

УДК 303.214

Дембіцький С.С., канд. соціол. наук, старший науковий співробітник відділу методології та методів соціології Інституту соціології НАН України, вул. Шовковична, 12, Київ, 01021, Україна, e-mail: e_forge@me.com,

Любива Т.Я., канд. соціол. наук, науковий співробітник відділу методології та методів соціології Інституту соціології НАН України, вул. Шовковична, 12, Київ, 01021, Україна, e-mail: tlubivaya@gmail.com

ВИМОГИ ДО РОЗРОБКИ ТА АДАПТАЦІЇ КОМПЛЕКСНИХ ВИМІРЮВАЛЬНИХ ІНСТРУМЕНТІВ У СОЦІОЛОГІЇ: НАДІЙНІСТЬ, ВАЛІДНІСТЬ І ДОСТОВІРНІСТЬ

Подано головні характеристики якості соціологічного вимірювання та загальну логіку їх перевірки. Виявлено принципову методологічну відмінність між вимірювальними шкалами та індексами. Надійність концептуалізована як відтворюваність вимірювання. Валідність розглянуто в її теоретичній та емпіричній складових. Достовірність інтерпретується як можливість та (або) бажання респондента повідомити інформацію, що цікавить дослідника.

Ключові слова: надійність, валідність, достовірність, соціологічне вимірювання.

Під комплексними вимірювальними інструментами в соціології (далі – КВІ) автори розуміють, насамперед, шкали й індекси, які є комплексними в тому сенсі, що вони складаються з набору індикаторів, які вирішують єдине наукове завдання – вимір виразності деякого феномену. Їх комплексність зумовлюється і тим, що в процесі розробки необхідні спеціальні методичні дослідження, покликані продемонструвати наявність у таких інструментів конвенційно затверджених показників якості: надійності та валідності.

Якщо говорити про відмінності між шкалами та індексами [1, р. 12], то в першому випадку йдеться про вимірювання деякої властивості, яка безпосередньо не спостерігається (латентної змінної), що обумовлює ті аспекти її прояву, які закладені в основу індикаторів методики. У разі ж інтегральних індексів латентна змінна або соціальний феномен зумовлюються тими факторами, на підставі яких сформульовані індикатори.

Перевірка надійності та валідності є невід’ємними етапами розробки чи адаптації КВІ. Ці компоненти нерозривно пов’язані між собою, і тому, говорячи про надійність, ми розглядаємо поняття “валідність”, а про валідність – поняття “надійність”. Саме вони є основними характеристиками вимірювання в західній соціології. Крім цих понять у пострадянській соціології та психології використовується поняття “достовірність”, що стосується не інструменту, а власне отриманих даних, а в альтернативному підході – характеристик респондента.

Мета статті – запропонувати термінологічне бачення таких характеристик соціального вимірювання, як надійність, валідність і достовірність, з урахуванням як класичного бачення даної проблематики, так і сучасних напрацювань у соціології.

В основу дослідження покладено твердження, що концептуальні побудови стосовно якості соціологічного вимірювання мають інструментальний характер, а отже, не можуть отримати остаточної й єдиної вірної методологічної інтерпретації. З огляду на це, зміст статті – керівництво до розгляду проблем якості соціального вимірювання, яке

© Дембіцький С.С., Любива Т.Я., 2017

ISSN 1681-116X. Український соціум, 2017, № 3(62): 45–57

використовують її автори. Оскільки стаття має теоретико-методологічний характер, всі релевантні джерела згадуються у прив'язці до відповідних частин цієї роботи.

Надійність. Поняття “надійність” пов’язане з відсутністю похибок вимірювання. Останні в соціологічній науці розділяють на випадкові та не випадкові (або систематичні)¹. З численних визначень надійності, які можна зустріти в методичній літературі, найлаконічніше належить П. Алреку і Р. Сетглу, які характеризують надійність як “свободу від випадкових похибок” (на відміну від валідності, яку визначають як “свободу і від випадкових, і від систематичних зміщень даних соціального дослідження”) [2, р. 64].

Говорячи про КВІ для масових опитувань, поняття “надійність” більш доцільно використовувати як відтворюваність вимірювання. Тут надійність пов’язана з повторюваністю у випадку неодноразового вимірювання одних і тих самих соціальних феноменів з використанням одного й того самого вимірювального інструменту. Чим стійкішими є результати вимірювання при повторних спробах, тим надійнішим є КВІ, і навпаки, чим менш схожими є результати вимірювання при повторному вимірюванні, тим менш надійним є інструмент [3, р. 11–12].

У загальному вигляді соціологічне вимірювання може бути подано такою формулою:

$$X = t + e, \quad (1)$$

де X – спостережуване значення, t – істинне значення, а e – випадкова похибка.

Надійність також може бути виражена у термінах похибки вимірювання:

$$\rho_x = 1 - [\text{VAR}(e)/\text{VAR}(X)]. \quad (2)$$

З формули видно, що надійність змінюється від 0 до 1: чим більшою є варіація похибки по відношенню до спостережуваної варіації, тим ближчою є надійність до 0, тобто тим ненадійнішим є вимірювання.

