

ОЛЕГ МАЗЯР

КУРС ЛЕКЦІЙ

«ОСНОВИ ПСИХОМЕТРИКИ»

**ДЛЯ ЗДОБУВАЧІВ ДРУГОГО (МАГІСТЕРСЬКОГО) РІВНЯ
ВИЩОЇ ОСВІТИ**

Мазяр О. В., Фещук В.В. Основи психометрики : методичні рекомендації до навчальних занять здобувачів другого (магістерського) рівня вищої освіти галузі знань 05 Соціальні та поведінкові науки спеціальності 053 Психологія. Житомир : Вид-во ЖДУ, 2022. 108 с.

Лекція № 1-2
ПСИХОМЕТРИКА ЯК ПРИКЛАДНА ДИСЦИПЛІНА
План

1. Предмет і завдання психометрії.
2. Історичні аспекти розвитку психометрії.
3. Психометрія на перехресті психологічної теорії та практики.
4. Шкала й типи шкал.

Мета: ознайомити із колом теоретичних і прикладних завдань, які вирішує сучасна психометрія.

Основні поняття: психометрія, психодіагностика, тести інтелекту, тести пам'яті, інтроспекція, теоретична модель, дослідницька модель, властивості чисел, психічна норма, шкала найменувань, шкала порядку, шкала інтервалів, шкала відношень.

Вчені: В. Вундт, Ф. Дондерс, І.О. Сікорський, Ф. Гальтон, Г. Еббінгауз, А. Біне, Т. Сімона, Дж. Кеттелл.

§ 1. ПРЕДМЕТ І ЗАВДАННЯ ПСИХОМЕТРІЇ

Психометрика, або психометрія (від грецького *psyche* – душа та *metron* – міра), є цариною знань теорії та практики вимірювання феноменів у галузі психології. Це прикладна галузь психології, яка спрямована на розробку вимірювальних процедур (тестів, опитувальників, анкет, проєктивних методик). Це галузь науки про такі властивості психологічних тестів як тип вимірювальних шкал, надійність, валідність, дискримінативність, внутрішня узгодженість тощо.

Предмет психометрії – організація, планування, розробка правил проведення й обробки результатів емпіричного дослідження у галузі психології. Це наука про умови та процес складання психологічних тестів, опитувальників, проєктивних методик.

Завданнями психометрії є

- аналіз оптимальних процедур дослідження того чи іншого психологічного феномену та розробка релевантного діагностичного інструментарію;
- формування основних принципів і правил проведення психологічного вимірювання та системної інтерпретації отриманих даних;
- виокремлення й адекватне застосування релевантних (відповідних) методів математичного аналізу.

Коротко, психометрія розробляє психологічний інструментарій, який дозволяє адекватно систематизувати відомості про актуальний стан психологічних феноменів. Це ключове завдання обумовлює постійний розвиток та удосконалення психометричних процедур. При цьому психометрія безпосередньо не продукує нової психологічної теорії, а здебільшого спрямована у практичному аспекті «обслуговувати» психологічну теорію, доводити її прикладну цінність, підтверджувати, коригувати та спростовувати науково-

психологічні гіпотези, що, своєю чергою, може сприяти виробленню нових концептуальних положень. Відповідно, психометрія забезпечує переважно завершальний акт становлення наукової теорії – *верифікацію* (перевірку істинності) теоретичних положень.

Науково підтверджені продукти психометрії стають надбанням практичної психодіагностики. Тож психодіагностику можна вважати психометрикою, яка довела свою дослідницьку цінність.

Актуальність вирішення психометричних завдань постає перед дослідниками при розробці оригінальних психологічних концепцій і теорій, які не вдається верифікувати за допомогою вже розроблених психометричних засобів. «Авторська психометрія» здебільшого представлена у дисертаційних дослідженнях.

Мета вимірювання – встановлення індивідуально-психологічних відмінностей індивіда шляхом зіставлення його показників із показниками решти людей. Це дозволяє пояснювати поведінку та прогнозувати розвиток індивіда. На першому етапі це робиться на підставі несистематизованого спостереження, інтуїтивно, «на око». Натомість психометрія розробляє низку правил, які покликані забезпечити обґрунтованість і коректність вимірювання та порівняння. Саме це дозволяє перейти на другий, власне науковий, етап оформлення психологічної теорії.

§ 2. ІСТОРИЧНІ АСПЕКТИ РОЗВИТКУ ПСИХОМЕТРІЇ

Психометрика стихійно розвивалася у донауковий період як спроба вичувати долю людини. Її праформами можна вважати хіромантію, астрологію, фізіогноміку, нумерологію, гадання, графологію. Перелічені праформи психометрики не отримали наукового підтвердження, дарма що залишаються популярними серед широких верств населення. На прикладі фізіогноміки можна побачити, що її положення потужно розроблялися у вітчизняній науці ще на початку ХХ століття (Гриньова Н. В. (Уклад), 2012; Коханова О. П. (Авт-уклад.), 2016; Мазяр, 2014б). Проблеми зв'язку почерку й особистості розробляються досі (Кайко, & Мацієвська, 2018).

У ХІХ столітті розпочинається науковий період становлення психометрики. Значний вплив на її розвиток справила щойно розроблений напрям *психофізики* (Fechner, 1882). Своїм основним завданням ця наукова галузь ставила вивчення співвідношення психічного (душі) та фізичного (тіла). Це, своєю чергою, передбачало зіставлення об'єктивних (числових) показників функціонування людини. У межах цих досліджень Г. Фехнером навіть був сформульований основний психофізичний закон, який вимірював інтенсивність відчуттів.

Окремої уваги заслуговують перші вимірювання часу психічного реагування. Зокрема, у працях Ф. Дондерса значущим є намагання дослідити не лише час простої реакції, а й складної, що передбачає врахування її когнітивних аспектів.

Епохальним для психологічної науки загалом став 1879 рік. По-перше, цього року була відкрита Лейпцігська лабораторія, з якою пов'язують зародження наукової психології, а її засновника В. Вундта називають «батьком психології». Той фактично запровадив системне психофізіологічне дослідження людини, для чого використовував спеціальні пристрої. По-друге, 1979 року розпочинаються суто психологічні дослідження, які організовували та проводили Г. Еббінгауз та І.О. Сікорський. Перший вивчав пам'ять, другий – втомлюваність (Мазяр, 2014б; Мазяр, 2015). Стимульним матеріалом для досліджень ставали беззмістовні склади та запам'ятовування незнайомих (іноземних) слів. Результатом дослідження, зокрема, стала крива забування Г. Еббінгауза.

Значущими виявились праці Ф. Гальтона, якого вважають **зачинателем психометрики**. Зокрема, він здійснив одну з перших спроб провести об'єктивне дослідження обдарованості за допомогою статистичного аналізу наявних даних (Тітов, 2018); також психологічна складова дослідження була інтегрована в його антропометричні вимірювання людей у лабораторії. Ним введені в науковий обіг поняття «коефіцієнт кореляції» (Galton, 1889), «вибірки досліджуваних» (Galton, 1902), «нормального розподілу даних» (Galton, 1907).

