

## **Соціологічні тести: сутність і валідизація**

### *Анотація*

*У статті здійснено спробу концептуалізації такого напрямку емпіричної соціології, як соціологічне тестування. Для цього виконується низка завдань. По-перше, з точки зору соціологічного тестування оцінюється релевантність чотирьох ключових теорій, що стосуються розроблення вимірювальних інструментів у соціальних науках, — класичної теорії тестів, математичної теорії вимірювання (Item Response Theory), теорії надійності й теорії валідності. По-друге, на підставі досвіду вітчизняної соціології окреслено сутнісні особливості соціологічного тестування (застосовність у масових опитуваннях, спрямованість на вивчення соціальних феноменів, діагностичний потенціал при аналізі соціальної системи). По-третє, розглядається специфіка валідизації соціологічних тестів, що включає три етапи: теоретична валідизація, вибір різновиду вимірювального інструменту, емпірична валідизація. На завершення статті розглянуто приклади соціологічних тестів: Інтегральний індекс соціального самопочуття (Є.Головаха, Н.Паніна), Типи політичної культури (Є.Головаха), SCL-9-NR (С.Дембіцький).*

**Ключові слова:** соціологічний тест, вимірювання, валідизація

У відомій “Енциклопедії вимірювання і статистики” за редакцією Н.Салкаїнда [Encyclopedia, 2007] термін “тест” зустрічається у назвах статей, що переважно належать до трьох тематичних розділів — поняття і проблеми у вимірюванні, статистичні техніки, статистичні тести. У пропонуваній статті соціологічні тести розглядаються першою чергою в контексті понять і проблем вимірювання психологічних, соціально-психологічних і соціальних феноменів.

Цікаво, що попри часте вживання терміна “тест” у назвах різноманітних статей згаданої енциклопедії, власне його визначення не наводиться. При цьому дається визначення шкалювання, під яким розуміють діяльність з метою вимірювання чи квантифікації психологічних настанов (атитюдів) або властивостей (атрибутів) [Parker, 2007: р. 862]. Саме вимірювання розглядається як із класичної (С.Стивенс), так і з сучасної точки зору. У першому випадку під ним розуміють “присвоєння числових значень об’єктам або подіям згідно із правилами”; при цьому вимірювання розглядається як процедура підготовки даних до статистичного аналізу. У другому випадку вимірювання і статистика вважаються фундаментально різними поняттями із власними допущеннями та результатами використання [Stemler, 2007: р. 582–84]. У рамках даної статті соціологічні тести розглядаються як вимірювальні інструменти, що задовольняють вимоги застосування в соціологічних опитуваннях.

Окреслюючи таку перспективу, слід коротко розглянути чотири ключові теорії, що стосуються розроблення вимірювальних інструментів у соціальних науках, — класичну теорію тестів, математичну теорію вимірювання<sup>1</sup> (Item Response Theory), теорію надійності та теорію валідності, щоб оцінити їхню релевантність у ракурсі соціологічного тестування.

*Класична теорія тестів.* З точки зору цієї теорії, значення, отримане в результаті використання тесту, включає два компоненти — істинне значення і помилку вимірювання. Істинне значення розглядається як середнє ідентичних значень, отриманих у результаті повторюваних без обмеження вимірювань. Таким чином, ключове допущення цієї теорії полягає в тому, що отримане в рамках тестування значення є сумою істинного значення і помилки. При цьому істинне значення завжди невідоме. Вимірювальні помилки — це будь-які аспекти вимірювання, крім істинного значення. Їхніми джерелами можуть бути добір пунктів методики, інструкція тесту, підрахунок результатів та будь-які систематичні помилки вимірювання. Перші три джерела ведуть до несподіваних і неузгоджуваних ефектів, тобто справляють несистематичний вплив на результати вимірювання і, відповідно, є випадковими помилками. Такі помилки не пов’язані як з істинним значенням, так і з помилками інших вимірювань такого стибу. Своєю чергою, систематичні помилки виникають у тих випадках, коли тест вимірює щось відмінне від того, що він має вимірювати (проблема валідності). Друге допущення стосується паралельних тестів. Воно полягає в тому, що кожен індивідуальний пункт методики сам по собі можна розглядати як тест. Це пов’язане з тим, що кожен із них є похідним від латентної змінної. При цьому вплив латентної змінної і величина випадкових помилок є однаковими в усіх випадках [Alexopoulos, 2007: р. 140–142].

Не відкидаючи загального фокусування теорії (істинне значення + помилка вимірювання), важко погодитися з тими акцентами, що їх зроблено в запитаннях про випадкові помилки. Систематичний зсув може бути внесений будь-яким елементом тесту, а не тільки його невалідністю. Сумнівним

<sup>1</sup> Цей варіант перекладу запозичено зі статті В.Аванесова [Аванесов, 2007].

виглядає і допущення про однаковий вплив латентної змінної на всі індикатори тесту, особливо коли йдеться не про вимірювальну шкалу, а про інтегральний індекс (це питання розглядається далі).

*Математична теорія вимірювання.* У рамках цієї теорії стверджується ймовірнісний характер взаємозв'язку між відповіддю на той чи інший пункт тесту і латентною змінною. Математична теорія вимірювання включає статистичні моделі оцінювання якості пунктів тесту. Найпоширенішими і найбільш популярними є моделі для оцінювання тестів, що призначені для вивчення одновимірних характеристик і застосовують пункти з дихотомічною шкалою. Більш комплексні моделі існують і для випадків, коли пункти можуть включати шкали відповідей із більш як двома категоріями і/або коли досліджувані характеристики є багатовимірними [Keller, 2007: р. 493–494].

