

1. ПОНЯТТЯ ОСВІТНЬОГО ВИМІРЮВАННЯ

Три теорії вимірювань. Людина, яка вперше зустрічається з ідеєю вимірювання таких феноменів, як «інтелект», «навченість», «здібність», відчуває збентеження. У повсякденному житті вимірювання застосовуються до фізичних властивостей предметів, таких як довжина, вага, об'єм, тощо. Фізичне вимірювання має ряд властивостей, які важко перенести на сферу психічного. Важливою властивістю фізичного вимірювання є адитивність, або існування одиниці вимірювання такої, що міра предмета є простою сумою цих одиниць, так, якби ми при вимірюванні відкладали послідовно якийсь «твердий брусок», рахуючи кількість відкладень.

Таке розуміння вимірювання існувало у фізиці століттями, аж до тих часів, коли фізична наука проникла у мікросвіт. Розвиваючи квантову механіку, фізики зіштовхнулися з питаннями: 1) що ми вимірюємо? 2) Як вимірювальний прилад може впливати на сам процес вимірювання?

Для психології та інших суспільних наук ці питання були актуальними від самого початку. Якщо у фізиці дослідник взаємодіє з суб'єктом вимірювання через вимірювальний прилад, то в суспільних науках таким «приладом» є вся організація експерименту. Потреба суспільних наук у вимірюваннях змусила дослідників переосмислювати саме поняття вимірювання. У 19-20 століттях спостерігались три підходи, або три теорії вимірювань, які використовувались у психології: класична, операціональна та репрезентативна.

У відповідності з *класичною теорією вимірювань*, психологія є кількісною наукою. Засновник психофізіології Густав Фехнер вважав, що в загальному випадку вимірювання кількості полягає у з'ясуванні, як часто одиниця кількості одного і того ж виду міститься у ній. У наш час класична теорія практично повністю витіснена репрезентативною та операціональною теоріями.

Згідно з *операціональною теорією вимірювань*, яка розвивалася у першій половині 20-го століття, під будь-яким концептом (тобто поняттям) розуміється не що інше як сукупність операцій. З операціональної точки зору будь-яке вимірювання – це операція,

що породжує числа. Числа є самодостатніми і не залежать від природи об'єкта, щодо якого відбувається вимірювання. В граничному розумінні для операціоналіста наука є просто вивченням виконуваних операцій, а не дослідженням реальності. Так, операціоналіст розглядає результати тестування як вимірювання вже тоді, коли вони є наслідком осмисленого приписування чисел, виконаного за допомогою строго визначених операцій.

Для послідовника *репрезентативної теорії вимірювань*, отриманих в результаті тестування оцінок недостатньо для того, щоб стверджувати, що вони є результатом вимірювання, хоча й отримані внаслідок чітко визначених операцій. Для того, щоб оцінки можна було вважати результатом вимірювань, необхідно, щоб відношення між ними (скажімо, відношення порядку) репрезентували якісні емпіричні відношення між успішністю виконання тесту різними особами, і дослідження мають полягати в ідентифікації таких відношень і описанні їх характеристик.

Останнім часом в психологічних і освітніх вимірюваннях більш широко використовуються положення саме репрезентативної теорії вимірювань.

Репрезентативна теорія вимірювань була розроблена американським психологом Стенлі Стівенсом в 40-х роках 20 століття. Центральним поняттям цієї теорії є поняття *вимірювальної шкали*.

Слід розрізнити *об'єкти вимірювання*, їх *властивості* і *ознаки*. Останні виражають у тій чи іншій числовій шкалі властивості об'єктів.

В залежності від того, яка операція лежить в основі вимірювання ознаки, виділяють *шкали вимірювання*, які часто називають ще *рівнями вимірювання*. Тип шкали задає групу *допустимих перетворень* шкали. Допустимі перетворення не міняють співвідношень між об'єктами вимірювання, тобто після допустимих перетворень дані залишаються вимірними у шкалі того ж самого типу. Наприклад, при вимірюванні довжини перехід від футів до метрів не міняє співвідношень між довжинами даних об'єктів – якщо перший об'єкт довший другого, то це буде встановлено і при вимірюванні у футах, і при вимірюванні в метрах. При цьому чисельне значення довжини в футах відрізняється від чисельного значення довжини в метрах – не міняється лише результат порівняння довжин двох об'єктів.