Перший серйозний психометричний продукт побачив світ на зламі століть, коли вийшла праця Дж. Кеттелла «Розумові тести та вимірювання». Саме дослідження інтелекту як психологічного феномена, що має чітку об'єктивізацію у продуктах розумової діяльності, стало початком бурхливого розвитку тестології. Французи А. Біне та Т. Сімона запропонували ідею вивчати інтелект як відношення біологічного та реального віку досліджуваного. Дещо пізніше ці положення втілилися у шкалі Стенфорд-Біне (Stanford-Binet Intelligence Scale), де була детально розроблена інструкція та система інтерпретації результатів (Bain, & Allin, 2005). Власне, з вимірюванням IQ асоціюється поняття психологічного тесту. Сьогодні психометрика значно розширила область свого застосування. Вона вийшла на рівень дослідження несвідомих аспектів психічного функціонування, особистісних якостей та психофізіологічних властивостей.

§ 3. ПСИХОМЕТРИЯ НА ПЕРЕХРЕСТІ ПСИХОЛОГІЧНОЇ ТЕОРІЇ ТА ПРАКТИКИ

Одна з ключових причин кризи психологічної науки полягає у тому, що запропоновані теорії та концепції не мали переконливого емпіричного підтвердження. Основним методом залишалася *інтроспекція*. Самоспостереженню як основній процедурі дослідження істотно бракувало систематизації рефлексій учених, об'єктивованих у кількісних показниках. Це необхідно було зробити, щоби мати підстави зіставляти висновки, до яких вони приходили.

Психометрика була покликана вирішити проблему емпіричного обґрунтування теоретичних положень й переведення психології у розряд наукових дисциплін. До цього часу вона залишалася «умоглядною» наукою, частиною філософії. Не дарма ж більшість вчених докризового періоду не бачили між філософією та психологією демаркаційної лінії. Доводилося покладатися лише на їхню авторитетну думку, яка підкріплювалася цікавими й подекуди влучними спостереженнями, але не об'єктивними статистично значущими даними.

Теорія пропонує ідею, положення, яке необхідно перевірити практично. Без відправної теоретичної точки розробляти діагностичний інструментарій неможливо. Своєю чергою, психометрика пропонує той матеріал, який шліфує теорію, змушує коригувати її окремі положення або приводить до формулювання нових припущень. У цьому плані психометрика стає ніби очима психологічної теорії.

Відповідно, розробці психодіагностичного інструментарію повинна передувати детально розроблена теоретична структура досліджуваного явища, що може об'єктивуватися у схематизованій *теоретичній моделі*. Співвідношення та зв'язки між структурними елементами у цьому випадку стають основою для структурної та змістовної побудови об'єктивного тесту (опитувальника). Останній по суті стає *дослідницькою моделлю*, що втілює послідовність та поетапність вивчення психологічного феномену. Однак при цьому теоретична і дослідницька модель не обов'язково та зрідка стають тотожними структурно й змістовно. Вони перетинаються у ключових аспектах.

§ 4. ШКАЛА Й ТИПИ ШКАЛ

Виміряти у психології означає присвоїти певному феномену числове значення і порівняти його з еталонним показником у межах психодіагностичного інструменту або окремої його шкали. Ці два показники пов'язані між собою математичним відношенням, яке стає еквівалентом психологічних відмінностей. Відтепер якісні відмінності можливо порівняти кількісно, що наближає дослідника до істинного розуміння індивідуальності.

Щодо отримуваних показників існує низка сутніх зауважень:

- 1) *міра числових показників* може бути різною в інструментах. Наприклад, один опитувальник може будуватися на засадах 3-бальних оцінок, інших – на 5-бальних або 7-бальних оцінках;
- 2) показники не можна вважати абсолютно точними, позаяк існують статистичні похибки;
- 3) *поняття нуля* (початку відліку) у психології відносно і почасти довільне, тобто не означає відсутність феномену;
- 4) *поняття одиниці вимірювання* здебільшого також умовне, довільне, що породжує проблему взаємозв'язку між одиницями вимірювання;

5) *поняття психічної норми* є нестабільним та потребує періодичної переоцінки. Прикладом можуть слугувати Діагностичне й статистичне керівництво з психічних розладів (Diagnostic and Statistical Manual of mental disorders), яке витримало 5 видань, та Міжнародна статистична класифікація хвороб і проблем, пов'язаних зі здоров'ям (International Classification of Diseases), яка витримала 11 видань.

Приписування числових значень психологічним феноменам відбувається за чіткими правилами. Дослідники виокремлюють три основних властивості чисел:

- *властивість ідентичності*, або фіксація тотожності, брак відмінностей у розвитку феномену (якщо досліджувані набирають однакові бали, ми вважаємо їх ідентичними у вияві певної психологічної характеристики);
- *властивість порядку*, або фіксація ієрархічного положення феномену (за результатами дослідження ми можемо прорангувати досліджуваних, розташувати їх у порядку вияву потужності психологічної характеристики);
- *властивість кількості*, або ступінь відмінностей феномену за величиною вияву (можемо відмітити ступінь відмінностей психологічної характеристики з-поміж показників решти досліджуваних).

Традиційно виокремлюють чотири типи шкал (Климчук, 2009; Furr, 2011):

1. **Шкала найменування**, або номінальна шкала вказує тільки на назву феноменів і має числову властивість ідентичності. Це базовий рівень вимірювання. Така шкала містить лише класифікацію певних об'єктів, натомість самі об'єкти жодним чином не порівнюються. Наприклад, перелік акцентуацій характеру, типів темпераменту.

2. **Шкала порядку**, або рангова шкала вказує на порядок розташування відносно інших об'єктів; має числові властивості ідентичності та порядку. Ранжування об'єктів відбувається за кількісними характеристиками, які почасти не відбивають якісні відмінності. Основним є порядок об'єкта у вибірці.

3. **Шкала інтервалів**, або інтервальна шкала, вказує на однакові проміжки між об'єктами та володіє числовими властивостями ідентичності, порядку та кількості. Відбувається фіксація кількості вимірюваної характеристики, що дозволяє здійснювати якісну інтерпретацію результатів. Наприклад, профіль особистості в опитувальнику.

4. **Шкала відношень**, або пропорційна шкала, вказує на ступінь прояву феномену та володіє всіма числовими відношеннями (ідентичності, порядку, кількості та наявністю абсолютного нуля). Прикладом може слугувати тест інтелекту, вимірювання сили нервової системи.

Питання та завдання для самоконтролю

У чому полягає ключове завдання психометрики?

Чим різняться психометрика та психодіагностика?

Дослідження якого психологічного феномену були покладені в основу тестології? З якої причини?

Що таке теоретична модель досліджуваного явища? Чим вона відрізняється від дослідницької моделі?

Які властивості нуля у психометриці?

Які властивості одиниці вимірювання у психометрії?

Наведіть приклади зрушень в уявленнях науковців про психічну норму.

Який тип шкал прийнятний для вимірювання рівня інтелекту? Свою відповідь обґрунтуйте.

Який тип шкал прийнятний для вимірювання особистісних якостей? Свою відповідь обґрунтуйте.