Ця теорія виглядає прийнятнішою аналітичною схемою порівняно з класичною теорією тестів. Разом із тим, беручи до уваги акцент математичної теорії вимірювання на таких властивостях індикаторів тесту, як рівень підготовленості опитуваного, показник рівня складності завдань, здатність тесту відрізнити підготовлених респондентів від непідготовлених тощо [Аванесов, 2007], вона є слушною насамперед для розроблення педагогічних, а не соціологічних тестів.

*Теорія надійності.* Надійність вимірювання пов'язана з його узгодженістю в різноманітних контекстах. Назагал виокремлюють три аспекти надійності: 1) результати надійного інструменту не залежать від того, хто проводить дослідження; 2) використання того самого надійного інструменту в різний час дає однакові результати; 3) усі частини надійного інструменту є взаємозалежними. Відповідно, можна вирізнити три види надійності: надійність оцінювача або кодувальника (результати оцінювання різними експертами того самого об'єкта за посередництва того самого інструмента ведуть до однакових результатів), тест-ретестова надійність (зміни в результатах повторних вимірювань відповідають теоретичним очікуванням), внутрішня узгодженість (взаємопов'язаність пунктів тесту є високою) [Juni, 2007: р. 834].

Звісно, перевірка надійності є важливою умовою розроблення соціологічного тесту. При цьому актуальність того чи іншого різновиду надійності може істотно варіювати відповідно до тих завдань, для виконання яких розроблено конкретний тест.

*Теорія валідності.* Розглядаючи цю концептуальну царину з точки зору кількісних досліджень, можна (але лише з певною мірою умовності) говорити про єдину теорію. Якщо ж брати до уваги імплікації якісного і змішаного підходів, то має йтися про цілий набір теорій валідності. І оскільки, на мій погляд, теорія валідності в рамках кількісного підходу є суттєво лімітованою, далі приділено увагу власне проблемі валідації.

### ***Поняття соціологічного тесту***

Поняття “соціологічний тест” було введено в науковий обіг радянським, а нині російським, соціологом і педагогом В.Аванесовим у 1982 році. Він дав

соціологічному тесту два визначення. Згідно із першим, соціологічний тест — це “система відібраних соціологічними методами висловлювань, запропонованих респондентам з метою отримання надійної й валідної інформації про ознаки, що становлять інтерес”, відповідно до другого — “система висловлювань, що дають змогу одержати обґрунтоване відображення емпіричної системи, що становить інтерес, з відношеннями в числову систему з відношеннями” [Аванесов, 1982: с. 41]. У рамках цієї статті я дотримуватимуся першого визначення.

Двома аргументами В.Аванесова на користь виокремлення соціологічного тесту як самостійного методу дослідження є такі [Аванесов, 1982: с. 38–39]:

1) необхідність вивчення в соціологічних дослідженнях не тільки психологічних (для цього використовують психологічні тести), а й соціологічних феноменів<sup>1</sup>, таких як ціннісні орієнтації, світоглядні настанови, соціальні потреби тощо;

2) актуальність цього напряму досліджень на сучасному етапі розвитку методології соціальних наук.

На мою думку, навіть за більш як 30 років обидва ці аргументи не втратили своєї ваги. Соціологічні або соціальні феномени, як і раніше, перебувають у фокусі інтересу дослідників, а методологія розроблення й використання соціологічних тестів за ці роки потрапила в поле інтересів лише двох українських соціологів — Наталії Паніної та Євгена Головахи, котрі розробили низку соціологічних тестів<sup>2</sup>, що їх нині використовують у масових опитуваннях.

Н.Паніна та Є.Головаха поширили розуміння соціологічного тесту й на психологічні методики, спеціально адаптовані для використання в масових опитуваннях. І це цілком виправдано, адже більшість психологічних тестів є дуже обсяжними, що накладає істотні обмеження на використання їх у соціологічних дослідженнях. Крім того, Н.Паніна і Є.Головаха сформулювали низку вимог, які має задовольняти соціологічний тест [Головаха, 1997: с. 5]<sup>3</sup>:

- універсальність — методика має включати найзагальніші сутнісні індикатори досліджуваного явища, що відкриває можливість для використання її практично в будь-якому соціологічному дослідженні;
- інтегральність — можливість зведення різноманітних вимірювань соціологічного тесту до одного інтегрального показника;

<sup>1</sup> Далі я говоритиму про соціальні, а не соціологічні феномени.

<sup>2</sup> Є.Головаха і Н.Паніна розробили такі тести, як Інтегральний індекс соціального самопочуття і Типи політичної культури, а також адаптували низку закордонних методик з метою застосування їх у масових опитуваннях в Україні.

<sup>3</sup> Подані далі вимоги сформульовано в рамках розроблення Інтегрального індексу соціального самопочуття, але, по суті, їх можна розглядати як універсальні принципи соціологічного тестування.

- *якісність* — соціологічний тест має бути “паспортизований” статистичними показниками надійності та валідності;
- *стандартизованість* — встановлення “норм” досліджуваного явища, що дасть змогу інтерпретувати отримані результати з точки зору вираженості досліджуваного феномену;
- *компактність* — соціологічний тест має бути достатньо компактним, щоб його можна було використовувати в масових опитуваннях;
- *“чутливість”* — з огляду на вимогу компактності соціологічний тест має передбачати прийнятну точність вимірювання;
- *об’єктивність* — мета вимірювання має бути прихованою від респондента, що дає змогу уникнути фальсифікації результатів тестування.

