

ТЕТЯНА НІКІТИНА,

соціолог відділу теорії, історії та методології соціології Інституту соціології НАНУ, аспірантка Київського національного університету ім. Тараса Шевченка

Теоретична та емпірична інтерпретація і операціоналізація поняття “валідність”¹

Abstract

Validity is one of the basic properties of empirical measurement. A common definition of validity is the extent to which the survey measure accurately reflects the intended construct. But in different sociological source of information there are different understandings of the domain of validity. Some researchers understand validity as a property of measurement, while others – as a property of measurement instrument. Researchers determine different types of validity. There is no standardization of notions. Such situation complicates the process of validation of particular measurement. In fact, there is no concrete pattern of measurements validation. The purpose of this article is to make theoretical and empirical interpretation and operationalization of the validity notion and as a result develop the concrete pattern of measurements validation.

Питання якості соціологічного вимірювання та соціологічної інформації як результату такого вимірювання є одним із найважливіших для емпіричної соціології. Без доведення якості соціологічного вимірювання ми не можемо бути впевненими у змістовних висновках, які робимо на основі даних, отриманих у перебігу вимірювання. В.Паніотто у праці “Якість соціологічної інформації” писав: “Претензії соціології на участь у розв’язанні реальних практичних задач, що стоять перед нашим суспільством, можуть

¹ Ця стаття була б неможливою без стипендії університету Констанц і допомоги Агнежки Войти, завдяки чому тут цитовано більшість першоджерел. Вдячна усім, хто погодився прочитати і прокоментувати ранні версії даної статті. Особлива подяка моему науковому керівнику Андрію Петровичу Горбачику, а також Ксенії Урсуленко, Тетяні Любивій і Олексію Сергійовичу Середі.

бути обґрунтованими тільки у тому разі, якщо рекомендації соціологів базуватимуться на надійній (насамперед валідній) соціологічній інформації. Підвищення надійності соціологічної інформації — важливий чинник підвищення ефективності соціологічної науки та підвищення її практичної віддачі” [Паниотто, 1986: с. 158]. Валідність є однією з характеристик якості соціологічного вимірювання. У найзагальнішому вигляді, **валідність — це міра того, наскільки ми вимірюємо те, що заплановано виміряти** [Bohrnstedt, 2000: с. 3207].

У соціологічній літературі існують різні погляди на те, що власне характеризує валідність — соціологічне вимірювання в цілому чи вимірювальний інструмент. Дослідники виділяють різні види валідності. Стандартизація понять відсутня. Різняться і погляди на співвідношення таких ключових показників якості соціологічного вимірювання, як валідність і надійність. Різні відповіді на ці питання диктують різні схеми перевірки валідності.

Ця ситуація термінологічної невизначеності, “розмитості меж” поняття валідності ускладнює перевірку валідності конкретного вимірювання. Отже, необхідно розробити чітку схему такої перевірки, що передбачає теоретичну й емпіричну інтерпретацію та операціоналізацію поняття “валідність”, тобто визначення його меж та змісту й окреслення методів її перевірки.

Теоретична інтерпретація поняття “валідність”

Сьогодні в українській соціології паралельно існують дві традиції оцінювання якості соціологічного вимірювання в цілому та оцінювання валідності — радянська та західна, які відрізняються термінологічно, але за глибшого аналізу стає зрозуміло, що за суттю вони дуже схожі.

Коротко окреслюю, що маю на увазі, говорячи про “радянську” і про “західну” традиції.

Під західною традицією оцінювання якості соціологічного вимірювання я розумію підходи американських та західноєвропейських авторів починаючи від 30-х років ХХ століття і до сьогодні. Об’єднати їхні праці в одну загальну традицію дає змогу їхня наступність і спільність у принципових питаннях.

Західна традиція бере початок із психологічних тестів, із праць з оцінювання якості цих тестів. До перших досліджень надійності та валідності належать розвідки Л.Терстоуна “Надійність та валідність тесту” [Thurstone, 1931] та П.Гілфорда “Психометричні методи” [Guilford, 1936]. Пізніше з’явилися статті у новоствореному журналі “Освітні та психологічні вимірювання” (Educational and Psychological Measurement). У цих двох книжках та журнальних статтях формально було визначено поняття “валідність”. Тут його пов’язували з кореляцією даних із певним критерієм. У сучасній термінології такий вид валідності називають валідністю за критерієм.

Наприкінці 1940-х — на початку 1950-х років у науковий обіг був введений ще один вид валідності — валідність за змістом, або змістова валідність (content validity). Визначення змістової валідності було наведено Е.Куретоном 1951 року у виданні “Вимірювання у сфері освіти” (Educational Measurement) [Cureton, 1951].

1955 року Л.Кронбах та П.Мел опублікували статтю під назвою “Конструктна валідність у психологічних тестах” [Cronbach, Meehl, 1955]. З цієї статті у науковий обіг увійшов ще один вид валідності — конструктна валідність. У 1959 році Д.Кемпбел та Д.Фіске розробили стратегію перевірки конструктної валідності, виокремивши два підвиди конструктної валідності — конвергентну та дискримінантну. Вони проводили процедуру перевірки конструктної валідності з використанням мультиознакової-мульти-методової матриці (МТММ)¹.

У 1954 році Американська психологічна асоціація видала перші стандарти психологічного тестування, в яких було дано визначення валідності та надійності й видів валідності та надійності [Technical recommendations, 1954]. Далі ці стандарти перевидавали 1966, 1974, 1985 та 1999 років [Standards, 1966; 1974; 1985; 1999]. На ці стандарти і досі посилаються як психологи, так і соціологи. На підставі Стандартів вирізняють три основні види валідності: критеріальну (яка має два підвиди: прогностичну та конкурентну), змістову і конструктну.

У 1979 році у популярній серії “Університетські видання “Сейдж” вийшла друком книжка “Оцінювання надійності та валідності” Дж.Кармінеса та А.Зелера. Автори її теж вирізняють три види валідності, як і в Стандартах. Більшість сучасних західних авторів працюють у рідчизні зазначених вище класичних праць, розширюючи та поглиблюючи головні напрацювання класиків.