Тести для самоконтролю

- 1.** Наука про розробку психодіагностичного інструментарію –
а) палеопсихологія; б) психометрика; в) загальна психологія; г) юридична психологія.
- 2.** Одна з праформ психометрики –
а) хіромантия; б) психофізіологія; в) алхімія; г) астрономія.
- 3.** Властивостями психологічних тестів є
а) чесність та влучність; б) невідповідність теорії; в) надмірна суб'єктивність; г) надійність та валідність.
- 4.** Батьком психометрики є
а) В. Вундт; б) І.О. Сікорський; в) Ф. Гальтон; г) Г. Еббінгауз.
- 5.** Поняття «коефіцієнт кореляції» та «нормальний розподіл даних» у науковий обіг ввів:
а) В. Вундт; б) І.О. Сікорський; в) Ф. Гальтон; г) Г. Еббінгауз.
- 6.** А. Біне та Т. Сімона запропонували методологію дослідження:
а) пам'яті; б) інтелекту; в) втомлюваності; г) часу реакції.
- 7.** Ф. Дондерс запропонував методологію дослідження:
а) пам'яті; б) інтелекту; в) втомлюваності; г) часу простої та складної реакції.
- 8.** Методологію дослідження пам'яті першим запропонував:
а) В. Вундт; б) І.О. Сікорський; в) Ф. Гальтон; г) Г. Еббінгауз.
- 9.** Методологію дослідження феномену втомлюваності 1879 року запропонував:
а) В. Вундт; б) І.О. Сікорський; в) Ф. Гальтон; г) Г. Еббінгауз.
- 10.** Закономірності процесу забування (крива забування) встановив:
а) В. Вундт; б) І.О. Сікорський; в) Ф. Гальтон; г) Г. Еббінгауз.
- 11.** Досліджуване явище, яке представлене у структурно-схематизованому вигляді називається:
а) теоретичною моделлю; б) епіфеноменом фізичної реакції; в) малюнком вченого; г) апробацією дослідження.
- 12.** Дослідницька модель спрямована поетапно вивчати:
а) математичну ймовірність отриманих результатів; б) актуальність планування дослідження; в) психологічний феномен; г) реєстр науково-психологічних праць з теми.
- 13.** Одиниця вимірювання у психодіагностиці є величиною
а) абсолютною; б) довільною; в) уніфікованою; г) безумовною.

- 14.** Поняття психічної норми у психології вважається
а) умовним; б) непорушним; в) безальтернативним; г) непохитним.
- 15.** Властивість числа, яка виражається у відсутності відмінностей, позначає:
а) ідентичність; б) дискретність; в) порядок; г) кількість.
- 16.** Властивість числа, яка виражається у показі ієрархічного положення, позначає:
а) ідентичність; б) дискретність; в) порядок; г) кількість.
- 17.** Властивість числа, яка виражається у ступені або рівні відмінностей, позначає:
а) ідентичність; б) дискретність; в) порядок; г) кількість.
- 18.** Шкала найменування володіє такими числовими властивостями як
а) ідентичність; б) порядок; в) кількість; г) наявність нуля.
- 19.** Абсолютний нуль представлено тільки у шкалі
а) найменування; б) порядку (ранжування); в) інтервалів; г) відношень (пропорцій).
- 20.** Шкала, що володіє всіма фундаментальними числовими властивостями, –
а) найменування; б) порядку (ранжування); в) інтервалів; г) відношень (пропорцій).

Лекція № 3
ОРГАНІЗАЦІЙНИЙ ЕТАП У РОЗРОБЦІ
ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТАРІЮ

План

1. Психологічне випробування та емпіричне вимірювання.
2. Вибірка респондентів.
3. Статистичні процедури при розробці психологічного інструментарію.
4. Планування проєкту психологічного вимірювання.

Мета: ознайомити з плануванням та організаційними аспектами розробки оригінального психодіагностичного інструментарію.

Основні поняття: випробування у психологічній практиці, емпіричне вимірювання, типи тестів, об'єктивні тести, опитувальники, проєктивні методики, вибірка досліджуваних, репрезентативна вибірка, середнє арифметичне, мода, медіана, дисперсія, стандартне відхилення, субнормальна показники, супернормальні показники, коефіцієнт кореляції, графік нормального розподілу, розмах даних.

§ 1. ПСИХОЛОГІЧНЕ ВИПРОБУВАННЯ ТА ЕМПІРИЧНЕ ВИМІРЮВАННЯ

Випробування у психологічній практиці – це штучне створення дослідницької ситуації, в якій суб'єкт потенційно може виявити певну обізнаність, зорієнтованість, здібності, навички, характеристики. Результати випробування являють собою зіставлення показників людей, встановлення індивідуальних і типологічних відмінностей.

Назагал мета випробування – систематизоване визначення психологічних характеристик індивіда, які показують його відмінності (Furr, 2011; Cronbach, 1960).

Емпіричне вимірювання у психології – ідентифікація певних психологічних характеристик індивіда, які спостерігаються або рефлексуються самим суб'єктом. Йдеться про систематизовану фіксацію властивих суб'єкту форми поведінки, когнітивних стратегій, емоційних реакцій тощо.

Емпіричне вимірювання передбачає чітко визначену процедуру збору психологічних характеристик індивіда. В залежності від цього розробляються різні **типи тестів**.

За предметом дослідження прийнято розрізняти: 1) тести досягнень; 2) власне психологічні тести: інтелектуальні тести; тести здібностей; соціально-психологічні опитувальники; особистісні опитувальники. За іншою класифікацією виокремлюють 1) тести вільні, структуровані, шкальовані, графічні та з пробілами; 2) тести з оцінювання об'єкту; 3) особистісні тести; 4) тести здібностей; 5) соціально-психологічні тести. За класифікацією P. Kline

(2000) існують 1) об'єктивні тести; 2) проєктивні тести; 3) опитувальники мотивації; 4) шкали настрою та станів; 5) вимірювання установок.

Спираючись на наявні класифікації психодіагностичного інструментарію, певно, найбільш сутнісним критерієм поділу можна вважати *емпіричний формат психологічної оцінки* досліджуваного феномену:

- об'єктивні випробування (**тести**) – здібності індивіда;
- самозвіти (**опитувальники**) – структуровані самохарактеристики суб'єкта;
- інтерпретаційні методи (**проєктивні методики**) – тлумачення несвідомих аспектів розвитку особистості.

Відповідно до цієї класифікації, *тести* покликані порівнювати людей за об'єктивними показниками (наприклад, тип нервової системи, інтелектуальні здібності); *опитувальники* систематизують усвідомлювані суб'єктом самооцінки (наприклад, особистісні опитувальники); *проєктивні методики* виходять за межі свідомості суб'єкта та прагнуть виявити логіку його несвідомого (наприклад, графічні методики). Власне, така послідовність відбиває складність розробки психологічного інструментарію та характеризується зниженням рівня його валідності та надійності (табл. 1).

Необхідно зауважити, що тести та опитувальники активно використовуються в межах практичної психології, а проєктивні методики – переважно у психологічній практиці. Відмінність між практичною психологією та психологічною практикою полягає у тому, що перша зорієнтована на вивчення дискретних психічних характеристик і встановлення психічної норми та відхилень у рідній академічній психології; друга – на «пізнання психологічних закономірностей в умовах взаємин «тут і тепер», які спонтанно виникають, на розуміння закономірностей цих взаємин та діагностику внутрішніх причин конкретної поведінки суб'єкта, першопричин його особистісної проблеми» (Яценко, 1996, С. 11). Це, з-поміж іншого, означає, що розробка проєктивних методів вимагає від дослідника ґрунтовної теоретичної підготовки з глибиннопсихологічних концепцій особистості, що виходить за межі академічної підготовки.