Зважаючи на викладене вище стосовно сутності соціологічного тестування, можна виокремити два його виміри. Перший і принциповий стосується необхідності забезпечити методичні вимоги, що дадуть можливість повноцінно використовувати вимірювальну методика (не важливо — психологічну чи соціологічну) у рамках масових опитувань. Вже в цьому разі можна говорити про соціологічне тестування. Другий вимір пов’язаний із тими аспектами соціальної дійсності, на вивчення яких спрямоване використання соціологічного тесту. Методики, що відповідають цьому виміру, спрямовані на вивчення соціальних феноменів. У цьому разі підготовка концептуальної бази соціологічного тесту здійснюється переважно соціологічними засобами.

Крім зазначених, напевно, варто вирізнити і третій вимір соціологічного тестування, пов’язаний із потенціалом використання отриманих результатів. В ідеалі соціологічний тест має бути не просто розроблений як валідна й надійна методика, спрямована на вивчення соціального або психологічного феномену, а й бути пов’язаним із ширшими теоретичними положеннями про будову і функціонування соціальної системи. Завдяки цьому вможливується досягнення діагностичних висновків про стан її різних частин, можливі перспективи її функціонування та розвитку. Проте навіть якщо відповідні теоретичні положення недоступні, але при цьому методика задовольняє вимоги першого виміру соціологічного тестування, відповідні положення можна сформулювати в процесі використання методики в емпіричних дослідженнях.

Останнє, про що хотілося б сказати в цій частині статті, пов’язане зі світовою практикою використання соціологічних тестів. Насамперед слід зазначити, що такого терміна в зарубіжній соціології я не зустрів. Разом із тим, застосовуючи окреслені вище виміри соціологічного тестування, можна легко виявити, що низка використовуваних у західній соціології вимірювальних методик задовольняє вимоги соціологічного тестування. Наведу два загальновідомі приклади. Першим вимірювальним інструментом, котрий можна розглядати як соціологічний тест, є методика вивчення цінностей Ш.Шварца<sup>1</sup>. Вона має прийнятний для масових опитувань розмір, спря-

---

<sup>1</sup> Докладніше про методика Ш.Шварца див.: [Магун, 2007].

мована на вивчення соціального феномену, а результати її застосування повсюдно використовують для оцінювання стану різних суспільств. За другий приклад може правити відомий тест для оцінювання психологічного дистресу, розроблений Р.Кеслером (К6)<sup>1</sup>. Цей тест повсюдно використовують в епідеміологічних опитуваннях. Цікаво, що незважаючи на психологічну спрямованість, через свою стислість ця методика придатна саме в соціологічних, а не психологічних дослідженнях.

Тут у читача може виникнути питання, чи треба в такому разі<sup>2</sup> взагалі говорити про соціологічне тестування. Моя ствердна відповідь на нього пов'язана з тим, що специфіка соціологічного вимірювання повною мірою не охоплюється методологічними засобами сучасних вимірювальних підходів, які спираються переважно на психологію і педагогіку. Відповідно, специфічні для соціології проблеми розв'язуються здебільшого інтуїтивним шляхом, а таку ситуацію для методології емпіричних досліджень не можна назвати прийнятною.

### *Процес валідизації соціологічного тесту*

Тут я маю намір заторкнутися саме специфіку валідизації соціологічних тестів, а не численні техніки, котрі можна застосовувати при розробленні їх. Це пов'язане з тим, що до згаданих технік належать численні процедури формулювання й добору індикаторів, формування вибірки (від пілотних опитувань до репрезентативних досліджень з метою виокремлення норм), статистичного опрацювання результатів використання тесту на різних етапах його розроблення. Ці техніки надто численні й не можуть бути охоплені в рамках однієї статті. Крім того, чимало їх є специфічними й аналіз ситуацій їх застосування просто неможливий. Тому я виходжу з того, що читач ознайомлений із багатьма з них, а також має все необхідне для подальшого їх опанування<sup>3</sup>.

Комплексний підхід до процесу валідизації соціологічного тесту передбачає виокремлення трьох його етапів: теоретичної валідизації, вибору різновиду вимірювального інструменту, емпіричної валідизації.

*Теоретична валідизація.* Щоб уникнути проблеми “дурного собаки, який ганяється за своїм хвостом” (де “дурний собака” — це теорія, що її мають намір верифікувати за посередництва вимірювальних процедур, а “хвіст” — вимірювальна методика, побудована на базі ще не перевіреної теорії), в основу соціологічного тесту треба покласти вже валідні теоретичні положення. Така постановка питання пов'язана з тим, що “ламані” теоретичні конструкції цілком здатні приводити до прийнятних результатів статистичного аналізу. Задля підтвердження цього положення я здійснив ме-

<sup>1</sup> Докладніше про методику Р.Кеслера див.: [K10 and K6 Scales, s.a.].

<sup>2</sup> Це ситуація, коли соціологічні тести розробляють попри відсутність відповідного поняття і спеціальної концептуальної царини.

<sup>3</sup> Докладніше про відповідні техніки див., напр.: [DeVellis, 2012; Анастаси, 2007; Клайн, 1994].

тодичне дослідження, в рамках якого порівнювалися результати використання методики, побудованої на підставі вже валідних теоретичних положень, із результатами тієї самої методики, що була свідомо спотворена таким чином, щоб порушити її змістову валідність [Дембицький, 2014]. На підставі цих двох варіантів вимірювальної методики викладачі та старости одного з київських ВНЗ оцінювали поведінкові стратегії своїх студентів (чи одногрупників). Далі отримані результати були зіставлені із середніми балами студентів, що їх вони одержали на всіх сесіях бакалаврату (див. табл.). Як можна бачити, в плані статистики обидва варіанти рівнозначні, чого не можна сказати про теоретичні конструкти, які за ними стоять.