Коли ми говоримо про визначення надійності КВІ, необхідно пам’ятати про те, що надійними також мають бути окремі пункти такого показника. У технічному плані окремі пункти завжди є змінними, які традиційно розділяють на номінальні, порядкові та метричні. Відповідно до цього повинні варіюватись і методи оцінки надійності (табл. 1).

Фактично, при встановленні надійності КВІ виникає завдання визначення надійності самого показника та показника через надійність усіх вихідних ознак [4, с. 67].

Таблиця 1

Методи оцінки надійності для змінних різних типів

Шкала вимірювання	Метод оцінки надійності
Номінальна	Коефіцієнт Крамера
Порядкова	Коефіцієнт Гамма (табличний формат), коефіцієнт рангової кореляції Спірмена
Метрична	Коефіцієнт кореляції Пірсона

Джерело: складено автором.

¹ Випадкові похибки – такі похибки, що є результатом взаємодії великої кількості факторів, які окремо є незначними. Випадкові похибки можна оцінити, вони зменшуються при збільшенні вибірки. Систематичні похибки виникають при неправильному відборі одиниць дослідження.

Тому перейдемо до підходів перевірки надійності КВІ, вважаючи, що різні види надійності можуть розглядатись як різні методи її перевірки. У соціологічній літературі можна зустріти такі методи надійності [5]:

- тесту-ретесту (Test-Retest Reliability);
- паралельних форм (Parallel-Forms Reliability);
- внутрішньої узгодженості (Internal Consistency Reliability).

Надійність тесту-ретесту визнається рівнем стійкості вимірювання через деякий проміжок часу з використанням тієї самої вибірки. При реалізації цього методу замір проводиться двічі з визначеним інтервалом часу для одних і тих самих респондентів та з використанням того самого інструментарію. Після цього розраховується коефіцієнт кореляції між двома замірами: даними, отриманими при першому опитуванні, та даними, отриманими при другому опитуванні. Чим вищою є кореляція між результатами цих замірів, тим вищою є надійність вимірювання. Цей метод має ряд обмежень. По-перше, його результати є дуже чутливими до величини проміжку часу між замірами: чим більше часу минуло, тим менш надійною виявиться шкала. Рекомендований проміжок часу між замірами – від двох тижнів до одного місяця [3, р. 40]. По-друге, сам факт першого заміру може змінити ставлення респондентів до об'єкта, що вивчається. По-третє, бувають випадки, коли повторний замір взагалі неможливий, наприклад, якщо потрібно заміряти першу реакцію респондентів стосовно певного об'єкта дослідження. По-четверте, при другому замірі респонденти можуть намагатися згадати, що вони говорили під час першого. По-п'яте, за час, що минув, погляди респондентів за певними питаннями можуть змінитися. Проте даний метод перевірки надійності підходить для КВІ, результати використання яких з плином часу значно не змінюються.

Надійність паралельних форм (також можна зустріти назви “еквівалентних” чи “альтернативних” форм) може бути визначена як міра еквівалентності двох форм одного й того самого тесту (шкали) [6, р. 112]. При цьому використовується така процедура: ті самі респонденти з вибірки дають відповіді за основним варіантом тесту (шкали), а потім – за еквівалентним або альтернативним варіантом. Паралельна форма тесту підбирається таким чином, щоб обидва варіанти були максимально співставними, з точки зору отримуваних результатів. Кореляція між результатами еквівалентних форм одного КВІ є оцінкою надійності. Однією з особливостей є те, що дослідникам слід мати досить велику кількість пунктів-складових даного тесту (шкали). Крім того, цей метод базується на припущенні, згідно з яким дві частини тесту (шкали) є паралельними, тобто ідентичними за змістом.

Внутрішня узгодженість шкали визначається, переважно, шляхом розбиття даних, отриманих з використанням певного інструменту, на частини та пошуком у цих підвибірках однакової структури.

У цьому випадку аналізується, наскільки узгодженими є результати для різних індикаторів (пунктів) одного й того самого КВІ в рамках одного й того самого вимірювання. Розрізняють такі методи для встановлення внутрішньої узгодженості:

1. Середня кореляція між пунктами шкали (Average Inter-item Correlation) – розраховується як середнє кореляцій кожної пари пунктів шкали на основі кореляційної матриці.

2. Середня кореляція з підсумковим значенням (Average Itemtotal Correlation) – фактично, це є тим самим, що й попереднє, але як один з компонентів є загальна змінна, що поєднує всі пункти шкали (і з цією змінною також розраховуються кореляції з окремими пунктами шкали).

3. Розщеплення шкали (Split-Half Reliability) – кореляція між значеннями пунктів шкали, які випадковим чином розділені на дві групи. Іноді робиться поділ за парними та непарними пунктами шкали. Цей метод є близьким до надійності паралельних форм за винятком того, що потрібен один замір (зазвичай у цьому випадку використовують коефіцієнт кореляції Спірмена-Брауна).

4. Коефіцієнт Альфа Кронбаха (Cronbach's Alpha (α)). Цей коефіцієнт оцінки надійності є математично еквівалентним середньому значенню всіх кореляцій усіх можливих розщеплень шкали однієї й тієї самої вибірки. При додаванні пунктів у шкалу завжди підвищується значення коефіцієнта. Водночас додавання пунктів ускладнює шкалу, тому варто звертати увагу на коефіцієнт Альфа Кронбаха при видаленні певного пункту шкали.