Таблиця 1
Властивості основних типів
психодіагностичного інструментарію

Тип психодіагностичного інструменту	Область дослідження	Предмет дослідження	Ступінь надійності та валідності інструменту	Сфера застосування
тести, або об'єктивні випробування	здібності	знання, вміння, навички	порівняно висока	практична психологія
опитувальники, або самозвіти	свідоме	самопрезентація	порівняно середня	практична психологія
проективні методики, або інтерпретаційні методи	несвідоме	логіка несвідомого	порівняно низька	психологічна практика

Зауважимо, що в межах кваліфікаційної роботи більш доречною є розробка тестів та опитувальників, а не проєктивних методик. Адже перші будуються на положеннях академічної психології, яка має вищий рівень верифікації у порівнянні з неакадемічними напрямками психології. Натомість показники проєктивних методик слабо піддаються кількісному (статистично-математичному) аналізу, що знижує ступінь їхньої доказовості, обґрунтованості.

§ 2. ВИБІРКА РЕСПОНДЕНТІВ

Розробка психодіагностичного інструментарію відбувається з урахуванням її стандартизації, що передбачає проведення дослідження на репрезентативній вибірці. Вибірка вважається **репрезентативною**, якщо має високий ступінь відповідності генеральній сукупності (Климчук, 2009). Тобто ця вибірка повинна відбивати основні характеристики генеральної сукупності. Наприклад, вивчення ефективності мотиваційного тренінгу, призначеного для роботи зі школярами, передбачає перевірку ефективності з учнями, які мають *пропорційну* представленість за рівнем академічної успішності, статевого і вікового розподілу та інших характеристик, які можуть здійснювати вплив на розвиток мотиваційної сфери школярів. Основна задача на етапі формування репрезентативної вибірки знання пропорцій.

В окремих випадках (наприклад, при розробці тесту інтелекту) доцільно вибірку диференціювати на підгрупи, щоби вивчати діагностичну цінність інструменту на окремих специфічних групах. У цьому випадку слід знаходити найбільш важливі змінні, які, власне, визначають специфіку групи. Причому за будь-яких умов варто прагнути до збільшення кількості обстежених. Якщо групи суттєво різняться між собою, необхідно розробити математичну процедуру нормування показників – переведення так званих сирих балів у стени або станайни, що дозволить коректно порівнювати групи між собою (Kline, 2000).

Якщо вибірка не є репрезентативною, тоді для перевірки надійності психодіагностичного інструментарію необхідно проводити дослідження на великих вибірках.

§ 3. СТАТИСТИЧНІ ПРОЦЕДУРИ ПРИ РОЗРОБЦІ ПСИХОЛОГІЧНОГО ІНСТРУМЕНТАРІЮ

Мета психологічного дослідження – встановлення індивідуальних або типологічних відмінностей. Ці відмінності можуть існувати на інтеріндивідуальному рівні (між окремими індивідами) та інтраіндивідуальному рівні (між показниками одного індивіда впродовж певного часу, по суті психодинамічні відмінності).

Статистика говорить за те, що переважна більшість людей має середні показники за всіма психічними феноменами. Власне, це продиктовано вимогами процесу соціальної адаптації. Зрозуміло, що ця обставина не виключає ймовірності численних випадків невідповідності аж до рівня явних психічних акцентуацій. Можливо, саме тому серед статистичних показників середнє арифметичне використовують значно частіше, ніж *моду* (значення, яке у множині стрічається найбільш часто) та *медіану* (значення, яке ділить множину навпіл). Ані мода (M_0), ані медіана (M_d) не здатні задовільним чином вказати на типовість, звичайність, пересічність, нормальність психологічного феномену. Пов'язано це, вочевидь, з розмитістю та варіабельністю психічної норми, яку краще сприймати не в точних показниках, а в діапазоні. Мода та медіана конкретні, натомість **середнє арифметичне** – саме така міра центральної тенденції, яка за замовчуванням передбачає варіабельність, достатньо широке коливання показників. Мода та медіана, безумовно, можуть виявити свою доречність, але до них постійно будуть зауваження дослідників. Натомість середнє арифметичне не претендує на точність, а, навпаки, інтегрує невраховані модою та медіаною оцінки досліджуваного.

Середнє арифметичне обраховується таким чином:

$$\bar{x} = \frac{\sum x}{N}, \text{ де } \sum x \text{ – сума всіх оцінок, а } N \text{ – кількість всіх випадків.}$$

Однак мета розробки психодіагностичного інструментарію полягає у здатності виявляти відмінності, тобто відхилення від середнього арифметичного та від оцінок окремих респондентів. Це дозволяє з'ясувати дисперсія та стандартне відхилення.

Дисперсія – це відхилення кожної індивідуальної оцінки від середнього арифметичного у групі досліджуваних. Дисперсія розраховується за формулою:

$$s^2 = \frac{\sum_{i=1}^n (X - \bar{X})^2}{n}$$

де s^2 – це вибіркова дисперсія, X – окремі значення, \bar{x} – середнє арифметичне, n – кількість випадків.

Стандартне відхилення вираховується як квадратний корінь з дисперсії: $s = \sqrt{s^2}$. Відмінність дисперсії та стандартного відхилення полягає у тому, що

другий показник дозволяє оцінити відхилення у сирих (первинних) балах, позаяк дисперсія є квадратом цієї величини, а відтак спотворює розуміння.

Яким чином треба інтерпретувати дисперсію та стандартне відхилення? По-перше, ці показники непрямо вказують на розмах даних (різниця максимальної та мінімальної оцінки у множині, що можна окремо порахувати: $R = X_{\max} - X_{\min}$). Розмах даних вказує на діапазон відмінностей у ряду, а відтак визначає підстави для точності висновків. Занадто широкий розмах даних може вказувати на необхідність внесення коректив у пункти та завдання психодіагностичного інструмента.

По-друге, інтерпретація цих показників повинна відбуватися у конкретному дослідницькому контексті. Більш інформативною вона стає у випадку порівняння з іншим показником дисперсії.

Втім, більш «робочим» математичним показником вважають **коефіцієнт кореляції Пірсона (r)**, який вказує досліднику на тісноту лінійного взаємозв'язку між двома вибірками. Такими вибірками можуть бути оцінки за двома методиками або показники однієї методики, які встановлені через певний проміжок часу (наприклад, при встановленні ретестової надійності методики). Власне, коефіцієнт кореляції стає основним інструментом вимірювання валідності та надійності розробленого діагностичного інструментарію.

Систематизація емпіричних показників передбачає нормування, яке має своїм наслідком стандартизацію психодіагностичного інструментарію. Введення критерію норми дозволяє у кількісних показниках визначити ступінь вияву досягнень або розвитку психічних якостей у порівнянні з рештою. Нормування дозволяє адекватно використовувати інструмент, максимально знизити фактор суб'єктивного тлумачення результатів дослідження.