*Таблиця*

**Сила і напрям зв'язку результатів експертного оцінювання із середніми значеннями оцінок, отриманих на всіх сесіях бакалаврату ( $N = 23$ )**

Експерти	Сила зв'язку	
	Застосування "ламаной" 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**

\*  $p < 0,05$ ;

\*\*  $p < 0,01$ .

Звісно, з точки зору статистичного аналізу (наприклад, пошук взаємозв'язків з іншими конструктами або завбачення значень інших змінних) обидва варіанти шкали можна оцінити як рівноцінні. Але з позицій соціологічної інтерпретації, "ламана" шкала втрачає будь-яку цінність, адже відмова від змістових елементів веде до втрати інформації про два типи студентів із трьох<sup>1</sup>.

Наступне питання тут стосується того, якими є джерела тих теоретичних положень, котрі можна розглядати 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), тому 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, властиві представникам певної соціальної спільноти. Розгляньмо їх докладніше.

<sup>1</sup> З допомогою залишеної частини шкали всіх студентів можна віднести тільки до "байдужого" типу, тоді як "декларативно зацікавлений" і "відповідальний" типи випадають з фокусу дослідження.



Якщо говорити про наукові концепції, то мову слід вести про ті з них, побудова яких спиралася на дослідницьку стратегію кейс-стаді<sup>1</sup>. При цьому я залічую до таких і деякі концепції історичної макросоціології. До останніх, наприклад, належать результати, подані у працях Ч.Тили [Тилли, 2009] та Б.Андерсона [Андерсон, 2001]. Ці теорії хоч і видаються дуже далекими від рівня соціологічного вимірювання, але задають правильне фокусування щодо найбільш релевантних процесів у сьогодишньому світі національних держав, чим підвищують теоретичну чутливість дослідника.

Відмітною рисою проведення кейс-стаді з метою отримання валідних теоретичних положень є акцент на вивченні соціальних феноменів, що виявляються у соціальній дійсності<sup>2</sup>, з подальшим переходом до теоретичних положень, що потребують не верифікації, а інспектування і збагачення в рамках тих соціальних контекстів, що не були охоплені ініціальним дослідженням.

Що стосується попередніх комплексних досліджень, то йдеться, знову ж таки, про використання кейс-стаді. Відмінність лише в тому, що тут дослідження “загострене” під побудову необхідної для подальшого соціологічного вимірювання теоретичної бази. Відповідним прикладом є моє дослідження поведінкових практик студентів, котрі поєднують роботу з навчанням [Дембицкий, 2016: с. 99–124].

У разі професійного досвіду все значно простіше. Виконання щоденних функціональних обов’язків дає змогу фахівцям осягнути різні аспекти 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.Дерогатиса [Derogatis, 2004], котрий відштовхувався від тих симптомів різноманітних психологічних розладів, що були зафіксовані відповідною професійною спільнотою.

Зрештою, за відправну точку можуть правити й аксіоматичні наукові положення, валідність яких не викликає сумнівів. Так, це можуть бути твердження, розглядувані як базисні в ключових соціологічних теоріях (функціоналізм, символічний інтеракціонізм, теорія конфлікту, феноменологія, теорія обміну) або універсальні соціологічні аксіоми. Прикладом останніх є положення, використане Є.Головахою та Н.Паніною для конструювання Інтегрального індексу соціального самопочуття: “соціальне самопочуття людини визначається рівнем задоволеності її соціальних потреб, що, своєю чергою, є похідними від наявної в суспільстві системи соціальних благ, їх виробництва і розподілу” [Головаха, 1997: с. 23–24].

Розглянуті джерела можуть робити різний внесок у досягнення теоретичної валідності. Тому природно використовувати кілька з них у рамках одного дослідницького проекту.

<sup>1</sup> Докладніше див., напр.: [Yin, 2009; Дембицкий, 2016: с. 125–149].

<sup>2</sup> Під “соціальною дійсністю” я розумію людські вчинки на різних рівнях соціальної взаємодії, спрямовані на створення й підтримку соціальних інститутів.



Коли говорять безпосередньо про теоретичну валідність, варто розуміти, що теоретична база дослідження має забезпечувати три її компоненти: очевидний, змістовий і композиційний. Очевидна валідність пов'язана з аргументацією, в рамках якої досліджуваний феномен вписується в соціальну дійсність і, відповідно, не є “витвором” чийось домислів; змістова валідність забезпечує значимість і репрезентативність елементів теоретичного конструкта; композиційна — слухність його внутрішньої структури [Дембіцький, 2016: с. 22–31]. У такому вигляді теоретична база надає все необхідне для подальшої операціоналізації теоретичних положень, а також оцінювання отриманих у рамках емпіричної валідації результатів. Але перш ніж переходити до останньої, необхідно визначитися зі специфікою вимірювального інструменту.

*Вибір різновиду вимірювального інструменту.* Після утвердження валідної теоретичної бази специфіка розроблення вимірювального інструменту зумовлюється вибором його різновиду. На підставі праць С.Садмена і Н.Бредберна [Садмен, 2002] та Р.ДеВелліса [DeVellis, 2012], до різновидів вимірювальних інструментів у соціальних науках я відношу такі:

- а) анкетні запитання різного рівня складності (предмет вивчення — думки, поведінка, знання);
- б) вимірювальні шкали (предмет вивчення — латентні змінні);
- в) інтегральні індекси (предмет вивчення — латентні змінні, соціальні феномени).