Перш за все, шкали поділяються на метричні і неметричні (за іншою термінологією, шкали якісних ознак і шкали кількісних ознак, або шкали низького рівня і шкали високого рівня).

Шкала є *метричною*, якщо для неї може існувати одиниця вимірювання. Також шкали поділяються на *дискретні* та *неперервні*.

Типи шкал. Розглянемо основні чотири типи шкал.

1. *Номінальна шкала (шкала найменувань, категоріальна шкала)*. У шкалі найменувань допустимими є тільки взаємно-однозначні (тотожні) перетворення. Використання цієї шкали має на меті класифікацію об'єктів. В цій шкалі числа якщо й використовуються, то лише як мітки для розрізнення об'єктів. В шкалі найменувань виміряні, наприклад, номери телефонів, автомашин, паспортів, студентських квитків. Номери страхових свідоцтв державного пенсійного страхування, медичного страхування, штрих-коди товарів теж виміряні в шкалі найменувань. Стать людей також вимірюється в шкалі найменувань, результат вимірювання приймає два значення – чоловіча, жіноча. Раса, національність, колір очей, волосся – номінальні ознаки. Нікому не прийде в голову додавати або множити номери телефонів, такі операції не мають для номінальної шкали сенсу. Єдине, для чого годяться результати вимірювань в шкалі найменувань – для розрізнення об'єктів. У багатьох випадках цього достатньо. Наприклад, шафки для одягу в роздягальнях для дорослих розрізняють за номерами, тобто числами, а в дитячих садках використовують малюнки, оскільки діти ще не знають чисел.

2. *Порядкова (рангова) шкала*. Вище ми розглянули приклад з шафками для одягу у дорослій і дитячій роздягальнях. У тому випадку, коли ми вирішимо позначати шафки не малюнками, а числами, ми можемо закласти в ці позначення додаткову інформацію, яка буде допомагати швидше відшукати потрібну шафку. А саме, ми можемо впорядкувати розташування шафок за зростанням чи спаданням чисел, написаних на них (іншими словами, пронумеруємо шафки). Очевидно, зменшення невизначеності дозволяє стверджувати, що ми переходимо таким чином до шкали вищого типу.

У порядковій шкалі числа використовуються не тільки для розрізнення об'єктів, але і для встановлення порядку між об'єкта-

ми. Простим прикладом є оцінки знань. Символічно, що оцінки студентів за чотирибальною національною шкалою можуть виражатись як числами (2, 3, 4, 5), так і словесно – незадовільно, задовільно, добре, відмінно. Цим підкреслюється "нечисловий" характер оцінок. Не можна стверджувати, що учень, який отримав двійку, знає рівно вдвічі менше, ніж той, що отримав четвірку, або різниця між тим, хто отримав трійку і тим, хто отримав четвірку, дорівнює різниці у знаннях між тими, хто отримав відповідно четвірку і п'ятірку.

Оцінки експертів найчастіше слід вважати вимірними в порядковій шкалі, адже людина більш правильно (і з меншими утрудненнями) відповідає на питання якісного, наприклад, порівняльного, характеру, ніж кількісного. Так, наприклад, легше сказати, який з двох відрізків довший, ніж вказати їх довжину в міліметрах.

У різних областях людської діяльності застосовується багато видів порядкових шкал. Так, наприклад, в мінералогії використовується шкала Мооса, по якому мінерали класифікуються згідно критерію твердості. Так, згідно з цією шкалою, тальк має бал 1, гіпс – 2, а алмаз – 10. Мінерал з більшим номером є більш твердим, ніж мінерал з меншим номером, при натисненні дряпає його.

Порядковими шкалами в географії є – бофортівська шкала вітрів («штиль», «слабкий вітер», «помірний вітер» і т.д.), шкала сили землетрусів. Очевидно, не можна стверджувати, що землетрус в 2 бали (лампа гойднулася під стелею – таке буває і в нашій місцевості) рівно в 5 разів слабкіший, ніж землетрус в 10 балів (повне руйнування всіх споруд на поверхні землі).