Поняття норми формується на підставі статистичних показників, які встановлюються на базі репрезентативної або рідше генеральної вибірки. Її сутність складають показники *середнього арифметичного* та *дисперсії* (відхилень від середнього). При цьому еталонним вважається графік нормального розподілу даних (рис. 1), коли більшість оцінок відбивають середні кількісні показники, а супернормальні (вище середнього з тенденцією до максимального) та субнормальні (нижче середнього з тенденцією до мінімального) показники становлять меншість.

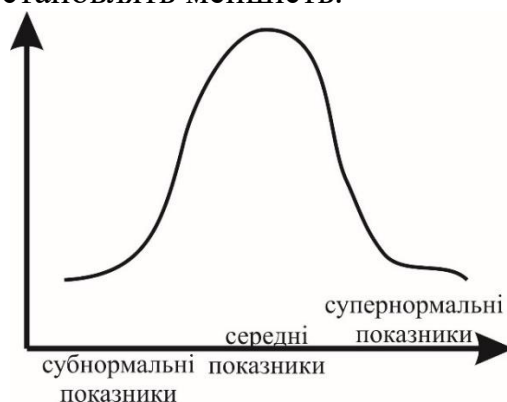


Рис. 1. Закон нормального розподілу (крива Гаусса)

Певне уявлення про відповідність вибірки нормам репрезентативності можна зробити на підставі мір середньої тенденції. У найбільш оптимальному варіанті мода, медіана та середнє арифметичне мають збігатися у репрезентативних вибірках.

Орієнтація на норму у кількісних показниках представлена у тестах та опитувальниках (відповідно до кривої Гауса), значно меншою мірою – у проєктивних методиках. Останні зорієнтовані на якісну інтерпретації даних, на встановлення індивідуальних відмінностей.

§ 4. ПЛАНУВАННЯ ПРОЄКТУ ПСИХОЛОГІЧНОГО ВИМІРЮВАННЯ

У циклі статей М.О. Батурин та Н.М. Мельникова пропонують такий алгоритм підготовки до розробки психодіагностичного інструментарію:

визначення попередніх характеристик діагностичного інструментарію ⇒ визначення основних джерел валідності ⇒ вибір психометричної та дослідницької моделі ⇒ складання детального плану розробки інструментарію ⇒ організаційні рішення.

Важливість підготовчого етапу продиктована визначенням загального вектору психометричної роботи. Помилки можуть суттєво позначитися на кінцевих результатах, адже може, приміром, виявитись брак валідності психодіагностичного інструменту, нерелевантність теоретичної та дослідницької моделі. Назагал центральна проблема валідності починає вирішуватися саме на організаційному етапі. Шлях до встановлення задовільного рівня математичної кореляції забезпечується саме тоді, коли інструмент лише замислюється та проєктується.

Насамперед маємо з'ясувати **мету** запланованого психодіагностичного інструмента. Що він дозволить з'ясувати? Чим вигідно вирізнятиметься з-поміж решти подібних інструментів? Чи не буде така розробка фабрикацією того, що вже існує? Відповідно, постає питання про джерела валідності.

Сфера застосування інструмента: його діагностичне призначення (для чого створений) та діагностичні обмеження (що не може виміряти).

Контингент, який ставатиме об'єктом дослідження (статевовікова та професійна специфіка, особливості психічного, психофізичного розвитку тощо).

Тип діагностичного інструменту з урахуванням поставлених задач: тест, опитувальник чи проєктивна методика? При цьому важливо одразу планувати формат отриманого продукту, щоби зіставити його з дослідницькими цілями, які можуть бути науковими та прикладними (ступінь релевантності). Зокрема, наскільки зручно буде користатися даними проєктивної методики у науковому дослідженні? Чи можна згрупувати показники проєктивних малюнків, щоби

показати специфіку сформованості емоційного феномену певної вибірки досліджуваних?

Структура діагностичного інструменту: кількість та назви шкал, які вимірюватимуть окремі властивості досліджуваного феномену. Важливо передбачити навантаження шкал (кількість пунктів і завдань), щоби планувати математичні процедури з переведення «сирих» балів у стени або станайни.

Важливо враховувати **психогігієнічні умови** дослідження: чи здатні досліджувані конкретної вікової чи професійної групи виконати заплановану кількість завдань за певний час, чи не будуть вони занадто легкими або складними у зв'язку зі специфікою їхнього когнітивного розвитку тощо. При цьому не слід забувати також про **етичні аспекти** дослідження, які полягають у дотриманні етичних норм при формулюванні питань, тверджень та завдань психодіагностичної методики. Йдеться про расові, національні, гендерні відмінності, особливості сексуальної орієнтації та сексуальної самоідентифікації.

Далі варто скласти **послідовний реєстр дій**, який дозволить планомірно розробити діагностичний інструмент і перевірити його валідність та надійність, дискримінативність та внутрішню узгодженість пунктів (завдань). Його має вирізняти послідовність, несуперечливість окремих етапів, що може нашкодити його розробці.

Питання та завдання для самоконтролю

Що таке випробування у психологічній практиці?

Чим відрізняється емпіричне вимірювання від тривіального спостереження за індивідом чи групою?

Які сутнісні відмінності тестів досягнень (здібностей) та особистісних опитувальників?

Який психодіагностичний інструментарій здебільшого використовується у практичній психології та психологічній практиці?

На підставі яких показників формуються тестові норми?

На що вказують показники графіку нормального розподілу даних (крива Гауса)?

Який тип психодіагностичного інструментарію зазвичай має вищий ступінь валідності та надійності? Чому?

Якою буде репрезентативна вибірка для розробки опитувальника, який вимірює стан стомленості школярів після робочого дня?

Які переваги підрахунку середнього арифметичного, моди та медіани?

Чим відрізняється дисперсія від стандартного відхилення?

Що дозволяє з'ясувати коефіцієнт кореляції Пірсона?

Які основні джерела валідності психодіагностичного інструмента можна виокремити?

Тести для самоконтролю

- 1.** Встановлення психологічних відмінностей індивіда або групи є
а) психокорекцією; б) психологічною профілактикою; в) випробуванням; г) психологічною просвітою.
- 2.** Ідентифікація психологічних якостей відбувається в межах
а) біографічної довідки; б) емпіричного вимірювання; в) медичного діагнозу; г) атестату академічної успішності.
- 3.** Вимірювання «особистісного профілю» індивіда відбувається за допомогою
а) особистісних опитувальників; б) тестів інтелекту; в) тестів здібностей; г) вимірювання установок.
- 4.** Когнітивні здібності індивіда вимірюються за допомогою
а) особистісних опитувальників; б) тестів інтелекту; в) соціально-психологічних тестів; г) вимірювання установок.
- 5.** Мотиваційні аспекти розвитку особистості досліджуються
а) особистісними опитувальниками; б) тестами інтелекту; в) тестами здібностей; г) мотиваційними опитувальниками.
- 6.** Тип нервової системи індивіда вивчається за допомогою
а) опитувальників; б) тестових процедур; в) проєктивних методик; г) вимірювання життєвих установок.
- 7.** Несвідома сфера психіки вивчається за допомогою
а) проєктивних методик; б) опитувальників; в) тестів; г) анкетування.
- 8.** Здібності індивіда досліджуються шляхом
а) вивчення самозвітів; б) тестування; в) несистематизованого спостереження; г) анкетування.
- 9.** Найвищий ступінь надійності та валідності діагностичного інструментарію зазвичай виявляють
а) тести; б) особистісні опитувальники; в) графічні методики; г) анкети з відкритими питаннями.
- 10.** Тестові норми ґрунтуються на показниках
а) середнього арифметичного та дисперсії; б) медіани; в) розмаху даних; г) моди.
- 11.** Якісні, а не кількісні показники лежать в основі норм
а) тестів інтелекту; б) особистісного опитувальників; в) проєктивних методик; г) анкет із закритими питаннями.
- 12.** Графік нормального розподілу даних на графіку відповідає кривій
а) Купера; б) Стьюдента; в) Фішера; г) Гауса.
- 13.** Показники за результатами тестування, які є нижче середнього арифметичного, коректно називають
а) нормальними; б) супернормальними; в) субнормальними; г) ненормальними.
- 14.** Показники за результатами тестування, які є вище середнього арифметичного, коректно називають
а) нормальними; б) супернормальними; в) субнормальними; г) ненормальними.