Спинятися на тому, що таке анкетні запитання, вважаю зайвим. Якщо ж говорити про відмінність між вимірювальними шкалами й інтегральними індексами, то в першому випадку йдеться про вимірювання деякої безпосередньо не спостережуваної властивості, що зумовлює аспекти її прояву, закладені в основу індикаторів методики; натомість у разі інтегральних індексів латентна змінна або соціальний феномен зумовлюються тими чинниками, на підставі яких сформульовані індикатори. Принциповим методичним результатом такого поділу є нерелевантність внутрішньої надійності в рамках обґрунтування якості інтегральних індексів. У цьому випадку між різними причинними факторами може бути зв'язок, а може й не бути.

*Емпірична валідація.* В ідеалі на підставі результатів емпіричного дослідження або їх серій перевіряють конструктну, передбачувальну, критеріальну та конкуретну валідність.

Конструктну валідність я розглядаю в термінах, запропонованих у праці Д.Кемпбела і Д.Фіске [Campbell, 1959]. На рівні простих анкетних запитань, а також окремих індикаторів вимірювальних шкал та інтегральних індексів її перевірка здійснюється шляхом порівняльного аналізу розподілу відповідей і базових описових статистик, отриманих при використанні різноманітних формулювань і шкал відповідей. На рівні підсумкового інструменту використовується кореляційний аналіз (у широкому сенсі), якщо йдеться про вимірювальну шкалу (з увагою до внутрішньої узгодженості) або інтегральний індекс (без уваги до внутрішньої узгодженості). Якщо ж апробується соціологічна анкета, акцент робиться на пілотних опитуваннях із

метою перевірки контекстуальної зумовленості відповідей при використанні різних послідовностей тих самих запитань.

Критеріальна валідність<sup>1</sup> ґрунтується на отриманні очікуваних відмінностей в перебігу опитування “контрастних” груп і, таким чином, показує загальну “чутливість” вимірювального інструментарію.

Передбачувальна валідність пов’язана з можливістю методики надавати інформацію, необхідну для прогнозування майбутніх станів об’єкта дослідження. З точки зору соціологічного тестування, це, можливо, найбільш важкодосяжний вид валідності — як внаслідок необхідної стислості інструментарію, так і з огляду на комплексність досліджуваних у соціології процесів та явищ. У зв’язку з цим результати соціологічного тестування — лише один із можливих джерел емпіричної інформації, на підставі якої будують соціологічний прогноз.

Конкурентну валідність розуміють як порівняльну перевагу вимірювального інструменту серед низки схожих методик. Таким чином, її обґрунтування пов’язане з необхідністю порівняння показників конструктивної, критеріальної і передбачувальної валідності різноманітних соціологічних тестів, спрямованих на вивчення одного явища.

### ***Приклади соціологічних тестів: соціальне самопочуття, політична культура, психологічний дистрес***

Далі аналізуються три соціологічні тести, конструювання яких так чи інакше пов’язане з іменами відомих київських соціологів Є.Головаха та Н.Паніної. Для кожного тесту стисло розглянуто його призначення, особливості конструювання, статистичної перевірки і валідизації, а також соціологічна специфіка в розрізі трьох описаних раніше вимірів соціологічного тестування (застосовність у масових опитуваннях, спрямованість на вивчення соціальних феноменів, діагностичний потенціал, з точки зору аналізу соціальної системи).

*Соціологічний тест “Інтегральний індекс соціального самопочуття”* (далі — ІСС). ІСС (автори — Є.Головаха, Н.Паніна<sup>2</sup>) являє собою стандартизовану тестову процедуру, що була розроблена з огляду на необхідність оцінювання масових емоційно-оцінних станів у постійно змінюваних умовах нестабільного перехідного суспільства [Головаха, 1997: с. 11]. Соціальне самопочуття автори методики розуміють як “емоційно-оцінне ставлення індивіда до системи соціальних відносин і до свого місця в цій системі” [Головаха, 1997: с. 20].

При конструюванні тесту було виокремлено 11 сфер соціальної життєдіяльності індивіда: соціальних відносин, соціальної безпеки, національних відносин, соціально-політична, професійно-трудова, інформаційно-

<sup>1</sup> Критеріальна, передбачувальна і конкурентна валідності інтерпретуються мною доволі вільно, що пов’язано із авторським розумінням специфіки соціологічних досліджень.

<sup>2</sup> Статистичний аналіз здійснено А.Горбачиком.

культурна, рекреаційно-культурна, матеріально-побутова (першого і другого рівня), міжособистісних відносин, особистісна. У середині кожної сфери було вирізнено по 4 індикатори, що дало в результаті 44 пункти методики, які описують соціальні блага. Така деталізація забезпечила змістову валідність, обґрунтування якої було здійснене на логічному рівні.

Як альтернатива відповіді була сформульована так звана “шкала достатності”: *не вистачає; важко сказати, вистачає чи ні; вистачає; не цікавить*. З метою статистичного опрацювання вирішили прирівняти четвертий пункт шкали (не цікавить) до другого (важко сказати, вистачає чи ні) [Головаха, 1997: с. 21–24].

На підставі результатів опитування ( $N = 1810$ ), репрезентативного для дорослого населення України (віком від 18 років), проведено оцінювання валідності й надійності методики загалом, проаналізовано використання кожного із запропонованих у початковому варіанті індикаторів соціального самопочуття, побудовано стислий варіант тесту, здійснено стандартизацію і нормалізацію шкали [Головаха, 1997: с. 27–28].

На підставі результатів кореляційного аналізу кожного пункту тесту із сумарним індексом, оцінювання впливу виключення окремих пунктів на внутрішню узгодженість і перевірки нормальності розподілу сумарного індексу всі 44 пункти методики були включені до фінального варіанта. Успішну перевірку пройшли такі види надійності, як розщеплення тесту навпіл, ретестова і внутрішня узгодженість.