Радянська традиція оцінювання якості соціологічного вимірювання ґрунтована на базових принципах та поняттях метрології — науки про вимірювання, методи та засоби забезпечення необхідного рівня його якості. Метрологія спирається на методологічні засади вимірювання та розв’язує такі завдання, як: створення еталонів, мір та інструментів вимірювання; встановлення засобів оцінювання точності та правильності вимірювань [Докторов, 1979: с. 7]. Більшість радянських авторів посилалися на метрологічну літературу. До радянської традиції оцінки якості соціологічного вимірювання належать праці таких авторів, як А.Здравомислов (1969), В.Ядов (1972), В.Волович (1974), Б.Докторов (1979), Г.Саганенко (1983), В.Паніотто (1986) та ін. У переважній більшості випадків радянські автори використовували термін “обґрунтованість” замість терміна “валідність” (за винятком Паніотто, який використовує термін “валідність”). Але за змістом валідність та обґрунтованість у цих розвідках означають приблизно те саме. У радянській традиції, на противагу західній, не приділяли великої уваги розгляду окремих видів валідності (обґрунтованості) (знову-таки за винятком праці Паніотто, де вирізняє теоретичну та емпіричну валідність).

Не можна говорити про пряму наступність західної і радянської традицій, оскільки вони виходили з різних джерел (психологічне тестування та метрологія) і розвивалися більш-менш автономно, але і про абсолютну ізоляваність радянських авторів від західних не можна говорити.

Що стосується сучасної української літератури, то тут зустрічаються посилання як на радянських, так і на західних авторів. У зв’язку з цим спостерігається певна термінологічна плутанина.

¹ З цим методом можна ознайомитись у статті: [Campbell, Fiske, 1959].

Спробую співвіднести терміни “радянської традиції” з термінами “західної традиції”. Знаходження термінологічної відповідності цих двох традицій — не самоціль, це тільки один крок на шляху до розуміння суті поняття валідності та окреслення конкретних методів її перевірки. Слід зазначити, що, як радянська, так і західна традиції розробили досить ефективні способи оцінювання якості вимірювання, і не можна стверджувати, що якась із цих традицій є кращою чи ефективнішою. Але сьогодні, коли Україна інтегрується у світовий соціологічний простір, бере участь у міжнародних порівняльних дослідженнях (наприклад, Європейське соціальне дослідження (European Social Survey — ESS) чи Міжнародній програмі соціального дослідження (International Social Survey Programm — ISSP), більш адекватним є використання підходів саме західної традиції, оскільки вони вже стали стандартними для багатьох країн і їх використання полегшує міжнародну дослідницьку комунікацію, адже “без уніфікації термінології ускладнюється взаєморозуміння між дослідниками” [Паниотто, 1986: с. 7]. Тим паче, що така тенденція вже окреслилася. Прикладом використання термінології західної традиції оцінювання якості соціологічного вимірювання є праці Н.Паніної [напр.: Паніна, 1996], а також останні статті в часопису “Соціологія: теорія, методи, маркетинг”: “Суб’єктивна надійність: теорія і методи вимірювання (ICN)” [Головаха, Горбачик, Любива та ін., 2008] та “Теоретична валідність вимірювальної процедури і зміщення даних у соціологічному дослідженні” [Дембіцький, 2008].

У радянській традиції інтегральною характеристикою якості соціологічного вимірювання й інформації¹ як результату цього вимірювання вважали надійність. Такої точки зору дотримувалися В.Ядов, Г.Саганенко, В.Волович, Б.Докторов, В.Паниотто. Б.Докторов у 1979 році та В.Паниотто у 1986-му проаналізували напрацювання своїх колег із питань якості соціологічного вимірювання. На підставі їх можна визначити основні показники надійності соціологічної інформації у радянській традиції: обґрунтованість, правильність, стійкість.

Обґрунтованість (валідність) — відсутність теоретичних, методологічних, логічних помилок, слушність референцій при розробленні методики дослідження, відповідність числової моделі досліджуваній соціальній системі [Паниотто, 1986: с. 27]. Б.Докторов називав це також відсутністю квазі-помилки узгодженості, тобто відповідністю моделі об’єкта самому об’єктові [Докторов, 1979: с. 28].

Правильність — міра відсутності систематичних помилок вимірювання [Паниотто, 1986: с. 31].

Стійкість — міра відсутності випадкових помилок вимірювання [Паниотто, 1986: с. 35].

У західній традиції немає інтегральної характеристики якості соціологічного вимірювання; тут найчастіше вирізняють два показники якості со-

¹ Інформацію, отриману в результаті якісного вимірювання, можна вважати якісною. Надалі питання розрізнення понять якості соціологічного вимірювання та якості соціологічної інформації не розглядатиметься, оскільки соціологічна інформація є результатом соціологічного вимірювання.

ціологічного вимірювання — валідність (validity) та надійність (reliability). Така позиція відображена в усіх Стандартах психологічного та освітнього тестування [Standards, 1966; 1974; 1985; 1999], у класичній праці “Оцінювання надійності та валідності” Дж.Кармінеса та А.Зелера [Carmines, 1979], у такому сучасному виданні, як “Довідник “Сейдж” з методів у соціальних дослідженнях” (Sage Handbook of social research methods) [Sage Handbook, 2008]. У західній традиції валідність та надійність розуміють так:

Валідність (validity) — це відповідність вимірювання його меті [Carmines, 1979: p. 12]. Валідність є мірою того, наскільки ми справді вимірюємо те, що заплановано [Bohrsted, 2000: p. 3207]. Валідність пов’язують з відсутністю систематичних помилок; вимірювання, відносно вільне від систематичних помилок, вважають валідним [Carmines, 1980: p. 14].