У медицині є багато загальноприйнятих порядкових шкал – шкала стадій гіпертонічної хвороби, шкала опіків тощо. Нумери будинків також виміряні в порядковій шкалі – вони показують, в якому порядку стоять будинки уздовж вулиці. Нумери томів в зібранні творів письменника або нумери справ в архіві підприємства зазвичай пов'язані з хронологічним порядком їх створення, тому теж є прикладами порядкової шкали.

До порядкових шкал допустимими є перетворення, які не порушують порядку між результатами вимірювань, тобто будь-які монотонні перетворення, наприклад, логарифмування. Слід пам'ятати, що спадне перетворення змінює порядок розміщення чисел на шкалі на протилежний.

При статистичному опрацюванні порядкових даних відшукання такої характеристики, як середнє арифметичне, взагалі кажучи, не має сенсу.

Порядкова шкала і шкала найменувань є основними шкалами якісних ознак і відносяться до *неметричних шкал*. В багатьох конкретних областях науки і практики результати якісного аналізу можна розглядати як вимірювання за цими шкалами.

Шкали кількісних ознак (метричні шкали, або шкали вищого типу) – це, в основному, шкала інтервалів і шкала відношень.

3. *Шкала інтервалів*. У цій шкалі числа відображають не тільки відмінність між об'єктами за ступенем вираженості ознаки, а й *на скільки* більше або менше виражена ознака. Характерною особливістю цієї шкали, яка відрізняє її від іншої метричної шкали – шкали відношень, – є *відсутність природного початку відліку і природної одиниці вимірювання*. Дослідник повинен сам задати точку (початок) відліку і сам вибрати одиницю вимірювання.

Допустимими перетвореннями в шкалі інтервалів є будь-які лінійні перетворення, тобто перетворення виду $y = ax + b$. Наприклад, температурні шкали Цельсія і Фаренгейта зв'язані залежністю $^{\circ}\text{C} = 5/9 (^{\circ}\text{F} - 32)$, де $^{\circ}\text{C}$ – температура (в градусах) за шкалою Цельсія, а $^{\circ}\text{F}$ – температура за шкалою Фаренгейта.

Для інтервальної шкали величина середнього арифметичного має цілком визначений зміст.

4. *Шкала відношень*. З кількісних шкал найпоширенішими в природничих науках і у повсякденній практиці є шкали відношень. У них є *природний початок відліку* – нуль, який означає відсутність величини. В такій шкалі числа відображають не тільки відмінність між об'єктами за ступенем вираженості ознаки, не тільки те, на скільки більше або менше, а й *у скільки разів* більше або менше виражена ознака.

За шкалою відношень вимірюється більшість фізичних одиниць: маса тіла, довжина, заряд, а також ціни (і різні вартісні характеристики) в економіці. Тільки для інтервальної шкали результати вимірювань є числами в звичному значенні слова. Допустимим перетворенням для шкали відношень є *масштабування* (розтягування або стискування) виду $y = ax$.

Слід зазначити, що за іншою, більш детальною класифікацією, шкалу відношень відрізняють від так званої *абсолютної* шка-

ли. Відмінність цих шкал полягає в тому, що для шкали відношень не існує природної одиниці вимірювання, а для абсолютної шкали така одиниця існує. Прикладом віднесення до абсолютної шкали з природною одиницею вимірювання є кількість людей в кімнаті.

Перераховані шкали можна характеризувати за їх диференційовною (роздільною) здатністю (чи потужністю) – здатністю розрізнати об'єкти як відмінні один від одного.

Шкали по мірі зростання потужності розташовуються так: шкала найменувань, порядкова, інтервальна, відношень. Тобто неметричні шкали менш потужні – вони несуть менше інформації про відмінності між об'єктами.

Шість рівнів наукового усвідомлення. У процесі розвитку відповідної області знання тип шкали може мінятися. Так, спочатку температура вимірювалася за порядковою шкалою (холодніше – тепліше). Потім – за інтервальною (шкали Цельсія, Фаренгейта, Реомюра). Нарешті, після відкриття абсолютного нуля температуру можна вважати виміряною за шкалою відношень (шкала Кельвіна). Зауважимо, що серед фахівців іноді виникають розбіжності з приводу того, за якими шкалами слід вважати виміряними ті або інші реальні величини. Допомогти тут можуть запитання на зразок такого: «якщо сьогодні температура 0 градусів, а завтра буде удвічі холодніше, то якою буде завтра температура?».