Також було перевірено конструктну валідність на підставі зіставлення сумарного індексу для різних категорій респондентів, виокремлених залежно від відповідей на запитання про різні аспекти життєвої задоволеності [Головаха, 1997: с. 40–43].

Скорочений варіант (20 із 44 пунктів) був побудований на базі формальних (збереження внутрішньої узгодженості) і змістових засад (зберігання в основних рисах початкової багатоаспектності вимірювань). У результаті було отримано вимірювальний інструмент, що відповідає стандартам надійності і валідності [Головаха, 1997: с. 46–47].

Проаналізуймо соціологічну специфіку тесту. ПСС розроблявся й оцінювався передусім як інструмент соціологічного вимірювання для масових опитувань. Тому він задовольняє принципову вимогу соціологічного тестування – застосовності у масових опитуваннях. Сам феномен соціального самопочуття включає як психологічні, так і соціальні компоненти, зі значним переважанням останніх. Отже, показники ПСС залежать насамперед від специфіки суспільних процесів, пов'язаних із задоволенням потреб індивідів. З одного боку, це є аргументом на користь діагностичного потенціалу тесту при аналізі соціальної системи нашого суспільства. Водночас тут залишається низка нез'ясованих питань, серед яких, наприклад, урахування зміни актуальності різноманітних потреб у контексті масштабніших соціальних процесів (так, війна на Сході України мала позначитися на актуальності сфер соціальної безпеки і національних відносин) і розроблення критеріїв оцінювання успішності різних соціальних інститутів, відповідальних за надання основних соціальних благ у нашому суспільстві (хоча

потенціал тесту повною мірою дає підстави сподіватися на реалізацію цих завдань).

Узагальнюючи подану інформацію, доходимо висновку, що ПСС є повноцінним соціологічним тестом, який задовольняє високі стандарти щодо розроблення вимірювальних інструментів у соціальних науках.

*Соціологічний тест “Типи політичної культури”* (далі – ТПК). ТПК (автор – Є.Головаха<sup>1</sup>) є інтегральним тестом, призначеним для визначення типу політичної культури респондента. При цьому під політичною культурою передбачають “сукупність усталених форм політичної свідомості і поведінки людей, що визначають особливості функціонування різних суб’єктів політичного процесу в рамках певної політичної системи” [Щербак, 2006: с. 155].

При деталізації теоретичного конструкта було виокремлено два виміри політичної культури: демократія / тоталітаризм, активність / пасивність (у сенсі громадянської участі). Для кожного з вимірів було сформульовано по 50 тверджень, при цьому для збалансованості було розроблено по 25 тверджень для демократії і тоталітаризму, а також для активності і пасивності.

Для реєстрації відповідей використано шкалу Лайкерта: “цілком згоден”, “радше згоден”, “важко сказати, згоден чи ні”, “радше не згоден”, “абсолютно не згоден”.

З метою скорочення методики до прийняттого для соціологічних опитувань розміру, перевірки конструктивної валідності та ретестової надійності було проведено пілотне опитування респондентів віком від 18 років ( $N = 130$ ). Для оцінювання ретестової надійності кожного пункту використовувався як кореляційний аналіз (надійним вважався пункт, що показував силу зв’язку не менше 0,3), так і розрахунок шкальних оцінок за формулою середньої арифметичної похибки (не більш як 1).

Далі на підставі факторного аналізу було визначено судження, що дають однофакторне рішення на шкалах демократія / тоталітаризм і активність / пасивність. Також була оцінена внутрішня надійність ( $\alpha$  Кронбаха) і сила зв’язку між окремими індикаторами й адитивним індексом. З огляду на всі отримані результати з вихідного набору суджень було відібрано по вісім для кожної зі шкал.

Цей проміжний варіант був включений в опитування в рамках “Європейського соціального дослідження”. Подальша апробація методики також проводилася із використанням перевірки внутрішньої узгодженості та факторної структури. Для обох шкал були отримані двофакторні рішення, із поділом на шість і два індикатори. Після видалення чотирьох тверджень, що не входили до числа основних факторів, був отриманий фінальний варіант тесту, що включав по шість суджень для шкал демократії / авторитаризму й активності / пасивності.

Фінальний варіант ТПК було використано для вирізнення типів політичної культури населення України із використанням кластерного аналізу методом *k-середніх*. Виходячи з теоретичних очікувань наявності чо-

<sup>1</sup> Статистичний аналіз здійснено А.Горбачиком.

тирьох типів політичної культури (демократична активна, демократична пасивна, авторитарна активна, авторитарна пасивна), було побудовано чотири кластери, що підтвердили попередні теоретичні припущення [Стегній, 2015: с. 29–32].

Розгляньмо тест з точки зору соціології. Як і в попередньому випадку, ТПК був спроектований і реалізований як соціологічний тест. У плані стислості він навіть перевершує ПСС, що забезпечує його відповідність першому виміру соціологічного тестування. Дві концептуальні дихотомії, що лежать в основі цього інструменту, є цілком соціальними, що підтверджує його спрямованість на вивчення саме соціальних феноменів. Також немає жодних сумнівів, що ТПК вимірює результативні ефекти інституціональної динаміки і демонструє потенціал демократичних процесів у суспільстві. Відповідно, його результати найвищою мірою застосовні для оцінювання стану соціальної системи.