Характеристики підходів до перевірки надійності показано у табл. 2.

Таблиця 2

Підходи до перевірки надійності КВІ

Підхід	Кількість замірів	Необхідність панелі	Статистичні критерії	Що взяти до уваги
Тест-ретест	2	Так	Кореляція	Вибір проміжку часу між тестами
Надійність паралельних форм	2	Так	Кореляція	Побудова паралельних форм, які є ідентичними за змістом
Надійність внутрішньої узгодженості	1	Ні	Кореляція, коефіцієнт Альфа Кронбаха	Вибір методу встановлення узгодженості

Джерело: складено автором.

Валідність. В найбільш загальному розумінні під валідністю ми розуміємо ступінь відповідності результатів дослідницького процесу дійсності [7, р. 190–191]. Щодо комплексного підходу до процесу валідації вимірювальних шкал й інтегральних індексів, то він включає два етапи: теоретичну та емпіричну валідацію. Теоретична валідація передбачає досягнення достатнього рівня відповідності теоретичного конструкту досліджуваному феномену, а емпірична – перевірку відповідності вимірювальної методики та отриманих на її основі результатів теоретичному конструкту. Таким чином, спочатку досягається теоретична валідність, а вже потім перевіряється емпірична.

Теоретична валідація. Щоб уникнути проблеми “дурної собаки, що ганяється за своїм хвостом” (де “дурна собака” – це теорія, яку мають намір верифікувати за допомогою вимірювальних процедур, а “хвіст” – вимірювальна методика, побудована на базі ще неперевіреної теорії), в основу КВІ повинні бути покладені теоретичні положення, що вже є валідними. Така постановка питання пов’язана з тим, що “ламані” теоретичні конструкти цілком можуть приводити до прийнятних результатів статис-

тичного аналізу. З метою підтвердження цього положення проведено методичне дослідження, в рамках якого порівнювалися результати використання методики, побудованої на підставі вже валідних теоретичних положень, з результатами тієї ж методики, яка навмисно була викривлена ² таким чином, щоб порушити її змістовну валідність [8]. За допомогою двох цих варіантів вимірювальної методики викладачі та старости одного з київських вузів оцінювали поведінкові стратегії своїх студентів (або одноступіньників). Далі отримані результати були зіставлені з середніми балами студентів, які вони отримали на всіх сесіях бакалаврату (табл. 3). Отже, обидва варіанти, з точки зору статистики, є рівнозначними, чого не можна сказати про теоретичні конструкти, які за ними стоять.

З точки зору статистичного аналізу (наприклад, знаходження взаємозв'язків з іншими конструктами або передбачення значень інших змінних), обидва варіанти шкали можна оцінити як рівноцінні. Проте, з точки зору соціологічної інтерпретації, “викривлена” шкала втрачає цінність, оскільки відмова від змістовних елементів призводить до втрати інформації про один з трьох типів студентів ³.

Наступне питання стосується того, що є джерелом тих теоретичних положень, які можуть розглядатись як валідні. Оскільки феномени, що потрапляють у фокус соціологічного вимірювання, можуть бути соціальними (наприклад, ставлення до владних інститутів), соціально-філософськими (наприклад, справедливість), соціально-психологічними (наприклад, психологічний клімат у колективі) та психологічними (наприклад, тривожність), то до таких джерел можуть бути віднесені окремі концепції в рамках соціальних наук, попередні спеціальні дослідження, професійний практичний досвід, а в певних випадках і універсальний досвід, тобто знання, що властиві всім представникам певної соціальної спільності. Розглянемо їх детальніше.

Таблиця 3

Сила та напрям зв'язку результатів експертного оцінювання з середніми значеннями оцінок, отриманих на всіх сесіях бакалаврату (N = 23)

Експерти	Сила зв'язку	
	Використання “викривленої” шкали	Використання валідної шкали
Викладач № 1	0,52*	0,63**
Викладач № 2	0,80**	0,74**
Старости	0,75**	0,74**
Середнє значення трьох оцінок	0,80**	0,81**

* $p < 0,05$;

** $p < 0,01$.

Джерело: складено автором.

² Змістовно валідний і навмисно викривлений варіанти методики наведені у Додатку.

³ За допомогою частини шкали, що була залишена, усі студенти можуть бути віднесені або до “декларативно зацікавленого”, або до “відповідального” типу, в той час як “байдужий” тип випадає з фокусу дослідження. Про перелічені типи студентів і розробку методики для їх визначення див.: [9, с. 109–124; 10].

Щодо наукових концепцій, то слід говорити про ті з них, побудова яких спиралася на дослідницьку стратегію кейс-стаді [9, с. 125–149; 11]. При цьому ми вважаємо такими і деякі концепції історичної макросоціології. До останніх, наприклад, належать результати, представлені в роботах Ч. Тіллі [12] та Б. Андерсона [13]. Ці теорії хоч і виглядають вельми далекими від рівня соціологічного вимірювання, але задають правильне фокусування щодо найрелевантніших процесів у сьогоdnішньому світі національних держав, тим самим збільшують теоретичну чутливість дослідника.