Надійність (reliability) — це відтворюваність вимірювання. Надійність стосується того, якою мірою експеримент, тест чи будь-яка процедура вимірювання дає ті самі результати за повторного вимірювання. Надійність пов’язана з тенденцією повторюваності в разі кількаразового вимірювання одних і тих самих соціальних феноменів одним і тим самим вимірювальним інструментом. Чим стійкіші результати вимірювання за повторних спроб, тим надійніша вимірювальна процедура, і навпаки, чим менше схожі результати вимірювання за повторного вимірювання, тим менш надійною є процедура вимірювання [Carmines, 1979: p. 11–12]. Надійність пов’язують із відсутністю випадкових помилок — вимірювання, відносно вільне від випадкових помилок, вважається надійним [Carmines, 1980: p. 14].

З наведеного аналізу радянської та західної традицій видно, що за змістом показники якості вимірювання у них не суперечать один одному. Основні відмінності полягають у термінології. Відмінність перша — у радянській традиції є інтегральна характеристика якості соціологічної інформації — надійність, у західній традиції такої інтегральної характеристики немає. Але це не змінює сутності розуміння якості вимірювання. У радянській традиції показниками надійності як інтегральної характеристики якості вимірювання є обґрунтованість, правильність і точність, у західній показниками якості вимірювання слугують валідність та надійність. Виходячи зі змісту наведених вище понять, спробуємо їх співвіднести (табл.).

Таблиця

Термінологічне зіставлення радянської та західної традицій

РАДЯНСЬКА ТРАДИЦІЯ	ЗАХІДНА ТРАДИЦІЯ
Надійність — інтегральна характеристика якості вимірювання (інформації)	Якість інформації (quality of measurement)
Обґрунтованість	Валідність (validity)
Правильність	—
Стійкість	Надійність (reliability)

У західній традиції не розмежовують показники відсутності теоретичних та систематичних помилок, об’єднуючи їх в одну категорію — валідність. У радянській традиції ці показники розмежовують — обґрунтованість

та правильність. На мій погляд, адекватніше розділяти ці категорії. Але з іншого боку, як Б.Докторов, так і В.Паніотто відзначають, що чітко розмежувати обґрунтованість та правильність інформації не завжди вдається, оскільки в деяких випадках теоретичні помилки збігаються із систематичними, а в інших випадках ця розбіжність виражена досить яскраво [Паніотто, 1986: с. 31, Докторов, 1979: с. 28]. Як зазначив В.Паніотто “відмежувати обґрунтованість від правильності — це, зрештою, означає вказати, де закінчується методологія і починається методика... Образно кажучи, обґрунтованість характеризує напрям, в якому ми рухаємося, а правильність — траєкторію руху в даному напрямі” [Паніотто, 1986: с. 32]. Стійкість у радянській традиції за змістом відповідає надійності у західній традиції.

Як зазначалося, нині, враховуючи інтегрованість української соціології у міжнародний контекст, зручніше використовувати міжнародну термінологію, але при цьому впроваджувати в неї напрацювання радянської традиції. Ще 1986 року В.Паніотто писав: “Із двох варіантів — використовувати як “ім’я” поняття “надійність” чи загалом не присвоювати імені якості інформації — більш зручним я вважаю останній варіант (використовувати надійність і обґрунтованість як рівноправні складові якості). Але щоб не виникало розбіжностей з уже усталеним у вітчизняній літературі слововжитком..., в інтересах пришвидшення процесу стандартизації соціологічної термінології тут і надалі ... я розумітиму надійність як інтегральну, узагальнювальну характеристику якості соціологічної інформації, як її “ім’я”” [Паніотто, 1986: с. 25–26]. Я, з тих самих міркувань, що й В.Паніотто, пропоную задля узгодженості вітчизняної термінології з міжнародною розуміти валідність та надійність як рівнозначимі аспекти якості вимірювання.

У перебігу розвитку поняття “валідність” у науковий обіг було введено різні види валідності. Але валідність як показник міри відповідності вимірювання тому, що заплановано виміряти, є унітарним концептом. Якщо валідність — це унітарний концепт, то коректніше говорити не про різні види валідності, а про різні підходи до перевірки валідності. Аналізуючи різні види валідності, описані в літературі, ми бачимо, що види валідності передбачають різні підходи, різні методи перевірки однієї властивості — валідності.

Отже, слід розглядати валідність як унітарний концепт, а різні види валідності, описані в літературі, — як різні методи перевірки валідності.

Визначивши зміст поняття валідність шляхом співвіднесення термінів радянської та західної традицій, показавши співвідношення валідності з надійністю та показавши, що валідність є властивістю, наявність якої можна перевіряти різними методами, я завершую теоретичну інтерпретацію поняття валідність і перехожу до емпіричної інтерпретації даного поняття.

Емпірична інтерпретація поняття “валідність”

У рамках емпіричної інтерпретації належить показати, до яких етапів вимірювання слід застосовувати поняття валідність, конкретніше — які етапи вимірювання слід перевіряти на валідність, а які етапи виходять за межі валідності.

Межі будь-якого поняття можна окреслити у тому чи тому більш загальному контексті. Таким контекстом для валідності є вимірювання. За основу візьмемо визначення вимірювання Г.Блейлока, на сьогодні найбільш цитоване у західній літературі. Блейлок розумів вимірювання як теоретично навантажений процес і вбачав принципову методологічну проблему в аналізі зв'язку між змістовною соціологічною теорією і відкрито застосовно до цієї теорії концепцією вимірювання. За його визначенням, “вимірювання — це процес співвіднесення абстрактних понять, які не можна прямо спостерігати, з емпіричними індикаторами, які піддаються прямому спостереженню” [Blalock, 1969(a): p. 6]. Блейлок вказує на неможливість проведення вимірювання без наявності теоретичних рамок і виступає проти протиставлення теорії та емпіричних досліджень [Blalock, 1969(b), p. vi].

Спираючись на це визначення, виокремлю етапи вимірювання (на прикладі кількісного опитування).

1. Створення теоретичної моделі¹ досліджуваного явища, яка адекватно відображає саме досліджуване явище.
2. Добір емпіричних індикаторів відповідно до розробленої теоретичної моделі.
3. “Переклад” індикаторів у конкретні питання та вибір найадекватніших шкал.
4. Розроблення репрезентативної вибірки.
5. Підготовка необхідної для проведення польового етапу документації (інструкції інтерв'юерам, маршрутні листи), інструктаж інтерв'юерів.
6. Проведення польового етапу (збирання інформації), кодування отриманої інформації.