Можна прослідкувати зв'язок між рівнями вимірювання і рівнями науковості мислення, слідуючи запропонованій Ч. Пірсом класифікації, яку він розглядав в рамках створеної ним семіотики – науки про знаки. Ця класифікація відображає еволюцію людського мислення – від перших «диких» спалахів уяви до репродуктивних кількісних законів, які є інструментами і вищими проявами науки.

Власне, класифікація знаків у Пірса налічує десять рівнів, але для нас є суттєвими вищі шість шаблів. Охарактеризуємо коротко кожен з них.

1. *Уявлення.* Перший рівень наукового усвідомлення – це просте інтуїтивне уявлення, спалах думки, «дика» гіпотеза. Це «зерна» творчості. Таке уявлення належить тільки даній особі, не може нею бути переосмислене чи передане комусь іншому. Пірс називає цей перший рівень обізнаності «можливою іконою» (possible icon).

2. *Думка*. Деякі уявлення вимагають осмислення. З'являються думки, які тепер можуть відтворюватися у нашому мозку неодноразово. Ми також можемо тепер ідентифікувати ті ж самі думки у інших осіб. Думка, таким чином, стає тим, до чого ми можемо повертатися і чим ми можемо обмінюватися з іншими людьми. Це цілком якісна характеристика. Eddington (1946) та Kinston (1985) називають цей рівень «суттю» (entity). Пірс називає це «можливим індексом» (possible index).

3. *Об'єкт*. Наступний крок – усвідомлення того, як ми можемо вказати на нашу думку. Ми присвоюємо нашій думці ім'я, і тепер ми можемо поводитися з нею, як з річчю, частиною світу, екземпляром нашої ідеї. Ми тепер можемо вказувати на ідею, і можемо підраховувати ідеї. На цьому рівні з'являється поняття кількості. Це той рівень, на якому може існувати перший, за Стівенсом, рівень (шкала) вимірювань, який називається номінальним чи категоріальним. На цьому рівні ми можемо віднести даний об'єкт до певного класу, наприклад, віднести особу до класу чоловіків чи до класу жінок. Клас може позначатися числом, але це тільки символ, який не має сенсу числа. Eddington (1946) та Kinston (1985) називають цей рівень «спостережуваним» (observable). Пірс називає це «дійсним індексом» (factual index).

4. *Порівняння*. Далі приходимо до ідеї надання переваги. Деякі речі можуть подобатися нам, інші – ні. Ми починаємо враховувати «кращість» чи «гіршість», присвоюючи їм відповідні числа.

Хоча ми ще не кажемо, на скільки щось краще за інше, чи більше за інше, чи важче за інше. Стівенс назвав це порядковим рівнем вимірювання. За Пірсом, це «можливий символ» (possible symbol). Саме тут починає з'являтися ідея подрібнюваності, квантування. Ми переходимо від підрахунку «солодких яблук» до ідеї «солодкості».

5. *Вимірювання*. Підрахунок – це початок вимірювань. Однак проста кількість не завжди приводить до того, що ми називаємо мірою. Якщо цеглини мають різні розміри, то ми не можемо сказати, скільки цеглин нам потрібно, щоб збудувати задумане. Три яблука не завжди будуть оцінені у ту ж саму ціну. На цьому рівні ми приходимо до вигадування абстрактної змінної, вздовж якої ми можемо відкладати рівні частки, чи інтервали.

Стівенс назвав цей рівень інтервальним вимірюванням. Пірс назвав це «дійсним символом» (factual symbol).

б. *Залежність*. Шостий, найвищий рівень – це рівень виникнення теорії. Що дає нам якась змінна, якщо вона не веде до іншої? Як змінні описують процес? Якби не існувало цього рівня наукового усвідомлення, то не існувало б і поняття статистичного аналізу даних вимірювання, не існувало би дослідження мінливості, кореляції, не існувало би поняття моделі і порівняння моделей.

Пірс називає це «символом, який вимагає доведення» (arguable symbol), а Кінстон – «пов'язуванім».