Відмінною рисою проведення кейс-стаді з метою отримання валідних теоретичних положень є акцент на вивченні соціальних феноменів, які проявляються в соціальній дійсності⁴, з подальшим переходом до теоретичних положень, які вимагають не верифікації, а інспектування та збагачення в рамках тих соціальних контекстів, які не були охоплені ініціальним дослідженням.

Що стосується попередніх комплексних досліджень, то йдеться про використання кейс-стаді. Відмінність тільки в тому, що тут дослідження “заточене” під побудову необхідної для подальшого соціологічного вимірювання теоретичної бази. Відповідним прикладом є дослідження поведінкових практик студентів, які суміщають роботу з навчанням [9, с. 99–124].

У випадку професійного досвіду все значно простіше. Виконання щоденних функціональних обов’язків дозволяє фахівцям отримати розуміння різних аспектів соціальних феноменів природним шляхом. Тому такий досвід є потенційним джерелом валідних теоретичних положень. Прикладом використання професійного досвіду при побудові вимірювальної методики є шкала психологічного дистресу SCL-90-R Л. Дерогатіса [14], який відштовхувався від тих симптомів різних психологічних розладів, які були зафіксовані професійним співтовариством.

Як відправна точка можуть використовуватися й аксіоматичні наукові положення, валідність яких не викликає сумнівів. Так, це можуть бути твердження, які розглядаються як базисні в ключових соціологічних теоріях (функціоналізм, символічний інтеракціонізм, теорія конфлікту, феноменологія, теорія обміну) або ж універсальні соціологічні аксіоми. Прикладом останніх є положення, використане Є. Головахою та Н. Паніною для конструювання інтегрального індексу соціального самопочуття: “соціальне самопочуття людини визначається ступенем задоволення його соціальних потреб, які, в свою чергу, є похідними від існуючої в суспільстві системи соціальних благ, їх виробництва та розподілу” [15, с. 23–24].

Розглянуті джерела можуть робити різний внесок у досягнення теоретичної валідності. Тому природно використовувати кілька з них, якщо це можливо, в рамках конкретного дослідницького проекту.

Говорячи безпосередньо про теоретичну валідність, слід розуміти, що теоретична база дослідження повинна забезпечувати три її компоненти: очевидний, змістовний і композиційний. Очевидна валідність пов’язана з аргументацією, в рамках якої досліджуваний феномен вписується в соціальну дійсність і, відповідно, не є “плодом”

⁴ Під “соціальною дійсністю” розуміються людські вчинки на різних рівнях соціальної взаємодії, що спрямовані на створення та підтримання соціальних інститутів.

чийхось домислів; змістовна валідність забезпечує значущість і репрезентацію елементів теоретичного конструкту; композиційна – правильність його внутрішньої структури [9, с. 22–31]. У такому вигляді теоретична база надає все необхідне для подальшої операціоналізації теоретичних положень, а також оцінки результатів, отриманих у рамках емпіричної валідації.

Після підготовки валідної теоретичної бази дослідник конструює на її основі вимірювальну шкалу або інтегральний індекс. Перевірка правильності переходу від теоретично валідної бази до інструменту вимірювання здійснюється за допомогою емпіричної валідації, що спирається на результати методичних опитувань.

Емпірична валідація. Емпірична валідність має три незалежних і один залежний компонент. Незалежні охоплюють такі види валідності як конструктивна, критеріальна та передбачувальна, залежний – порівняльну валідність. Ці чотири види слід розглядати як основні у контексті емпіричної валідації комплексних вимірювальних інструментів. Крім них у конкретних випадках може знадобитися перевірка додаткових (або неосновних) видів емпіричної валідності. Так, популярна сьогодні факторна валідність належить до цієї категорії. Тут ми приділимо увагу тільки основним видам емпіричної валідності.

Конструктивна валідність розглядається в термінах, запропонованих в роботі Д. Кемпбелла та Д. Фіске [16]. Так, для її підтвердження необхідно продемонструвати конвергентну й дискримінантну складові. Перша ґрунтується на тому, що результати методики пов'язані з результатами тих шкал або індексів, які спрямовані на вимірювання явищ, теоретично пов'язаних з феноменом, що знаходиться у фокусі нашого інтересу. Більш просто – вимірювання конструктів, які пов'язані в теорії, повинні бути пов'язаними й фактично. У разі дискримінантної валідності логіка є тією ж самою, тільки до уваги беруться ті аспекти, які теоретично не мають бути пов'язаними. Відповідні вимірювання повинні це продемонструвати. Якщо говорити про статистичні процедури, то найприйнятнішим є використання кореляційного (в широкому сенсі) аналізу.

Критеріальна валідність належить до ступеня, в якому результати вимірювання пов'язані з якимось об'єктивним результатом. Зазвичай її поділяють на конкурентну та передбачувану (або прогностичну, або предиктивну) валідності. Проте з міркувань, які наведені далі, під критеріальною валідністю вимірювальної методики розуміється виключно конкурентна, а прогностична є окремим видом. У цьому сенсі критеріальна валідність перевіряється шляхом порівняння результатів вимірювання з іншими спостережуваними результатами, що є об'єктивними показниками досліджуваного феномену. Одним з простих, але переконливих способів демонстрації критеріальної валідності комплексного вимірювального інструменту є отримання очікуваних відмінностей при опитуванні “контрастних”, з точки зору досліджуваного явища, груп, що серед іншого свідчить і про загальну “чутливість” методики.