- 15.** Якщо вибірка відбиває основні характеристики генеральної сукупності досліджуваних, її називають
- а) репрезентативною; б) нерепрезентативною; в) рандомною; г) ненормальною.
- 16.** Значення, яке ділить множину значень навпіл, називається
- а) модою; б) медіаною; в) середнім арифметичним; г) стандартним відхиленням.
- 17.** Значення, яке у множині значень зустрічається найбільш часто, називається
- а) модою; б) медіаною; в) середнім арифметичним; г) дисперсією.
- 18.** Стандартне відхилення обчислюється як корінь квадратний з
- а) моди; б) медіани; в) середнього арифметичного; г) дисперсії.
- 19.** Різниця між максимальними та мінімальними значеннями у множині значень називається
- а) розмахом значень; б) стандартним відхиленням; в) середнім арифметичним; г) дисперсією значень.
- 20.** Тіснота зв'язку між двома рядами даних вимірюється за допомогою
- а) середнього арифметичного; б) коефіцієнта кореляції; в) встановлення розмаху значень; г) медіани.

Лекція № 4

ЗМІСТ

ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТАРІЮ

План

1. Вихідний теоретичний концепт і феноменологія досліджуваного явища
2. Системотворча дефініція.
3. Теоретична модель.
4. Припущення щодо структурних зв'язків.

Мета: ознайомити з теоретичними передумовами розробки оригінального психодіагностичного інструментарію.

Основні поняття: вихідний концепт, теоретична модель, дефініція, принцип ієрархії, принцип інваріантності, принцип координації та взаємодії, принцип індивідуальної єдності.

§ 1. ВИХІДНИЙ ТЕОРЕТИЧНИЙ КОНЦЕПТ І ФЕНОМЕНОЛОГІЯ ДОСЛІДЖУВАНОГО ЯВИЩА

Будь-якій діагностичній роботі передуює робота теоретична. Власне, психодіагностика, дарма що має своє виразне прикладне значення, передовсім «обслуговує» теорію, підтверджуючи, коригуючи або спростовуючи її положення. Психодіагностика не повинна проводитися «просто так». В її «послугах» має виникнути нагальна потреба. Психодіагностика має стати наступним кроком після теоретичної розробки психологічної проблеми, або «теоретизації предмета дослідження» (Киричук & Роменець (Ред.), 2006, С. 49).

Кожен психодіагностичний інструмент спрямований на вивчення конкретного феноменологічного поля – психологічного стану, якості, явища, процесу. Розробник ставить перед собою питання «Що вивчаємо?». Відповідь повинна бути чіткою та вичерпною, у згорнутому варіанті, з визначенням вихідного концепту. **Вихідний концепт** дослідження, за влучним визначенням М.О. Батурина та Н.М. Мельникової, можна вважати початковою точкою конструювання тесту та наріжним каменем у фундаменті майбутньої будівлі валідності.

Завдання (пункти) психодіагностичного інструмента мають відбивати психологічний феномен у штучно створеній дослідником ситуації. Наприклад, в об'єктивному тесті йтиметься про задачу, яку суб'єкт повинен вирішити (вивчатимуться когнітивні здібності, стратегії і тактики інтелектуальних дій); в опитувальнику – питання і твердження, які стосуватимуться ситуацій, в яких суб'єкт має певним чином відреагувати (актуалізація особистісних рис на підставі декларації індивіда). При цьому необхідно вивчати всі можливі аспекти психологічного феномену, забезпечити змістовну повноту дослідження. У протилежному випадку діагностика виявиться однобічною. Власне, адекватність та повнота змісту забезпечують валідність психодіагностичної методики.

Феноменологія неодмінно повинна певним чином об'єктивуватися – в емоційних реакціях, поведінці, думках, способах рішення, часі виконання завдання, виборі правильного рішення з-поміж запропонованих, кількості спроб тощо. Розробник методики ставить перед собою питання «*Як вивчаємо?*». Йдеться про набір діагностичних процедур, які релевантно виявлятимуть феномен. Це стає змістом методики, наповнює її завдання. Задача розробника – адекватно зіставити таку об'єктивацію з формою завдань (пунктів). Наприклад, інтелект не може вивчатися шляхом декларацій суб'єкта щодо власного інтелекту у форматі опитувальника («Ви завжди правильно вирішуєте задачі на кмітливість?»), а тільки шляхом виконання комплексу когнітивних завдань різного ступеня складності, які кмітливість перевірятимуть.

Найбільш важливий момент полягає у тому, що попередньо у розробника повинна існувати *гіпотеза* (припущення, істинність якого вимагає доведення) або *система теоретичних положень*, з яких він виходитиме при складанні діагностичного інструменту. Якщо йдеться про складання особистісного опитувальника, то попередньо повинна існувати розроблена автором або іншими вченими теорія або концепція особистості (її структура, класифікація особистостей). Виходячи з цих положень та змісту теорії, розробник особистісного опитувальника складатиме його пункти, об'єднані у шкали. При цьому шкали відповідатимуть структурним елементам феномену.

§ 2. СИСТЕМОТВОРЧА ДЕФІНІЦІЯ

Дефініцію можна представити як «коротке визначення змісту якогось поняття» (Бочелюк & Бочелюк, 2008, С. 336). Однак ця скороченість не тотожна спрощеності, поверхневості. В ній мають відбиватися основні *системні властивості поняття*, які вирізняють його з-поміж синонімічних. Авторська дефініція закладає основу для наукової новизни дослідження, що виявляється у системному розумінні досліджуваного феномену, що не виключає чималої кількості уточнень (нюансів). Останнє може бути представлено більш розгорнуто в окремих публікаціях. Користувачі психодіагностичним інструментом мають бути поставлені до відома щодо цього.

Коротше, йдеться про змістовне наповнення центрального поняття в авторському розумінні. Проблема полягає у тому, що існує достатньо велика кількість дефініцій одного й того самого поняття, які можуть різнитися у суті та нюансах. Відтак не кожна теорія чи концепція зможе однаково успішно користатися розробленим психодіагностичним інструментом. Причому мова йде як про класичні сфери дослідження (інтелект, особистість), так і про відносно нові (емоційний інтелект, цінності).