Якщо розглядати валідність як характеристику соціологічного вимірювання чи як характеристику соціологічної інформації (результату соціологічного вимірювання), то до процесу перевірки валідності слід включати перевірку правильності проведення всіх етапів соціологічного вимірювання, означених вище. В такому разі валідність стає інтегральною характеристикою якості соціологічного вимірювання і соціологічної інформації. Але загальноприйнято, що якість вимірювання передбачає і надійність, і репрезентативність вибірки, і якісний вибір та забезпечення роботи інтерв'юерів. А це, своєю чергою, надто розширює та “розмиває” рамки процедури перевірки валідності конкретного вимірювання.

Таким чином, виходячи з прагнення чітко окреслити “межі” поняття валідності та розробити схему перевірки валідності, щоб не вносити терміно-

¹ Оскільки більшість об'єктів соціологічного дослідження не можна безпосередньо вимірювати, то дослідники, перш ніж проводити дослідження, мають чітко описати досліджуваний об'єкт – вирізнити його основні властивості, диференціювати складові частини, якщо це можливо. Саме такий теоретичний опис, з виокремленням характерних властивостей та складників, я маю на увазі, говорячи про створення теоретичної моделі досліджуваного об'єкта. Синонімом теоретичної моделі досліджуваного явища можна назвати створений дослідником теоретичний концепт цього явища.

логічну плутанину в уже й без того заплутаний дискурс соціологічного вимірювання, а також зважаючи на попередній досвід, вважаю доречним розглядати валідність як характеристику адекватності вимірювального інструментарію та коректності застосування певного інструменту в конкретній ситуації дослідження.

Згідно із цим висновком “межі” перевірки вимірювання на валідність включають перевірку таких аспектів:

- наскільки адекватно розроблена теоретична модель досліджуваного явища відображає саме досліджуване явище;
- наскільки адекватно підібрані емпіричні індикатори;
- наскільки адекватно емпіричні індикатори “перекладені” у конкретні питання та наскільки адекватно підібрані шкали.

Тобто валідність — це показник якості вимірювання, яку слід забезпечити до проведення самого процесу вимірювання, до збирання даних. Валідності належить досягати на етапах розроблення теоретичної моделі досліджуваного явища та інструментарію вимірювання, а перевіряти після завершення процесу вимірювання на основі зібраних даних.

Якщо прийняти такий погляд на валідність, то валідність, поряд із надійністю, репрезентативністю та якістю проведення польового етапу є одним із показників якості соціологічного вимірювання, відповідно і якості соціологічної інформації (рис. 1).

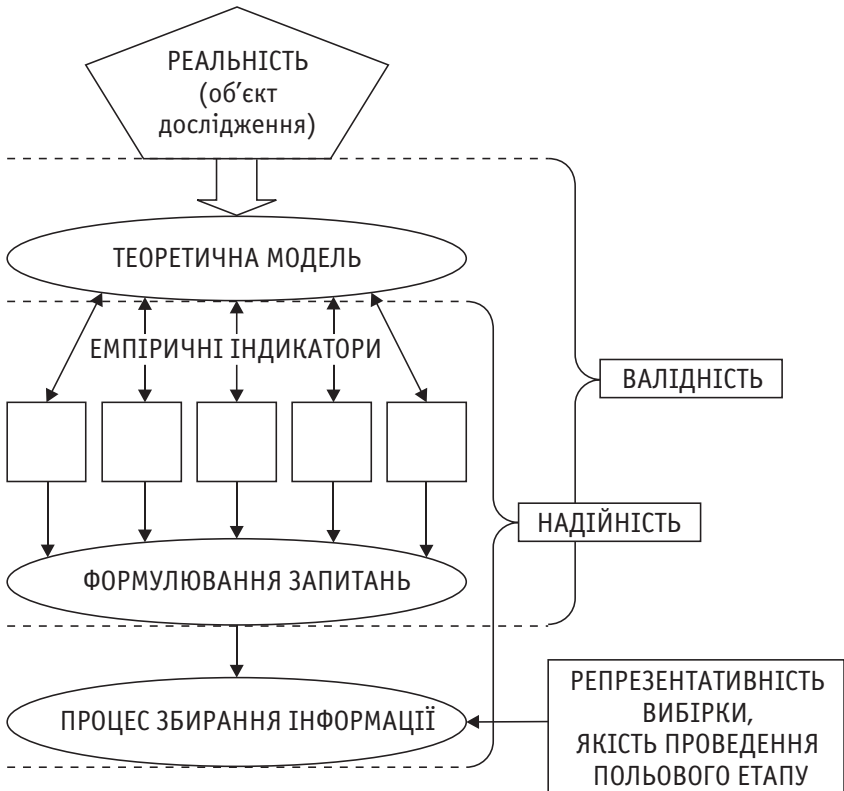


Рис. 1. Етапи процесу вимірювання

Визначивши “межі” поняття валідності шляхом окреслення етапів вимірювання, на яких вона має досягатись, і окресливши, що саме має піддаватись валідизації, ми завершуємо емпіричну інтерпретацію і переходимо до операціоналізації.

Операціоналізація поняття валідність

Як зазначалося, валідність — це показник властивості вимірювання — міри того, наскільки вимірювальний інструмент вимірює те, що заплановано виміряти. Операціоналізація поняття валідність — це пошук послідовних операцій, що дають змогу встановити наявність та інтенсивність вияву властивості у носія (носієм валідності є вимірювальний інструмент). Фактично йдеться про пошук конкретних методів перевірки валідності.

Слід зауважити, що словосполучення “методи перевірки валідності” не дуже часто зустрічається в літературі. Частіше пишуть про “види валідності”. На мій погляд, коректніше використовувати саме вираз “методи перевірки валідності”. Незважаючи на це, для опису наявного досвіду перевірки валідності тут буде застосовано термін “види валідності”, оскільки я посилатимусь на джерела, в яких його використано.