Передбачувальна валідність пов'язана з можливістю інструменту надавати інформацію, яка є необхідною для прогнозування майбутніх станів об'єкта дослідження. З точки зору соціологічного вимірювання, це, можливо, найбільш важко досяжний вид валідності – як через необхідність стислості інструментарію, так і через комплексність досліджуваних у соціології процесів і явищ. Саме тому ми проводимо сутнісну від-

мінність між критеріальною та прогностичною валідністю. У цілому результати соціологічного вимірювання (якщо ми говоримо про конкретну методику) – лише одне з можливих джерел емпіричної інформації, на підставі якої будується соціологічний прогноз. Тому автор методики повинен розуміти, яким чином результати його шкали, індексу або тесту вбудовуються в певну галузь соціології з тим, щоб поліпшити висновки про можливі сценарії розвитку соціальної дійсності.

Порівняльна валідність розуміється як порівняльна перевага вимірювального інструменту у ряді схожих методик. Таким чином, її обґрунтування пов'язане з необхідністю порівняння показників конструктивної, критеріальної та прогностичної валідності різних соціологічних методик, спрямованих на вивчення одного й того самого явища. Саме тому даний різновид валідності є залежним компонентом емпіричної валідації – результати її перевірки пов'язані як з попередніми результатами емпіричної валідації, так і з тією методикою (і результатами її валідації), яка виступає в ролі головного “конкурента”.

Достовірність. Існує два підходи до розгляду поняття “достовірність”. У рамках першого вона розуміється як ступінь відповідності отриманого знання дійсності [17, с. 22], а в рамках другого – як можливість і (або) бажання респондента повідомити інформацію, що цікавить дослідника [18, с. 154–155]. Тут ми дотримуємося другого підходу, оскільки перший у контексті сучасного термінологічного поля пов'язаний з валідністю.

При такій постановці питання достовірність є вкрай важливою характеристикою збору даних за допомогою КВІ, оскільки навіть використання надійної та валідної вимірювальної методики саме по собі не забезпечує отримання необхідних досліднику даних у випадку, коли респондент не хоче повідомляти інформацію чи з яких-небудь причин не може цього зробити. Відповідні зміщення належать до категорії помилок, які не пов'язані з вибіркою (*nonsampling error*) [19, р. 42–46].

Для вирішення цієї методологічної проблеми співробітники Інституту соціології НАН України у 2006–2007 рр. розробили й апробували індекс суб'єктивної надійності респондента [20], який розуміється як інтегральна оцінка респондентом самого себе як достовірного джерела соціальної інформації. Концепт суб'єктивної надійності містить такі компоненти, що підлягають самооцінюванню: 1) зрозумілість інструменту опитування або його частини; 2) достатність знань для відповіді на запитання чи блоку запитань; 3) інтерес до теми опитування; 4) впевненість у виборі найліпшого варіанта відповіді на поставлені запитання; 5) сталість власної думки; 6) незалежність думки; 7) можливість широко відповідати на поставлені запитання. При цьому вимірювання концепту суб'єктивної надійності може будуватися на самооцінках, які стосуються окремих запитань, блоків запитань або опитувальника у цілому.

Висновки. Обираючи показники надійності, валідності та достовірності з метою підвищення якості інструменту вимірювання, дослідник може орієнтуватися на ряд рекомендацій (їх методологічне обґрунтування виходить за межі цієї статті).

Насамперед, необхідно забезпечити теоретичну валідність вимірювання, оскільки доцільність усієї подальшої роботи залежить від її забезпечення. Стосовно рішення цього завдання континуум усіх можливих ситуацій попадає в діапазон між розробкою

скороченої версії вже існуючої вимірювальної методики та конструюванням нового інструменту “з нуля”. При роботі з готовою методикою необхідно здійснити оцінку її теоретичної валідності принаймні в двох компонентах – очевидному та змістовному. Композиційну валідність можна відтворити в процесі отримання емпіричного матеріалу. У випадку конструювання нового інструменту слід подбати про всі компоненти теоретичної валідності.

Ідея, на якій ґрунтується перевірка емпіричної валідності, полягає в демонстрації адекватності отриманого інструменту вимірювання тій теоретичній базі, яка покладена в його основу. Використовуючи відповідні результати, дослідник вирішує, чи правильно підібрані та оформлені індикатори, а також обрана загальна логіка побудови вимірювального інструменту. З урахуванням цього, більша кількість перевірених видів валідності (конструктивна, критеріальна, прогностична) дає більше підстав для позитивної оцінки емпіричної валідації.

Якщо ж можливості перевірки емпіричної валідності обмежені, то план-мінімум – це перевірка конструктивної валідності, що не тільки продемонструє коректність інструменту, але й дозволить з’ясувати його вимірювальні властивості. У цьому випадку достатньо невеликої вибірки ($N = 100$), що є гетерогенною за основними соціально-демографічними показниками (наприклад, статтю і віком) та є близькою до характеристик генеральної сукупності. Останнє дозволить отримати приблизні результати щодо типового рівня досліджуваної властивості (звісно, якщо вибірка не зміщена за певним релевантним параметром). Перевірка цього виду валідності на великій репрезентативній вибірці виконує ті ж самі завдання, але надійність отриманих результатів, як і можливості аналізу, в такому випадку будуть найвищими.

