіта*

Саме по собі підтвердження гіпотез мало про що говорить, якщо спирається лише на кабінетні міркування та результати вимірювання (нехай його і заведено називати в соціології спостереженням). Навіть використання “ламаних” теоретичних конструктів може забезпечити отримання статистично значимих результатів, а також демонстрацію високих коефіцієнтів кореляції або значимих різниць (відповідний приклад наведено далі). Тому вимірювання, “замкнуте саме на себе”, годиться для розв’язання багатьох задач, але найкраще — для одержання наукоподібних артефактів.

Я вже не кажу про вельми поширені в соціології ситуації, коли статистично значимі результати, отримані на великих вибірках (скажімо, 1500 і більше одиниць), розцінюються як цінні з наукової точки зору (або як такі, що підтверджують щось) лише за фактом своєї статистичної значимості. А на великих вибірках, як відомо, практично все є статистично значимим.

Такі міркування висувають на передній план проблему теоретичної валідності, досягнення якої покликане забезпечити дослідника валідними стосовно досліджуваних соціальних феноменів теоретичними конструктами. Отже, теоретична валідизація багато в чому є автономною і визначальною для подальшої операціоналізації й емпіричної інтерпретації сферою. Легко уявити ситуацію, коли теоретичний конструкт описаний достатньо вичерпно, проте розроблена методика не забезпечує задовільного відображення цього конструкта. Згідно із “кращими” традиціями кількісного підходу в такій ситуації можна сміливо відмовитися від теоретичного конструкта, котрий не пройшов перевірку. І протилежний випадок — хоча методика дає можливість цілком адекватно відобразити конструкт, але цей конструкт змістовно не валідний (наприклад, охоплює тільки частину концептуального поля, що перебуває у фокусі дослідження). У цьому випадку висновки будуть хибними — цього разу вже стосовно правильності невалідного конструкта.

Для докладнішого опису останньої ситуації звернімося до такого прикладу. В одній зі своїх недавніх публікацій [Дембицкий, 2012] я описав специфіку теоретичної валідизації вимірювальних шкал. За демонстраційний приклад було взято шкалу, призначену для визначення типу студента, що вказує на його ставлення до навчання (байдужний, декларативно зацікавлений або відповідальний) на підставі застосовуваних студентом поведінкових стратегій.

Ця шкала має такий вигляд:

Припустімо, що в узагальненому вигляді студентів можна розподілити на три типи залежно від їхнього ставлення до навчання:

- А) найменш сумлінні студенти, які здебільшого не докладають зусиль для навчання, сподіваються на вдалий збіг обставин і на те, що всі борги можна буде закрити наприкінці семестру, а іспити та заліки скласти завдяки допомозі товаришів по групі або навіть якщо практично нічого не робити;*
- В) які балансують між підготовкою найважливіших завдань та ігноруванням завдань, що не мають принципового значення, які використовують ранжування викладачів за суворістю й вимогливістю з метою відповідної концентрації зусиль і які воліють відвідувати ті пари, що мають більшу вагу в рамках дисципліни;*
- С) найбільш старанні, які ставляться до навчання максимально відповідально, намагаються готувати всі завдання, передбачені навчальними вимогами, здобути максимум знань і постійно відвідують пари.*

А тепер, використовуючи наведену далі шкалу, ВКАЖІТЬ МІСЦЕ, НА ЯКЕ Б ВИ ПОСТАВИЛИ ПЕРЕЛІЧЕНИХ НИЖЧЕ СТУДЕНТІВ, порівнюючи із цими трьома типами. При цьому не можна обирати “чистий” тип (А, В чи С), а необхідно обрати ОДНЕ зі значень між двома “чистими” типами (між А і В АБО між В і С) .

А	1	2	3	4	5	В	6	7	8	9	10	С
	A-90% B-10%	A-70% B-30%	A-50% B-50%	A-30% B-70%	A-10% B-90%		B-90% C-10%	B-70% C-30%	B-50% C-50%	B-30% C-70%	B-10% C-90%	

У перебігу валідазації цієї шкали було проведено низку опитувань, у тому числі з допомогою експертів, якими виступали викладачі та самі студенти (у цьому разі вони оцінювали не тільки себе, а й своїх товарищів по групі). Опитування у трьох вишах, у рамках яких залучалися експерти ($N_1 = 20$, $N_2 = 24$, $N_3 = 23$), я й використовую для того, щоби показати, як “ламаний” конструкт може привести до надійних статистичних результатів.

У першому у другому опитуванні кожен з експертів (два викладачі та один студент) оцінили кожного зі студентів однієї навчальної групи за 10-бальною шкалою. Не вдаючись у подробиці, зазначу лише, що значення від 1 до 5 відповідали байдужому типу, від 6 до 7 – декларативно зацікавленому, від 8 до 10 – відповідальному. Далі було обчислено: а) узагальнену оцінку на підставі середнього арифметичного оцінок усіх експертів; б) коефіцієнт кореляції між узагальненою оцінкою кожного студента і його середнім балом успішності складання сесій за чотири роки навчання (див. табл. 1).

Таблиця 1

**Сила і напрям зв'язку результатів експертного оцінювання
із середніми значеннями оцінок, отриманих на всіх сесіях
(перший і другий ВНЗ)**

Експерти	Сила зв'язку	
	Перший ВНЗ	Другий ВНЗ
Перший викладач	0,83**	0,93**
Другий викладач	0,72**	0,91**
Студент	0,80**	0,93**
Середнє значення трьох оцінок	0,85**	0,95**

** $p < 0,01$.

У найкращих традиціях кількісного підходу я міг би вирішити, що я отримав підтвердження (принаймні часткове) правильності застосовуваного теоретичного конструкта. Утім, я й без того знав, що він правильний, адже я використовував валідний теоретичний конструкт (валідність конструкта була досягнута на попередніх етапах дослідження завдяки якісним методам).