У західній традиції (див. рис. 2) найчастіше вирізняють *три види валідності*: змістову, критеріальну (конкурентну і прогностичну) та конструктну (конвергентну і дискримінантну). Деякі дослідники виокремлюють очевидну, або зовнішню, валідність (face validity).

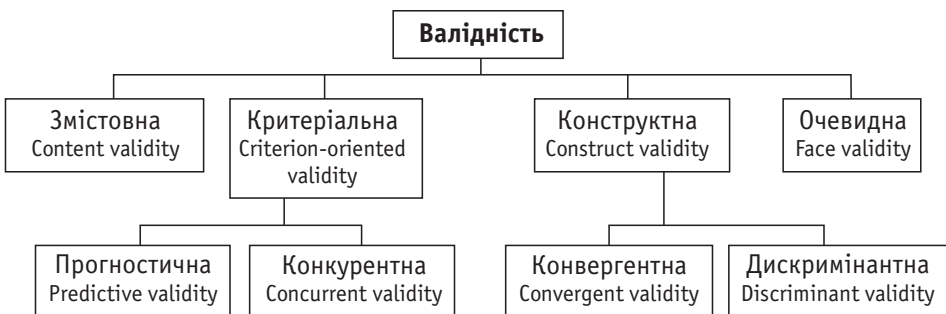


Рис. 2. Види валідності

У радянській традиції або взагалі не розмежовують різні види обґрунтованості (валідності), або вирізняють два види обґрунтованості (валідності): теоретичну й емпіричну [Паниотто, 1986: с. 109], розуміючи під теоретичною валідністю конструктну валідність, а під емпіричною — критеріальну валідність. Змістова валідність випадає з такої класифікації, позаяк її автор розглядає її як таку, що стосується тільки тестів, спрямованих на оцінювання деяких знань та вмінь [Паниотто, 1986: с. 106]. Оскільки я розглядаю змістовну валідність ширше (див. нижче), вважаю цей вид валідності дуже важливим. Фактично за змістом теоретична та емпірична валідності відповідають конструктній та критеріальній.

Змістова валідність. Існує багато визначень змістової валідності, більшість із яких за змістом відповідає такій дефініції: *змістова валідність є мірою релевантності й репрезентативності елементів¹ вимірювального інструменту вимірюваному конструкту* [Haynes, Richard, Kubany, 1995: р. 238]. При цьому *релевантність* означає відповідність елементів інструменту явищу, яке вимірюється. Релевантним є інструмент, елементи якого співвіднесені з усіма аспектами досліджуваного явища, а отже, після аналізу зібраних даних ми маємо отримати інформацію про всі сторони досліджуваного явища. Натомість *репрезентативність* вимірювального інструменту — це міра того, наскільки пропорційно елементи вимірювального інструменту співвідносяться з явищем, яке вимірюється, тобто елементи вимірювального інструменту мають пропорційно представляти різні аспекти досліджуваного явища.

Змістова валідність може різнитися для різних груп, зокрема національних чи культурних. Тому її слід оцінювати конкретно для групи, у якій проводитиметься дослідження. Наприклад, вимірювальний інструмент може бути валідним для американців і водночас невалідним для азійців, оскільки люди різних культур по-різному можуть розуміти зміст одного запитання [Marsella, Kameoka, s. a].

Змістова валідність завжди має визначатися для конкретної сфери дослідження, щодо якої було розроблено вимірювальний інструмент [Haynes, Richard, Kubany, 1995: р. 241]. Розроблення вимірювальних інструментів відбувається у межах наявних теорій. Оскільки теоретичні аспекти схоплення явища розвиваються, релевантність та репрезентативність елементів вимірювального інструменту також підлягають перегляду.

Таким чином, змістова валідність завжди прив'язана до конкретного проміжку часу. Динамічна природа явищ, досліджуваних у соціології, має низку наслідків для змістової валідності [Haynes, Richard, Kubany, 1995: р. 242]:

1. Показники змістової валідності не можуть бути стабільними впродовж довгого проміжку часу.
2. Змістову валідність вимірювальних інструментів належить періодично переглядати.
3. Вимірювальні інструменти мають адекватно відображати зміни у досліджуваному явищі.
4. Без перегляду (старих) вимірювальних інструментів можна зробити неправильні висновки щодо досліджуваного явища.

Багато авторів окреслювали рекомендовані методи змістової валідації. Наведу основні з них.

¹ У стандартизованому інтерв'ю елементами вимірювального інструменту (анкети) є запитання, у контент-аналізі — категорії аналізу, у глибинному інтерв'ю — обговорювані теми. Іноді для зручності пояснення замість терміна “елементи вимірювального інструменту” тут використовуватиметься термін “запитання вимірювального інструменту”, що реферує до стандартизованого інтерв'ю. Але наведені пояснення адекватні також для всіх інших методів збирання інформації.

Слід чітко визначити сферу належності явища та його аспекти, перш ніж переходити до формулювання питань вимірювального інструменту [Northorp, 1947; Suen, 1990]. Цей крок є основною та найважчою фазою у розробленні валідного за змістом вимірювального інструменту [Mурphy, 1994].

Слід дослідити думку експертів та респондентів для початкового розроблення питань (у випадку інтерв'ю), для чого можна використовувати глибинні інтерв'ю з експертами та респондентами. Це сприяє тому, щоб елементи (запитання) вимірювального інструменту були релевантними та репрезентативними щодо вимірюваного явища. Під час цього процесу можуть бути виявлені нові аспекти явища, що вимагатиме перегляду його теоретичної моделі.

Слід дослідити пропорційну репрезентативність елементів вимірювального інструменту, що представляють різні аспекти досліджуваного явища.

Елементи (запитання) у вимірювальному інструменті мають бути розподілені, або “зважені”, таким чином, щоб відображати важливість та інтенсивність різних аспектів досліджуваного явища. Якщо окремі питання надмірно чи недостатньо репрезентують аспекти об'єкта дослідження (тобто певні аспекти досліджуваного явища представлені у вимірювальному інструменті більшою або меншою, ніж потрібно, кількістю запитань), то в отриманих даних ми матимемо зсув, що призведе до зміщень у висновках.