Наступний пріоритет – перевірка критеріальної валідності з метою оцінки вимірювальної чутливості методики. Розмір кожної критеріальної групи починається приблизно з 50 спостережень. При цьому головна увага повинна бути приділена визначенню релевантних груп (тих, які демонструють теоретично очікувану специфіку у відповідях), а також тих груп, які, не дивлячись на свою специфіку в контексті досліджуваної властивості, демонструють типовий рівень досліджуваної властивості. Це дозволить визначити як сильні сторони методики, так і обмеження у її використанні.

Нарешті, прогностична валідність потребує тривалого використання вимірювального інструменту в широкомасштабних репрезентативних опитуваннях і врахування результатів його використання під час розробки сценаріїв розвитку подій на різних рівнях соціальної взаємодії. Розмір внеску відповідних результатів до таких сценаріїв, а також співставлення останніх з дійсними подіями, дозволить з’ясувати евристичний та інтерпретаційний потенціал методики.

Перевірка надійності зазвичай зводиться до демонстрації внутрішньої узгодженості за допомогою Альфа Кронбаха. Це зумовлено як загальною методологічною модою в сучасних соціальних науках, так і прагматичними міркуваннями: організаційною доступністю – необхідністю проведення лише одного методичного опитування. Проте перевірка внутрішньої узгодженості має сенс лише в тому випадку, коли розробляється або адаптується вимірювальна шкала. Відповідні результати можуть бути отримані під час конструктивної валідації без необхідності змінювати об’єм вибірки.

Що стосується інтегральних індексів, то перевірка внутрішньої узгодженості не є доцільною. Так, оскільки окремі частини інтегрального індексу є важливими, але не обов'язково пов'язаними між собою причинними факторами вимірювальної властивості, оскільки демонстрація взаємозв'язку між індикаторами індексу не має методологічної цінності.

У випадку перевірки ретестової надійності також треба розуміти доцільність такої процедури. Це пов'язано з тим, що у фокус соціології часто попадають досить мінливі характеристики, що накладає суттєві обмеження на можливість перевірки стабільності відповідних показників. У випадку ж, коли досліджується стабільна властивість, демонстрація її ретестової надійності становить більшу цінність, ніж демонстрація внутрішньої узгодженості, незалежно від того, який КВІ розробляється: вимірювальна шкала чи інтегральний індекс. При цьому прийнятний розмір вибірки становить 50 спостережень.

Додаток

Змістовно валідний варіант методики

Припустимо, що в найзагальнішому вигляді студентів можна розділити на три типи, залежно від їх ставлення до навчання:

А – найменш старанні студенти, які в основному не докладають зусиль для навчання, сподіваються на вдалий збіг обставин і на те, що всі борги можна буде закрити в кінці семестру, а іспити і заліки скласти за допомогою одногрупників або навіть нічого не робити;

Б – студенти, які балансують між підготовкою наважливіших завдань і ігноруванням завдань, що не мають принципового значення, використовують ранжування викладачів за суворістю та вимогливістю з метою відповідної концентрації зусиль і вважають за краще відвідувати ті пари, які мають більшу вагу в рамках дисципліни;

В – найбільш старанні, які ставляться до навчання максимально відповідально, намагаються готувати всі завдання, передбачені навчальними вимогами, отримати максимум знань і постійно відвідують пари.

Використовуючи наведену шкалу, *зазначте місце, на яке ви поставили б студентів, порівнюючи з цими трьома типами*. При цьому не можна вибирати “чистий” тип (А, Б або В), необхідно вибрати одне із значень між двома “чистими” типами (між А і Б або між Б і В).

А	1	2	3	4	5	Б	6	7	8	9	10	В
	А–90% Б–10%	А–70% Б–30%	А–50% Б–50%	А–30% Б–70%	А–10% Б–90%		Б–90% В–10%	Б–70% В–30%	Б–50% В–50%	Б–30% В–70%	Б–10% В–90%	

Змістовно невалідний варіант методики

Припустимо, що в найзагальнішому вигляді студентів можна розділити на два типи, залежно від їх ставлення до навчання:

А – студенти, які балансують між підготовкою найважливіших завдань та ігноруванням завдань, що не мають принципового значення, використовують ранжування викладачів за суворістю та вимогливості з метою відповідної концентрації зусиль і вважають за краще відвідувати ті пари, які мають більшу вагу в рамках дисципліни;

Б – найбільш старанні, які ставляться до навчання максимально відповідально, намагаються готувати всі завдання, передбачені навчальними вимогами, отримати максимум знань і постійно відвідують пари.

Використовуючи наведену шкалу, *вказуйте місце, на яке ви поставили б студентів, порівнюючи з цими двома типами*. При цьому не можна вибирати “чистий” тип (А або Б), а необхідно вибрати одне із значень між двома “чистими” типами.