При цьому я хочу наголосити один аспект, що викликає питання. Так, усі шість суджень для шкали демократія / авторитаризм сформульовані таким чином, що стосуються саме демократичних настанов. Отже, питання щодо того, якою мірою за ними можна судити про авторитарні настанови, залишається відкритим. Особливо актуальним це питання є з огляду на кризові процеси в сучасній Україні, що збільшують ймовірність розвитку авторитарних поглядів серед населення нашої країни.

Утім, тест ТПК є високоякісним соціологічним інструментом, що базується на теоретичних положеннях, які не підлягають сумніву, та емпіричних індикаторах, які пройшли багаторівневу процедуру добору, що робить його постійне використання в соціологічних опитуваннях посправжньому актуальним.

*Соціологічний тест “SCL-9-NR”.* SCL-9-NR (автор – С.Дембіцький) є модифікацією відомого психологічного тесту SCL-90-R<sup>1</sup>, розробленого Л.Дерогатисом [Derogatis, 2004]. Відмітними рисами SCL-9-NR порівняно із SCL-90-R є його стислість (для кожного із дев'яти симптоматичних вимірів залишено тільки один індикатор), а також принципові зміни в завданні методики і шкалі відповідей (але не у формулюваннях самих індикаторів). Головне призначення тесту – оцінювання психологічного дистресу респондентів у рамках масових опитувань.

Оригінальна методика, запропонована Л.Дерогатисом, базується на таких симптоматичних вимірах, як ворожість, obsесивно-компульсивний розлад, депресія, міжособистісна сензитивність, параноїдальні ідеї, тривожність, соматизація, фобійна тривожність, психотизм, кожен з яких включає набір індикаторів (від шести до тринадцяти пунктів). При доборі індикаторів для кожного симптоматичного виміру враховувалася сила його зв'язку з індексом своєї підшкали, а також з інтегральним індексом SCL-90-N [Дембіцький, 2016а: с. 54–55].

---

<sup>1</sup> З метою адаптації методики в Україні SCL-90-R було включено Н.Паніною до моніторингових опитувань Інституту соціології НАН України у 1997 і 1999 роках.

Зміна формулювання тестового завдання і, відповідно, самої шкали відповідей<sup>1</sup> здійснена на підставі рекомендацій, наведених у відомій праці С.Садмена і Н.Бредберна з розроблення соціологічного інструментарію [Садмен, 2002: с. 65–96]. Головною причиною такого коригування послугував передбачуваний зсув у відповідях респондентів, що виникає при використанні оригінального варіанта формулювання: респонденти були схильні обирати менші значення на шкалі відповідей, що зменшувало і фінальний рівень індексу психологічного дистресу. Як показала подальша перевірка, таке припущення виявилось слушним, і використання нових формулювань суттєво поліпшило фінальний розподіл індексу [Дембицкий, 2016а: с. 59–61].

На підставі онлайн-опитувань у Києві ( $N = 200$ ) та Львові ( $N = 200$ ) була перевірена внутрішня надійність і факторна валідність методики. У подальшому, оскільки тест спрямований на вивчення психологічного дистресу, було проведено опитування в Чернігівській обласній психоневрологічній лікарні ( $N = 181$ ) із метою перевірки критеріальної валідності та впливу ефекту соціальної бажаності. Для оцінювання останнього на звичайній вибірці спільно з кафедрою методології і методів соціологічних досліджень КНУ ім. Тараса Шевченка було проведено онлайн-опитування серед користувачів соціальних мереж ( $N = 367$ ). Отримані результати підтвердили як критеріальну валідність тесту, так і прийнятний рівень впливу ефекту соціальної бажаності на його результати. Крім цього, на підставі опитування в психоневрологічній лікарні виокремлено значення тесту, що характеризують підвищений і високий дистрес [Дембицкий, 2016б].

З метою перевірки конструктивної та конкурентної валідності, а також виокремлення норм тесту на підставі результатів репрезентативного для дорослого населення України опитування SCL-9-NR був включений у моніторингове дослідження “Українське суспільство” у 2016 році<sup>2</sup>. На момент написання статті відповідні результати ще не були доступні для аналізу.

Оцінюючи методику як соціологічний тест, слід визнати, що як таку її можна розглядати суто в контексті першого виміру соціологічного тестування — його застосовності в соціологічних опитуваннях. Власне, на поточний момент основна робота з розроблення методики здійснена саме в цьому напрямі. І хоча це найстисліший з-поміж розглядуваних у статті соціологічних тестів, він дещо поступається згаданому раніше тесту психологічного дистресу Р.Кеслера (К6). Однак SCL-9-NR базується на дев'яти симптоматичних вимірах, тоді як К6 тільки на двох — депресії та тривожності.

Сам психологічний дистрес не можна назвати соціальним феноменом, проте можна припустити його залежність від низки соціальних характеристик, що потенційно робить SCL-9-NR інструментом аналізу соціальної сис-

1 Було використано шкалу: “ніколи”, “іноколи”, “періодично”, “майже постійно”.

2 Організатором дослідження виступає Інститут соціології НАН України. Результати опитування репрезентативні для населення України віком від 18 років (вибірка — не менш як 1800 осіб). Анкета опитування містить понад 300 запитань і охоплює різноманітні аспекти життя українського суспільства.



теми. Для перевірки цього припущення в подальшому здійснюватиметься додаткова дослідницька робота.

### *Стислі висновки*

З огляду на досвід вітчизняної та світової емпіричної соціології можна сказати, що кожен із трьох вимірів соціологічного тестування є цариною наукової практики, що включає силу-силенну наукових проблем, які потребують розв'язання.