Отже, підсумуємо основні методи перевірки змістової валідності [Haines, Richard, Kubany, 1995; p. 245]:

1. Детальне визначення досліджуваного явища.
2. Вивчення думки респондентів та експертів під час створення моделі досліджуваного об'єкта.
3. Експертне оцінювання елементів вимірювального інструменту на предмет релевантності та репрезентативності щодо вимірюваного явища.

Змістова валідність забезпечується на етапі побудови теоретичної моделі об'єкта дослідження та розроблення вимірювального інструменту. Перевірка змістової валідності означає перевірку відповідності створеної теоретичної моделі досліджуваного явища самому явищу та перевірку відповідності вимірювального інструменту розробленій моделі. Досягнення змістової валідності подібне до теоретичної й емпіричної інтерпретації та операціоналізації основних понять дослідження, а перевірка змістової валідності — до з'ясування коректності теоретичної та емпіричної інтерпретації й операціоналізації. Перевірка змістової валідності апробованого раніше інструменту може виявити необхідність коригування цього інструменту.

Критеріальна валідність полягає у відповідності інформації щодо досліджуваного явища, отриманої з використанням вимірювального інструменту, валідність якого перевіряється, й інформації, отриманої іншими засобами та з інших джерел, яка слугує критерієм валідності [Carmines, Zeller, 1980: p. 17]. Такий критерій має визначатися іншим валідним щодо того самого явища вимірювальним інструментом, яким є статистика.

Розрізняють два різновиди валідності за критерієм — конкурентну, тобто паралельну, та прогностичну.

Конкурентна валідність (concurrent validity) визначається шляхом співвіднесення даних, отриманих валідизовуваним інструментом, з даними, отриманими за критерієм валідності у той самий період часу [Carmines, Zeller, 1980: p. 18].

Прогностична валідність (predictive validity) визначається шляхом вивчення кореляції між даними, отриманими валідизовуваним інструментом, і деяким критерієм, що дає змогу визначити стан досліджуваної властивості у майбутньому [Клайн, 1994: с. 27]. Наприклад, прогностична валідність інструменту, що прогнозує результати виборів, визначається шляхом вивчення кореляції результатів вимірювання з результатами голосування. Валідність тесту відбору працівників перевіряється шляхом вивчення кореляції результатів тесту з результатами роботи відібраних працівників.

Головна складність у здійсненні критеріальної валідизації полягає у виборі валідних зовнішніх критеріїв. Наприклад, для інструменту, що вимірює електоральні преференції, зовнішнім критерієм є вибори. Для інструменту, що вимірює матеріальний стан громадян, — дані податкової служби. Але такі критерії валідності не завжди доступні, а часто їх просто не існує. Тому даний метод перевірки валідності не завжди застосовний до соціологічних вимірювань.

В.Аванесов виокремлював ще один вид валідності — *валідність за відомими групами* [Паниотто, 1986: с. 104]. Суть методу полягає в тому, що деяким способом (наприклад, за експертними оцінками) обирають дві групи респондентів, які відрізняються за вимірюваною властивістю і проводять опитування цих респондентів з використанням інструменту, що валідизується. Якщо у цих двох групах результати вимірювання справді відрізняються, то це є свідченням на користь валідності вимірювального інструменту. Якщо ні — то або вимірювальний інструмент не є валідним, або гіпотеза про розбіжність груп не є слушною.

У популярній серед американських дослідників книжці “Методологія дослідження” [Survey Methodology, 2004] останній метод перевірки теж згадується і наводиться такий приклад: “Очікується, що рівень консерватизму американців, які голосують за республіканців, буде вищим, ніж у американців, які голосують за демократів. У разі отримання негативного результату можна зробити два висновки — або теорія не відповідає дійсності, або вимірювальний інструмент не є валідним. Взагалі, результати можна інтерпретувати тільки в рамках добре розробленої теорії” [Survey Methodology, 2004: p. 256].

Конструктна валідність — це міра того, наскільки вимірювальний інструмент вимірює конструкт, який має бути виміряний [Haynes, Richard, Kubany, 1995: p. 239]. Поняття конструкту було введено Л.Кронбахом та П.Мелом у 1955 році. За їхнім визначенням “конструкт — це деяка характеристика властивість людини, яка має бути відображена у результатах тесту. У рамках валідизації тесту характеристика, з приводу якої роблять висновок у процесі інтерпретації тесту, становить конструкт” [Stonbach, Meehl, 1955: p. 283]. У сучасній літературі існують різні визначення конструкту і різні

співвідношення між поняттями “конструкт” і “концепт”. Не буду заглиблюватися у це питання, оскільки воно є темою окремої статті. Тут розумітимемо поняття конструкту як певну описану теоретично характеристику досліджуваного явища, яку слід виміряти, позаяк це відповідає змісту конструктної валідизації.

Конструктна валідність базується на теорії вимірюваного явища. Дослідник має побудувати номологічну мережу — набір конструктів та зв'язків між ними — та “вписати” в цю мережу явище, вимірюване інструментом, що підлягає валідизації. На основі наявної літератури, результатів інших досліджень, логіки, суджень експертів та попередніх інтерв'ю з респондентами дослідник приймає гіпотези щодо того, з вимірюваннями яких явищ має корелювати вимірювання досліджуваного явища, а з якими не має корелювати, а потім тестує свої гіпотези на зібраних даних. Д.Кемпбел та Д.Фіске розробили стратегію перевірки конструктної валідності. Для цього вони ввели два підвиди конструктної валідності — конвергентну та дискримінантну валідності [Campbell, Fiske, 1959]:

- *конвергентна валідність* є оцінкою того, наскільки вимірювання, які мають корелювати, виходячи з теоретичних міркувань, корелюють між собою, згідно із зібраними даними;
- *дискримінантна валідність* є оцінкою того, наскільки вимірювання різних конструктів, які, за теорією, не мають корелювати між собою, не корелюють між собою з огляду на зібрані дані.