А	1	2	3	4	5	Б
	А–90%	А–70%	А–50%	А–30%	А–10%	
	Б–10%	Б–30%	Б–50%	Б–70%	Б–90%	

Список використаних джерел

1. DeVellis R. Scale Development: Theory and Applications. Los Angeles; London; New Delhi; Singapore; Washington DC: SAGE, 2012. 216 p.
2. Alreck P.L., Settle R.B. The Survey Research Handbook. Homewood, IL: Dow Jones-Irwin, 1985. 429 p.
3. Carmines E.G., Zeller R.A. Reliability and Validity Assessment. Beverly Hills, California: SAGE Publications, 1979. 72 p.
4. Паниотто В. Качество социологической информации. Киев: Наукова думка, 1986. 206 с.
5. Trochim W. Types of Reliability. URL: <http://www.socialresearchmethods.net/kb/relytypes.php>
6. Mueller R. Basic Principles of Structural Equation Modeling. An Introduction to LISREL and EQS. New York: Springer-Verlag New York, Inc., 1996. 229 p.
7. Johnston J., Pennypacker H. Strategies and tactics of human behavioral research. New Jersey: Erlbaum, 1980. 210 p.
8. Дембицкий С. Относительно проблематичности статистической проверки теоретических конструкторов. *Социология: теория, методы, маркетинг*. 2014. № 2. С. 190–194.
9. Дембицкий С. Теоретическая валидизация в социологическом исследовании. Москва: ЛЕНАНД, 2016. 200 с.
10. Дембицкий С. Теоретическая валидизация измерительных шкал. *Социология: теория, методы, маркетинг*. 2012. № 2. С. 53–65.
11. Yin R. Case Study Research: Design and Methods. London; Thousand Oaks; New Delhi: Sage, 2009. 219 p.
12. Тилли Ч. Принуждение, капитал и европейские государства. 1990–1992 гг. Москва: Территория будущего, 2009. 328 с.
13. Андерсон Б. Воображаемые сообщества. Размышления об истоках и распространении национализма. Москва: Канон-пресс-Ц: Кучково поле, 2001. 288 с.
14. Derogatis L., Fitzpatrick M. The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18. *The use of Psychological Testing for Treatment Planning and Outcomes Assessment. Vol. 3: Instruments for Adults* / ed. by Mark E. Maruish. New Jersey; London: Lawrence Erlbaum Associated Publishers, 2004. P. 1–42.
15. Головаха Е., Панина Н. Интегральный индекс социального самочувствия (ИИСС): конструирование и применение социологического теста в массовых опросах. Киев: Ин-т социологии НАН Украины, 1997. 64 с.
16. Campbell D., Fiske D. Convergent and Diskriminant Validation by the Multitrait-multimethod Matrix. *Psychological Bulletin*. 1959. № 2. P. 81–105.
17. Волович В. Надежность информации в социологическом исследовании. Проблемы методологии и методики. Киев, 1976. 136 с.
18. Шмелев А. Основы психодиагностики. Ростов-на-Дону: Феникс, 1996. 544 с.

19. McNabb D. Nonsampling error in social surveys. Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE, 2014. 272 p.

20. Головаха Є., Горбачик А., Любива Т., Панина Н., Серета В., Урсуленко К. Суб'єктивна надійність: теорія і метод вимірювання (ІСН). *Соціологія: теорія, методи, маркетинг*. 2008. № 1. С. 166–188.

Отримано 25.07.17

References

1. DeVellis, R. (2012). *Scale Development: Theory and Applications*. Los Angeles; London; New Delhi; Singapore; Washington DC: SAGE [in English]

2. Alreck, P.L., Settle, R.B. (1985). *The Survey Research Handbook*. Homewood, IL: Dow Jones-Irwin [in English]

3. Carmines, E.G., Zeller, R.A. (1979). *Reliability and Validity Assessment*. Beverly Hills, California: SAGE Publications [in English]

4. Paniotto, V. (1986). Quality of sociological information. Kiev: Naukova dumka [in Russian]

5. Trochim, W. Types of Reliability. URL: <http://www.socialresearchmethods.net/kb/relytypes.php> [in English]

6. Mueller, R. (1996). *Basic Principles of Structural Equation Modeling. An Introduction to LISREL and EQS*. New York: Springer-Verlag New York, Inc. [in English]

7. Johnston, J., Pennypacker, H. (1980). *Strategies and tactics of human behavioral research*. New Jersey: Erlbaum [in English]

8. Dembitskii, S. (2014). To the problem of statistical verification of theoretical constructs. *Sotsiologiya: teoriya, metody, marketing* [Sociology: theory, methods, marketing], 2, 190–194 [in Russian]

9. Dembitskii, S. (2016). Theoretical validation in sociological research. Moscow: LENAND [in Russian]

10. Dembitskii, S. (2012). Theoretical validation of measurement scales. *Sotsiologiya: teoriya, metody, marketing* [Sociology: theory, methods, marketing], 2, 53–65 [in Russian]

11. Yin, R. (2009). *Case Study Research: Design and Methods*. London; Thousand Oaks; New Delhi: Sage [in English]

12. Tilli, Ch. (2009). *Coercion, Capital, and European States, AD 990–1990. 1990–1992*. Moscow: Territory of the Future [in Russian]