Перший вимір (застосовність методики в масових опитуваннях) пов'язаний переважно з методичними аспектами розроблення методики, що передбачає технічну поінформованість вченого й організаційні навички, необхідні для проведення методичних досліджень. Разом із тим коли йдеться про другий вимір — вивчення соціальних феноменів, то технічні й організаційні навички мають доповнюватися надійними теоретичними знаннями і соціологічною уявою. Зрештою, якщо задіяні всі три виміри, тоді слід також додати достатні ресурси — як часові, так і фінансові, адже результати використання соціологічного тестування мають бути ретельно досліджені у прив'язці до відповідного соціального контексту. Усі ці аспекти роблять соціологічне тестування більш комплексною науковою практикою порівняно із психологічним.

Відповідно, методологія соціологічного вимірювання не мусить залежати від психологічного, нехай навіть останнє й має непохитний авторитет і є ядром вимірювального підходу в соціальних науках. Вихідним пунктом соціологічного тестування має бути досліджуваний феномен, а також той соціальний контекст, в якому він вивчається. Що стосується всієї сукупності підходів і технік, використовуваних психологами, то їй доцільно розглядати суто як доступний, але не визначальний інструментарій.

Ця вимога підтверджується й тим, що всі п'ять згаданих у статті соціологічних тестів розроблено із застосуванням дуже різних підходів і технік. Таким чином, за безумовної уваги до останніх ключовим показником якості соціологічного тесту є його ефективність для вивчення соціальної дійсності у фундаментальних, прикладних і оцінних дослідженнях: право ключового голосу належить прагматичним міркуванням.

### *Джерела*

*Аванесов В.* Истоки и основные понятия математической теории измерения (Item Response Theory) / В. Аванесов // Педагогические измерения. — 2007. — № 3. — С. 3–36.

*Аванесов В.* Тесты в социологическом исследовании / В. Аванесов. — М.: Наука, 1982. — 200 с.

*Анастаси А.* Психологическое тестирование / А. Анастаси, С. Урбина. — СПб.: Питер, 2007. — 688 с.

*Андерсон Б.* Воображаемые сообщества. Размышления об истоках и распространении национализма / Б. Андерсон. — М.: Канон-пресс-Ц: Кучково поле, 2001. — 288 с.

*Головаха Е.* Интегральный индекс социального самочувствия (ИИСС): конструирование и применение социологического теста в массовых опросах / Е. Головаха, Н. Панина. — К.: Ин-т социологии НАН Украины, 1997. — 64 с.



*Дембицкий С.* Критериальная валідація експресс-теста SCL-9-NR / С. Дембицкий // Социология: теория, методы, маркетинг. — 2016. — № 2. — С. 102–116.

*Дембицкий С.* Относительно проблематичности статистической проверки теоретических конструкторов / С. Дембицкий // Социология: теория, методы, маркетинг. — 2014. — № 2. — С. 190–194.

*Дембицкий С.* Теоретическая валідація в социологическом исследовании / Дембицкий С. — М. : ЛЕНАНД, 2016. — 200 с.

*Дембицкий С.* Экспресс-тест SCL-9-NR: методика оценки выраженности психологического дистресса для массовых опросов / С. Дембицкий // Социология: теория, методы, маркетинг. — 2016. — № 1. — С. 52–64.

*Клайн П.* Справочное руководство по конструированию тестов: Введение в психометрическое проектирование / Клайн П. — К. : ПАН ЛТД, 1994. — 283 с.

*Магун В.* Жизненные ценности населения: сравнение Украины с другими европейскими странами / В. Магун, М. Руднев // Украинское общество в европейском пространстве / под ред. Е. Головахи, С. Макеева. — К. : Ин-т социологии НАН Украины : Харьков. нац. ун-т им. В.Н. Каразина, 2007. — С. 226–273.

*Садмен С.* Как правильно задавать вопросы / С. Садмен, Н. Брэдберн. — М. : Ин-т Фонда “Общественное мнение”, 2002. — 382 с.

*Стегній О.* Соціологічні підходи до вивчення типів політичної культури / О. Стегній // Український соціум. — 2015. — № 2. — С. 26–35.

*Тилли Ч.* Принуждение, капитал и европейские государства. 1990–1992 гг. / Тилли Ч. — М. : Территория будущего, 2009. — 328 с.

*Щербак М.* Социально-классовые характеристики политической культуры в Украине: опыт применения социологического теста “Типы политической культуры” / М. Щербак // Социология: теория, методы, маркетинг. — 2006. — № 3. — С. 153–160.

*Alexopoulos D.* Classical Test Theory / D. Alexopoulos // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 140–143.

*Campbell D.* Convergent and Diskriminant Validation by the Multitrait-multimethod Matrix / Donald Campbell, Donald Fiske // Psychological Bulletin. — 1959. — № 2. — P. 81–105.

*Derogatis L.* The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18 / L. Derogatis, M. Fitzpatrick // 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.

*DeVellis R.* Scale Development: Theory and Applications / DeVellis R. — Los Angeles ; London ; New Delhi ; Singapore ; Washington DC : SAGE, 2012. — 216 p.

Encyclopedia of measurement and statistics / ed. by N. Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007.

*Juni S.* Reliability Theory / S. Juni // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 834–835.

K10 and K6 Scales [Electronic resource] — Access mode :  
[http://www.hcp.med.harvard.edu/ncs/k6\\_scales.php](http://www.hcp.med.harvard.edu/ncs/k6_scales.php).

*Keller L.* Item Response Theory / L. Keller // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 493–496.

*Parker S.* Scaling / S. Parker // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 582–584.

*Stemler S.* Measurement / S. Stemler, D. Birney // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 862–865.

*Yin R.* Case Study Research: Design and Methods / Yin R. — London ; Thousand Oaks ; New Delhi : Sage, 2009. — 219 p.