Конструктну валідність будь-якого вимірювального інструменту можна оцінити, якщо досліджуваний конструкт має місце у теоретичному контексті. Якщо результати вимірювання не розбіжні з теоретичними очікуваннями, то вимірювальний інструмент можна вважати конструктно валідним [Carmines, Zeller, 1980].

Але що робити у випадку, коли дослідник не виявляє очікуваного зв'язку?

На це питання дали відповідь самі автори поняття конструктної валідності Л.Кронбах та П.Мел [Cronbach, Meehl, 1955]. Решта дослідників підтримали цю стратегію і дотримуються її й по сьогоднішній день [28].

Причини розбіжностей результатів вимірювання з теоретичними очікуваннями можуть бути такими [Campbell, Fiske, 1959, p. 295]:

1. Інструменти, якими вимірюються інші конструкти номологічної мережі, в яку поміщений досліджуваний конструкт, не є валідними.
2. Експериментальний дизайн є хибним.
3. Вимірювальний інструмент, що валідизується, не вимірює потрібний нам конструкт.
4. Номологічна мережа (набір конструктів та гіпотези про зв'язки між цими конструктами) є хибною.

Коли ми не певні у правильності експериментального дизайну, ми повинні повторити дослідження з урахуванням попередніх помилок, щоб бути впевненими у правильності нашого дизайну.

У разі сумніву щодо валідності вимірювальних інструментів, що вимірюють інші конструкти номологічної мережі, ми маємо, мірою можли-

вості, використати інші, валідні, вимірювальні інструменти для цих конструктів і перевірити зв'язок на нових даних. Найкраще перевіряти конструктну валідність із використанням індикаторів, валідність і надійність яких досить висока. Таким чином, теоретично релевантні, валідні та надійні зовнішні індикатори є вирішальним моментом для оцінювання конструктної валідності вимірювального інструменту.

Якщо ми впевнені у валідності вимірювальних інструментів, що вимірюють конструкти номологічної мережі, та в експериментальному дизайні, ми можемо дійти висновку, що вимірювальний інструмент неадекватно вимірює конструкт. У такому разі треба розробити та валідизувати інший тест. Валідизацію нового інструменту належить здійснювати на нових даних.

Можна піддати сумніву номологічну мережу, яка лежить в основі вимірювання конструкту. Тоді дослідник має розробити нову номологічну мережу. Модифікуючи мережу, він має зібрати нові дані для тестування альтернативної гіпотези.

Зараз немає чіткої процедури, яка б давала змогу визначити, який із наведених варіантів є підставою негативного результату перевірки конструктної валідності вимірювального інструменту. Для вибору слід з'ясувати, наскільки добре теоретична схема підтверджувалася попередніми даними, зважити й докази на захист теоретичної схеми й зів'язати з останнім невдалим експериментом.

Після проведення процесу валідизації може знадобитися зміна вимірювального інструменту для конструкту, який нас цікавить, або перегляд теоретичної основи конструкту.

Результати валідизації можуть спричинити потребу в очищенні визначення конструкту або перегляді вимірювального інструменту чи процесу проведення дослідження [Standards, 1999].

Очевидна валідність. Деякі дослідники вирізняють також очевидну, або зовнішню, валідність (face validity). Тест (у випадку соціологічного вимірювання — вимірювальний інструмент) є очевидно валідним, якщо явно вимірює саме те, що мають на увазі респонденти. Очевидна валідність вимірювального інструменту допомагає досягти співпраці з респондентами, будучи пов'язана тільки з тим, як сприймають цей вимірювальний інструмент респонденти [Клайн, 1994: р. 26]. Очевидна валідність, по суті, не має стосунку до валідності у класичному розумінні; тут не йдеться про метод перевірки валідності.

Узагальнена схема перевірки валідності

На підставі викладеного огляду відомих методів перевірки валідності спробуймо окреслити узагальнену схему перевірки валідності конкретного вимірювального інструменту.

1. Перевірка адекватності моделі вимірювального явища з використанням результатів відомих досліджень, експертних оцінок та попередніх інтерв'ю з респондентами (у випадку опитування). На цьому етапі перевірки

валідності ще не проводиться збирання даних, і перевірку здійснюють логічним шляхом. Формалізувати такий процес доволі важко. Цей етап перевірки валідності відповідає описаній вище змістовій валідизації.

2. Перевірка відповідності вимірювального інструменту розробленій теоретичній моделі досліджуваного явища. Цей етап здійснюється як із використанням попереднього досвіду та на основі дослідження сприйняття розроблених індикаторів респондентами (змістова валідизація), так і на підставі зібраних даних (критеріальна та конструктна валідизація). Розпишімо детальніше процес перевірки відповідності вимірювального інструменту розробленій теоретичній моделі:
 - a) розроблення номологічної мережі — набору конструктів та зв'язків між ними;
 - b) проведення емпіричного дослідження для перевірки гіпотез про зв'язки між конструктами номологічної мережі;
 - c) добір критеріїв валідності (якщо вони існують) для критеріальної валідизації інструменту та дослідження зв'язків між даними, отриманими з використанням вимірювального інструменту, що його валідизують, та критерієм валідності;
 - d) виокремлення груп, які мають відрізнитися за досліджуваною ознакою (якщо такі групи можна вирізнити), та проведення дослідження в цих двох групах із застосуванням інструменту, що валідизується;
 - e) у випадку:
 - i) підтвердження гіпотез про зв'язки між конструктами номологічної мережі,
 - ii) виявлення зв'язку між даними, отриманими з використанням вимірювального інструменту, що валідизується, та критерієм валідності (якщо такий існує),
 - iii) отримання різних результатів вимірювання у виокремлених групах (якщо такі можна вирізнити)розроблений інструмент можна вважати валідним.
 - f) у випадку, коли:
 - i) гіпотези про зв'язки між конструктами номологічної мережі не підтверджено,
 - ii) зв'язок між вимірюванням і критерієм валідності не виявлено,
 - iii) різні результати у різних групах не отримано,треба переглядати розроблений вимірювальний інструмент та проводити валідизацію нових даних;
 - g) якщо після проведення повторної валідизації переглянутого вимірювального інструменту на нових даних теоретичні гіпотези не підтверджено і не виявлено зв'язку між вимірюванням та критерієм валідності, треба повертатися до розробленої моделі вимірювального явища та номологічної мережі, в яку “вбудоване” це явище, щоб переглянути їх, а потім проводити валідизацію вимірювального інструменту, скоригованого на основі оновленої моделі та оновленої номологічної мережі на нових даних.