13. Anderson, B. (2001). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Moscow: Kanon-press-Ts: Kuchkovo pole [in Russian]

14. Derogatis, L., Fitzpatrick, M. (2004). The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18. In M.E. Maruish (Ed.), *The use of Psychological Testing for Treatment Planning and Outcomes Assessment. Vol. 3: Instruments for Adults* (pp. 1–42). New Jersey; London: Lawrence Erlbaum Associated Publishers [in English]

15. Golovakha, E., Panina, N. (1997). Integral index of social well-being (IISWB): design and application of socio-logical test in surveys. Kiev: Institute of Sociology NAS Ukraine [in Russian]

16. Campbell, D., Fiske, D. (1959). Convergent and Discriminant Validation by the Multitrait-multimethod Matrix. *Psychological Bulletin*, 2, 81–105 [in English]

17. Volovich, V. (1976). Reliability of information in a sociological research. Problems of methodology and techniques. Kiev [in Russian]

18. Shmelev, A. (1996). *Fundamentals of Psychodiagnostics*. Rostov-na-Donu: Feniks [in Russian]

19. McNabb, D. (2014). *Nonsampling error in social surveys*. Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE [in English]

20. Golovakha, Ye., Gorbachyk, A., Lyubyva, T., Panina, N., Sereta, V., Ursulenko, K. (2008). Subjective Reliability: Theory and Method of Measurement (ISR). *Sotsiologiya: teoriya, metody, marketing*, 1, 166–188 [in Ukrainian]

Received on 25.07.17

Дембицкий С.С., канд. социол. наук, старший научный сотрудник отдела методологии и методов социологии Института социологии НАН Украины, ул. Шелковичная, 12, Киев, 01021, Украина, e-mail: e_forge@me.com,

Любивая Т.Я., канд. социол. наук, научный сотрудник отдела методологии и методов социологии Института социологии НАН Украины, ул. Шелковичная, 12, Киев, 01021, Украина, e-mail: tlubivaya@gmail.com

ТРЕБОВАНИЯ К РАЗРАБОТКЕ И АДАПТАЦИИ КОМПЛЕКСНЫХ ИЗМЕРИТЕЛЬНЫХ ИНСТРУМЕНТОВ В СОЦИОЛОГИИ: НАДЕЖНОСТЬ, ВАЛИДНОСТЬ И ДОСТОВЕРНОСТЬ

Представлены главные характеристики качества социологического измерения и общая логика их проверки. Выявлено принципиально методологическое отличие между измерительными шкалами и индексами. Надежность концептуализирована как воспроизводимость измерения. Валидность рассмотрена в ее теоретической и эмпирической составляющих. Достоверность интерпретируется как возможность и (или) желание респондента сообщить интересующую исследователя информацию.

Ключевые слова: надежность, валидность, достоверность, социологическое измерение.

Dembitskyi S.S., PhD. (Sociology), senior scientific fellow, department of methodology and methods of sociology, Institute of Sociology National Academy of Sciences of Ukraine, 12, Shovkovychna Str., Kyiv, 01021, Ukraine, e-mail: e_forge@me.com,

Liubuya T.Ya., PhD. (Sociology), scientific fellow, department of methodology and methods of sociology, Institute of Sociology National Academy of Sciences of Ukraine, 12, Shovkovychna Str., Kyiv, 01021, Ukraine, e-mail: tlubivaya@gmail.com

REQUIREMENTS FOR DEVELOPMENT AND ADAPTATION OF COMPLEX MEASUREMENT TOOLS IN SOCIOLOGY: RELIABILITY, VALIDITY AND TRUSTWORTHINESS

The article dwells upon main characteristics of sociological research quality (reliability, validity and trustworthiness) as well as general logic of their testing. Complex measurement instruments divided into scales and indexes are in the focus. Scales aim to measure certain qualities that are not explicitly observed (latent variables) and determine characteristics, which underlie the indicators of the measurement technique. When using integral indices, latent variables or social phenomenon are determined by the factors based on which indicators are formulated. Reliability is viewed as repeatability of measurement or absence of random errors. Authors distinguish such approaches towards reliability testing as a test-retest (level of rigidity of the measurement in some time), parallel forms (level of equivalence of two forms of the same scale), and internal consistency (consistency of results of different indicators). In the most general perspective, validity is viewed as correspondence degree of the research process results to the reality. Integrated approach towards validation of measurement scales and integral indices includes two stages – theoretical and empirical validation. Theoretical validation implies reaching the sufficient level of correspondence of theoretical construct to phenomenon under study, while empirical validation implies testing the accordancy of measurement technique and results that were obtained while using it with the theoretical construct. Thus, theoretical validity is achieved first, and then empirical validity is tested. Trustworthiness of measurement is viewed as a respondent's capability and desire to report information in which the researcher is interested. Usage of reliable and valid measurement technique does not guarantee obtaining required data when it is not possible to assure trustworthiness of measurement. It is recommended to use index of subjective reliability to test trustworthiness. Recommendations regarding general logic of validity and reliability testing of complex measurement techniques and requirements towards design and sample size of the survey are presented in the article.

Keywords: reliability, validity, trustworthiness, sociological measurement.