Дефініція і система вихідних теоретичних положень, яка сформульована розробником діагностичного інструменту, повинна вказувати на релевантність його використання. У цьому місці ми вимушені зауважити на певній суперечності: з одного боку, дефініція повинна володіти властивістю наукової новизни (у протилежному випадку це буде явним або прихованим плагіатом на

рівні ідеї), а з іншого – її повинні прийняти фахівці, які підготовлені на іншому теоретико-методологічному матеріалі. Відтак повна релевантність стає недосяжною. До результатів того чи іншого психодіагностичного інструменту спеціалісти ставитимуться з деякими поправками, теоретичними компромісами, недовірою або скепсисом.

У зв'язку з цим корисною стає розробка **теоретичної моделі** вихідного концепту. Це дозволить спеціалістам в економному режимі (схематично) ознайомитися з відмінностями у розумінні вихідних концептів, частково погодитися з ними чи заперечити їх і вважати неприйнятними для власного користування.

§ 3. ТЕОРЕТИЧНА МОДЕЛЬ

Власне, будь-яка теорія передбачає можливість її формалізації, що об'єктивується певною моделлю. Неможливість змоделювати теоретичні положення є вагомим ознакою неадекватності теорії. **Наукову теорію** (іншої, до слова, бути не може) слід розуміти як редуцію випадкових характеристик, ознак і, відповідно, *систематизацію сутнісних характеристик*. Системні характеристики можуть бути представлені у форматі *взаємопов'язаних структурних елементів*. Це може бути представлено як структурно замкнена система елементів (ланок), яка залишається відкритою для змін.

Теоретичне моделювання передбачає ідеалізацію об'єкта, його заміну копією, яку стає можливо вивчати з теоретичного боку і, як наслідок, планування емпіричних досліджень. Зрозуміло, що теоретична модель психологічного феномену не може змістовно вичерпати оригінал, але до цього повсякчас варто прагнути, досліджуючи різні його сторони.

Таким чином, перед розробником психодіагностичного інструменту постає задача створити таку *структурно-функціональну модель*, яка допоможе оцінити досліджуваний феномен за певною схемою. В ній мають бути представлені основні категорії (структурні елементи) та функціональні зв'язки між ними.

Наприклад, якби І.О. Сікорський більше, як сто років тому запланував розробити психодіагностичний інструментарій для вивчення особистості¹, він міг би спиратися на таку реконструйовану модель (рис. 2). У цій моделі структурно представлені основні чинники, які визначають розвиток особистості, їхні складові, взаємне розташування (субординація), механізми формування індивідуальності (Мазяр, 2013).

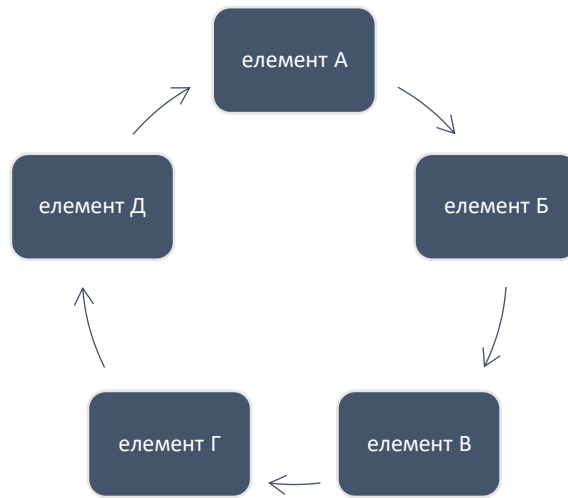
Власне, структурні елементи (ланки) теоретичної моделі можуть ставати окремими шкалами або факторами діагностичного інструменту, що дозволяє отримати диференційований погляд на досліджуваний феномен. В результаті це може бути представлено як характерологічний (особистісний) профіль.

¹ На той час психодіагностика тільки робила перші кроки у розробці тестів інтелекту, а технології розробки особистісних опитувальників не існувало.

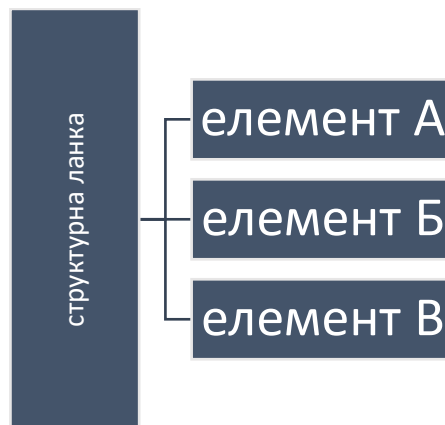
волю рівні ікі	Групова належність суб'єкта	Структу рні компоненти особистості	Структурні елементи особистості		М ми роз суб'
Свідоме	Професія Суспільн о-політичні погляди Релігійна ідентичність	Характер	Самосвідомість		Т актив
			Воля як поведінковий акт		
			Воля як бажання		
			Розум	Почуття	Ви н На я
Директив и	Вищі почуття				
Несвідоме	Вік	Темперамент	Вільні та складні уявлення	Душевні стани	Ге н детерм
			Уявлення як розумовий образ	Фізичні почуття як емоційний образ	
	Стать		Відчуття	Чуттєві відчуття	
	Нація		Інстинкти		Ев йно-в особл
			Складні рефлексии		
			Рефлексии		
Раса	Тропізми				

Рис. 2. Теоретична модель особистості у
концепції І.О. Сікорського

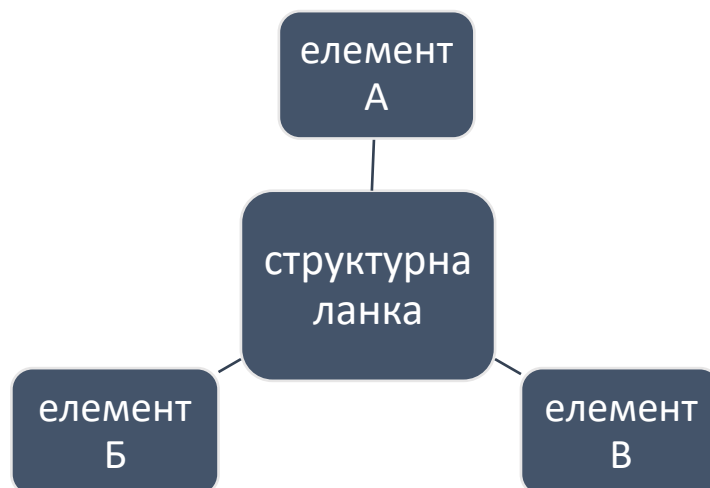
Теоретична модель може мати, до прикладу, такі типи структур або включати такі фрагменти структури:



Блочний цикл



Горизонтальна багаторівнева ієрархія



Радіальний кластер

Прикладом класичної моделі може бути така теоретична модель ситуації саморегуляції особистості (Мазяр, 2014а) (рис. 3).