Описане співвідношення рівнів вимірювання з рівнями наукового усвідомлення свідчать про те, що вимірювання у якійсь обраній шкалі – це не просто «окремий спосіб» вимірювання. Кожен наступний рівень вимірювання ґрунтується на попередніх, включає їх у себе, є їх подальшим розвитком і вдосконаленням.

Формалізація основних понять репрезентативної теорії вимірювань. Назвемо *емпіричною системою з відношеннями* сукупність реальних (емпіричних) об'єктів, які цікавлять дослідника, з визначеною для них системою відношень. Прикладом емпіричної системи з відношеннями може бути група учнів певного класу, які є «носіями» знань з певного предмету, скажімо, фізики, причому для учнів визначені бінарні відношення виду «учень Б знає фізику краще, ніж учень А». Для кожної пари учнів питання про те, хто з них володіє більшими знаннями з фізики, має сенс, тобто є осмисленим. Емпірична система формується дослідником довільно, у відповідності з його уявленнями про реальність, яка вивчається.

Назвемо *математичною системою з відношеннями* сукупність математичних об'єктів (часто такими об'єктами є числа), з виділеними відношеннями між ними. Вимірювання полягає в приписуванні об'єктам емпіричної системи з відношеннями символів математичної системи з відношеннями за допомогою певних правил.

Приписування чисел об'єктам створює числову шкалу. Створення шкали є можливим, оскільки існує ізоморфізм емпіричних систем і систем дій, виконаних над реальними об'єктами. Розрізняють декілька типів числових систем з відношеннями і відповідно кілька типів шкал. Операції, а саме – способи вимірювання об'єктів, задають тип шкали. Шкала, в свою чергу характеризуєть-

ся видом перетворень, які можуть бути віднесені до результатів вимірювання. Якщо не дотримуватися цього правила, то структура шкали порушиться, а дані вимірювання не можна буде осмислено інтерпретувати. Тип шкали також однозначно визначає сукупність статистичних методів, які можуть бути застосовані для обробки даних вимірювання.

Шкала (лат. *scala* – сходи) у буквальному значенні є вимірювальний інструмент. Нехай A – емпірична система з відношеннями, R – числова система з відношеннями, f – функція, яка гомоморфно відображає A в підсистему R . Зазвичай R є системою дійсних чисел або її підсистемою, хоча це можуть бути і будь-які символи нечислової природи. Назвемо *допустимими перетвореннями шкали* такі перетворення, відносно яких шкала є інваріантною, тобто при застосуванні цих перетворень не змінює свій тип. Нехай G – група допустимих перетворень шкали. Назвемо *шкалою* упорядковану четвірку $\langle A; R; f; G \rangle$. Відповідно до сучасних уявлень, внутрішньою характеристикою шкали виступає саме група G , а f є лише прив'язкою шкали до конкретної ситуації вимірювання.

Формування теоретичного конструкту і процес вимірювання. Розглянемо на прикладі процес формування теоретичного психологічного конструкту і як це веде до вимірювання. Припустимо, що шкільний психолог спостерігає за поведінкою дітей під час їхніх ігор. Він помічає, що дехто з дітей постійно намагається управляти діями інших. Після спостережень, виконаних протягом тривалого часу і в різних ситуаціях, наш психолог може дати спеціальну назву такій поведінці, наприклад, «соціальне домінування». Цим самим він винайшов *теоретичний конструкт*, який проявляється у різних схожих за змістом способах поведінки дітей. Поняття конструкту вживається нами для позначення ряду схожих між собою типів поведінки (проявів психічної чи інтелектуальної сфери людини). Конструкти є «цеглинами» у загальній теорії поведінки людини. Теорія потрібна для того, щоб уміти передбачати і при необхідності впливати на поведінку особи.