Тепер подивімося, що відбудеться, якщо зі змісту використовуваного теоретичного конструкта прибрати значну частину, а саме інформацію про байдужий тип (частина шкали від 1 до 5). У цьому випадку кожному студентові, який одержав узагальнену оцінку від 1 до 5,499 (таких було вісім, або 36,4% від загальної кількості), необхідно присвоїти 6, а інші оцінки залишити без змін. Наскільки сильно змінився коефіцієнт кореляції? Після перетворення шкали він став дорівнювати 0,70 (можливість помилки, як і раніше, не перевищувала 0,01%).

Утім, це лише уявний експеримент. Тому я здійснив додатковий методичний експеримент у навчальній групі іншого навчального закладу. Експертами виступили два викладачі, добре знайомі з групою, і двоє старост

(кожен староста давав оцінку тільки для студентів своєї підгрупи). На першому етапі вони оцінили всіх студентів за шкалою, що була спеціально спотворена. У ній була відсутня друга частина (від 6 до 10), тобто оцінювання здійснювалося на підставі тільки двох типів студентів — байдужого і декларативно зацікавленого, представлених пунктами шкали від 1 до 5 (див. додаток). Лише один з експертів (староста) поставив валідність шкали під сумнів, але після пояснення, що це методичний експеримент і так треба, заповнив анкету. Були отримані такі результати: індивідуальні й усереднені результати оцінювань експертів показали високий взаємозв'язок із середнім балом успішності складання сесій за чотири роки навчання. За місяць я попросив цих експертів повторити оцінювання, але вже з використанням повноцінної 10-бальної шкали. У результаті індивідуальні коефіцієнти в першому випадку збільшилися, у другому — дещо зменшилися, у третьому залишилися практично незмінними. Що стосується узагальненої оцінки, то вона також практично не змінилася (див. табл. 2).

Таблиця 2

**Сила і напрям зв'язку результатів експертного оцінювання
із середніми значеннями оцінок, отриманих на всіх сесіях
(третій ВНЗ)**

Експерти	Сила зв'язку	
	Застосування “ламаної” шкали	Застосування валідної шкали
Перший викладач	0,52*	0,63**
Другий викладач	0,80**	0,74**
Старости	0,75**	0,74**
Середнє значення трьох оцінок	0,80**	0,81**

* $p < 0,05$;** $p < 0,01$.

Відшотгтовхуючись суто від статистичної інформації, можна дійти висновку, що валідність двох варіантів шкали (принаймні в разі використання стосовно цієї конкретної групи студентів) є однаковою. Разом із тим цілком імовірно, що в разі використання неповноцінної шкали експерти передусім орієнтувалися на числові пункти, а не на зміст двох описаних у запитанні типів. На мою думку, така ситуація може мати місце тоді, коли експерти добре ознайомлені з досліджуваною проблемою. Відповідно, неврахована в тексті запитання інформація все одно враховується ними внаслідок їхніх знань та досвіду. Таким чином, статистична валідність шкали компенсується когнітивними можливостями експертів. Натомість змістова (а ширше — теоретична) валідність шкали є сумнівною, оскільки вона внеможливіює перехід до трьох типів студентів. Своєю чергою, повноцінний варіант шкали є валідним в усіх аспектах — і в статистичному, і в теоретичному.

На цьому прикладі я намагаюся показати лише одне — спотворені конструкти повною мірою можуть дістати “підтвердження” на підставі резуль-

татів вимірювання з подальшою статистичною перевіркою. І цю проблему, на мій погляд, не можна розв'язати в рамках кількісного підходу.

Першочергове значення в розробленні валідних теоретичних конструкторів має належати вивченню навколишньої соціальної дійсності шляхом ретельного спостереження процесів, що в ній відбуваються, а також через особисту взаємодію з учасниками таких процесів.

Проте емпіричні свідчення, подані в статті, є доволі фрагментарними і, відтак, радше окреслюють методологічну проблему, ніж доводять положення, викладені у статті. Утім, повноцінне розв'язання проблеми потребує проведення подальших методичних експериментів.

ДОДАТОК

Припустімо, що в узагальненому вигляді студентів можна розокремити на два типи залежно від їхнього ставлення до навчання:

А) ті, хто балансує між підготовкою найважливіших завдань та ігноруванням завдань, що не мають принципового значення, хто використовує ранжування викладачів за суворістю та вимогливістю з метою відповідної концентрації зусиль і воліє відвідувати ті пари, що мають більшу вагу в рамках дисципліни;

В) найбільш сумлінні, які ставляться до навчання максимально відповідально, намагаються виконувати всі завдання, передбачені навчальними вимогами, здобути максимум знань і постійно відвідувати пари.

А тепер, використовуючи наведену далі шкалу, ВКАЖІТЬ МІСЦЕ, НА ЯКЕ Б ВИ ПОСТАВИЛИ ПЕРЕЛІЧЕНИХ НИЖЧЕ СТУДЕНТІВ, порівнюючи із цими двома типами. При цьому не можна обирати “чистий” тип (А чи В), необхідно обрати ОДНЕ зі значень між двома “чистими” типами.

А	1	2	3	4	5	В
	A-90% B-10%	A-70% B-30%	A-50% B-50%	A-30% B-70%	A-10% B-90%	

Джерела

Дембіцький С. Теоретическая валидизация измерительных шкал / Сергей Дембицкий // Социология: теория, методы, маркетинг. — 2012. — №2. — С. 53–65.