Валідизація вимірювальних інструментів — це постійний, колективний процес [Cronbach, 1987: р. 163–164]. У ньому мають брати участь усі члени наукової спільноти, які займаються вимірюванням певного явища. Відповідальність за валідизацію лежить як на розробнику вимірювального інструменту, так і на дослідниках, які використовують цей інструмент. Розробник вимірювального інструменту відповідає за доведення валідності вимірювального інструменту і за інструкції щодо правильного його використання. Користувачі цього інструменту відповідають за оцінювання доказів його валідності в конкретній ситуації застосування [Standards, 1999]. Валідність наявних інструментів належить періодично знову перевіряти. Таку перевірку валідності можуть проводити інші дослідники, а не лише автор вимірювального інструменту. Це може бути зумовленим або значним проміжком часу з моменту розроблення певного інструменту, або прагненням використати його в іншому культурному контексті. Тільки постійно підтверджуючи у різних дослідженнях валідність вимірювальних інструментів, ми можемо бути впевненими в адекватності наших висновків про досліджувані явища.

Література

- Волович В.И.* Надежность информации в социологическом исследовании: Проблемы методологии и методики. — К., 1974.
- Головаха Є., Горбачик А., Любива Т., Паніна Н., Серета В., Урсуленко К.* Суб'єктивна надійність: теорія і методи вимірювання (ІСН) // Соціологія: теорія, методи, маркетинг. — 2008. — №1. — С. 166–188.
- Девятко И.Ф.* Методы социологического исследования. — М., 2002.
- Дембицький С.* Теоретична валідність вимірювальної процедури і зміщення даних у соціологічному дослідженні // Соціологія: теорія, методи, маркетинг. — 2008. — № 3. — С. 99–118.
- Докторов Б.З.* О надежности измерения в социологическом исследовании. — Л., 1979.
- Здравомыслов А.Г.* Методология и процедура социологических исследований. — М., 1969.
- Клайн П.* Справочное руководство по конструированию тестов. Введение в психометрическое проектирование. — К., 1994.
- Паніна Н.В.* Технологія соціологічного дослідження. — К., 1996.
- Паниотто В.И.* Качество социологической информации (методы оценки и процедуры обеспечения). — К., 1986.
- Саганенко Г.И.* Надежность результатов социологического исследования. — Л., 1983.
- Ядов В.А.* Социологическое исследование: Методология, программа, методы. — М., 1972.
- American Psychologist.* — 1975. — Vol. 30. — P. 955–966.
- Blalock H.M.* The measurement problem // *Methodology in Social Research* / Ed. by H.M.Blalock, A.Blalock. — N.Y., 1969.
- Blalock H.M.* Theory Construction: From verbal to Mathematical Formulation. Englewood Cliffs, N.J., 1969.
- Bohrsted G.W.* Validity // *Encyclopedia of Sociology, Second Edition* / 2nd ed. — Vol. 5. — N.Y., 2000.
- Campbell D.T., Fiske D.W.* Convergent and discriminant validation by the multitrait-multimethod matrix // *Psychological Bulletin.* — 1959. — № 58. — P.81–105.

- Carmines E.G., Zeller R.A.* Measurement in the social sciences. — Cambridge, 1980.
- Carmines E.G., Zeller R.A.* Reliability and Validity Assessment. — Beverly Hills, 1979.
- Cronbach L.J., Meehl P.E.* Construct Validity in Psychological Tests // Psychological Bulletin. — 1955. — № 52. — P. 281–302.
- Cronbach L.J., Meehl P.E.* Construct Validity in Psychological Tests // Psychological Bulletin. — 1955. — № 52. — P. 281–302.
- Cronbach L.J.* Construct validity after thirty years // Linn R.E. (ed.), Intelligence: Measurement, Theory, and Public Policy. — Urbana, 1987.
- Cureton E.* Validity // Educational Measurement. — Washington, D.C, 1951. — Ch. 16. — P. 621–695.
- Guilford P.* Psychometric Methods. — S.I., 1936.
- Haynes S., Richard D., Kubany E.* Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods // Psychological Assessment. — 1995. — № 3. — P. 238–247.
- Marsella A.J., Kameoka V.* Ethnocultural issues in the assessment of psychopathology // Wetzler S. (ed.). Measurement mental illness: Psychometric assessment for clinicians. — Washington, DC, 1995. — P. 231–256.
- Messick S.* The standard problem: Meaning and values in measurement and evaluation // American Psychologist. — 1975. — October. — P. 955–966.
- Myrphy K.R., Davidshofer C.O.* Psychological testing: Principles and applications (3rd ed.). — Englewood Cliffs, N.J., 1994.
- Northrop F.S.C.* The Logic of the Sciences and the Humanities. — N.Y., 1947.
- Sage Handbook of social research methods. — S.I., 2008
- Technical recommendations for psychological tests and diagnostic techniques // Psychological Bulletin [Supplement]. — 1954. — № 51 (2, part 2).
- Suen H.K.* Principles of test theories. — Hillsdale, N.J., 1990.
- Survey Methodology / Series W., Groves R.M., Fowler F.G., Jr, Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. — S.I., 2004.
- Standards for educational and psychological tests and manuals / American Psychological Association. — Washington, DC, 1966.
- Standards for educational and psychological tests and manuals / American Psychological Association. — Washington, DC, 1974.
- Standards for educational and psychological testing / American Educational Research Association. — Washington, DC, 1985.
- Standards for educational and psychological testing / American Educational Research Association, American Psychological Association and National Council on Measurement in Education. — Washington, DC, 1999.
- Thurstone L.L.* The Reliability and Validity of the Test. — Ann Arbor, MI, 1931.