Але винайдення конструкту ще не є його вимірюванням. Перед тим, як здійснювати вимірювання, потрібно визначити правила відповідності між множиною різних типів поведінки дітей і теоретичним конструктом. Цей процес називають *операційним визначенням*. У нашому прикладі, психолог повинен визначити, які типи

поведінки дітей в спостережуваних умовах є проявами домінування. Далі психолог повинен створити план отримання зразків таких типів поведінки в стандартних ситуаціях і фіксування цих зразків у деякому стандартизованому форматі для кожної дитини окремо. Цим самим психолог виготовляє інструмент для вимірювання соціального домінування, який ми будемо називати тестом. Взагалі кажучи, *тест* можна розуміти як стандартну процедуру отримання зразків поведінки в межах визначеної області. Якщо цією визначеною областю є навчальні досягнення, то тест має перевіряти *оптимальні дії* опитуваних, і їх спеціально налаштовують на якнайкраще проявлення досягнень під час тестування. Якщо досліджуваною областю є ставлення досліджуваних до чогось, чи їх інтереси, реакції на певні ситуації, то тест має перевіряти *типову поведінку*.

Процес отримання зразків поведінки досліджуваних може відбуватися у різних формах. В одному випадку це можуть бути записи спостережень, в іншому – використання приготованих заздалегідь переліків поведінки (тестових завдань). Вимірювання відбувається тоді, коли результати тестування за допомогою визначених правил переводяться у числа. Наприклад, вимірювання виникає під час підрахунку актів соціального домінування, які проявила дитина протягом 10-хвилинного спостереження за її поведінкою в спеціально створених умовах, або при підрахунку правильних відповідей на завдання тесту, який виконали учні класу на уроці хімії. Результатом вимірювання є висновок фахівця щодо рівня теоретичного конструкту, який проявляє кожен опитуваний. У випадку, коли таким фахівцем є педагог, передбачається, що після визначення результату вимірювання цей результат буде використаний ним (або самим опитуваним, чи батьками опитуваного) для прийняття *педагогічного рішення*.

Проблеми вимірювання психологічних конструктів. Оскільки психологічні конструкти – це абстракції, які не можуть оцінюватися безпосередньо (*є латентними*), створення інструментів вимірювання для них завжди пов'язане з рядом проблем. Виділимо п'ять основних проблем.

1. *Не існує одного універсального підходу до вимірювання певного конструкту.* Оскільки психологічний конструкт спостерігається лише через певні типи поведінки, то два теоретики можуть

запропонувати суттєво різні його операційні визначення. Нехай, наприклад, потрібно визначити здатність учня ділити числа у стовпчик. З цією метою можна дати учню ряд завдань на ділення у стовпчик; цієї ж мети можна досягти, якщо попросити учня пояснити, як потрібно виконувати ділення у стовпчик; можна також запропонувати йому знайти помилки у вирішених задачах. З різних операційних визначень випливають різні вимірювальні процедури.

2. *Психологічні вимірювання завжди базуються на обмежених вибірках спостережень.* Для нашого прикладу з діленням, ми не можемо дати учню розв'язати всі можливі задачі з цієї теми. Необхідно визначитися, скількох задач буде достатньо, щоб їх успішне розв'язання адекватно демонструвало вміння ділити у стовпчик і вся предметна область була охоплена.

3. *Вимірювання психологічних конструктів завжди супроводжуються помилками.* Навіть якщо дати одному учню один і той же набір задач, але в різний час, результати навряд чи будуть точно збігатися, оскільки дії учня завжди супроводжується такими важко передбачуваними ефектами як втома, неуважність, забудькуватість, спробами вгадування тощо. Ще більших розбіжностей у результатах слід очікувати, якщо один і той же конструкт вимірюється різними інструментами. Таким чином, завжди актуальною є питання про оцінку похибки вимірювання.

4. *Не існує природного нуля та одиниці вимірювання на обраній шкалі.* Чи означає нуль розв'язаних учнем під час тестування задач на ділення у стовпчик, що у нього відсутні знання з цієї теми? Нехай Володимир розв'язав 5 задач, Марія – 10, а Ольга – 15. Чи означає це, що відмінність у здатності ділити у стовпчик у Володимира і Марії така сама, як у Марії і Ольги? Практично завжди можна стверджувати, що первинні бали (тобто простий підрахунок розв'язаних завдань) не володіють всіма необхідними ознаками міри, і ці бали потрібно трансформувати за допомогою спеціальних процедур у обрану шкалу вимірювання.

5. *Психологічний конструкт не може визначатися лише в термінах операційного визначення, він також повинен проявлятися у зв'язках з іншими конструктами.* Цей другий рівень визначення конструкту дає змогу інтерпретувати результати вимірювання, і ця обставина має надзвичайно велике практичне значення.