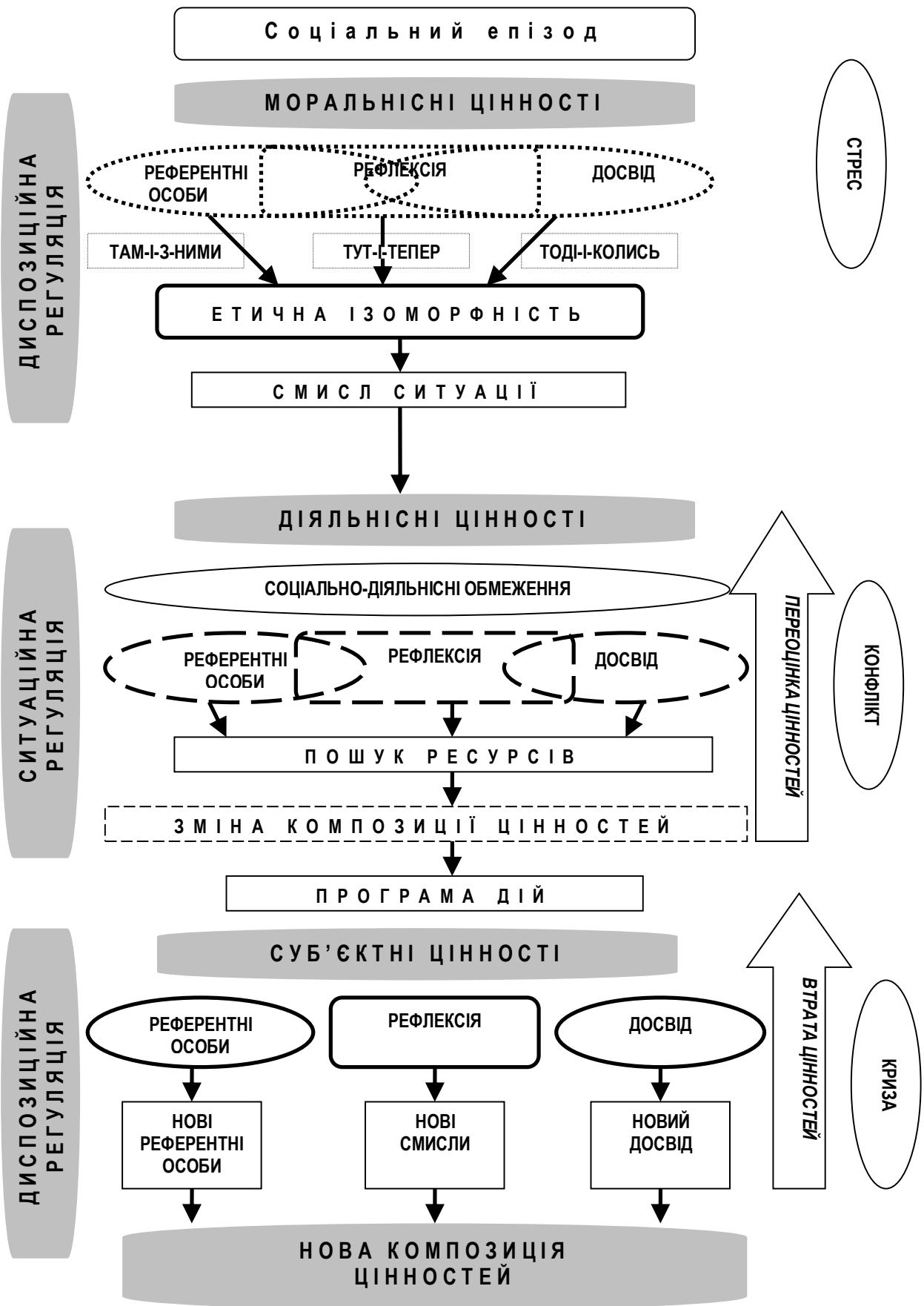


Рис. 3. Теоретична модель ситуаційної саморегуляції

§ 4. ПРИПУЩЕННЯ ЩОДО СТРУКТУРНИХ ЗВ'ЯЗКІВ

Якщо виокремлення структурних елементів феномену може бути покладене в основу шкал і факторів діагностичного інструменту, то структурні зв'язки – в основу інтерпретації результатів дослідження. Недостатньо лише поміряти вираженість окремих шкал. Їх необхідно показати у взаємних відношеннях, взаємовпливах. Власне, нерівність у вираженні шкал (факторів) вказує на індивідуальні відмінності суб'єкта, його самобутність, а відтак є кінцевою метою психодіагностики. Наприклад, повертаючись до концепції І.О. Сікорського (рис. 2), можна встановлювати, як у того чи іншого індивіда розвинуті вольові, розумові та почуттєві компоненти особистості, які вказують на її типологічні відмінності, співвідношення темпераменту та характеру тощо. А на прикладі теоретичної моделі ситуаційної саморегуляції (Мазяр, 2014а) можемо припускати психодинаміку формування суб'єктних цінностей.

Вивчення структурних зв'язків феномену базується на розумінні *принципів формального аналізу індивідуальності*. Зокрема, це дозволяє отримати відповідь на питання про:

- співвідношення елементів у певному ряду психологічних феноменів (**принцип ієрархії**). Наприклад, вираженість особистісних рис, що вивчаються у низці особистісних опитувальників;
- міру стабільності та змінюваності окремих психічних структур та їхніх елементів (**принцип інваріантності**);
- оптимальність та компенсацію структурних елементів (**принцип координації та взаємодії**);
- інтеграцію, адаптивність та результативність у функціонуванні окремих психічних характеристик (**принцип індивідуальної єдності**).

Кожен із цих принципів може бути закладений в якості основної мети психологічного вимірювання або виявлятися у зв'язку з рештою. Наприклад, для опитувальників провідними стають принципи ієрархії (субординації) і координації та взаємодії; для об'єктивних тестів – принцип інваріантності; для проєктивних методик – принцип індивідуальної єдності.

Питання та завдання для самоконтролю

Яку інформацію з'ясовуємо, відповідаючи на питання «Що ми будемо вивчати?» та «Як ми будемо вивчатимемо?»

Що називають вихідним концептом діагностичного інструмента?

Яку функцію виконує системотворча дефініція?

Що вирізняє теоретичну модель психологічного феномену з-поміж тривіального реєстру психічних якостей і властивостей?

Яким чином можуть бути використані структурні елементи (ланки) та їхні взаємозв'язки у теоретичній моделі психологічного феномену при розробці діагностичного інструменту?

Якими є провідні принципи формального аналізу психічних феноменів для укладання тестових завдань вимірювання інтелекту, особистісних якостей?

Тести для самоконтролю

1. Дефініція, яка стає початковою точкою розробки діагностичного інструменту – це
 - а) вихідний концепт; б) гіпотеза; в) наукова новизна; г) надійність методики.
2. Припущення науковця, яке вимагає доведення – це
 - а) вихідний концепт; б) гіпотеза; в) наукова новизна; г) апробація.
3. Характеристики психологічного феномену, представлені у форматі структурних елементів, які взаємопов'язані та взаємообумовлені, – це
 - а) система гіпотез; б) реєстр якостей та властивостей; в) теоретична модель; г) наукова новизна.
4. Структурні ланки теоретичної моделі психологічного феномену при розробці діагностичного інструмента можуть бути співвіднесені з
 - а) пунктами методики; б) надійністю методики; в) інтерпретацією результатів методики; г) шкалами методики.
5. Структурні взаємозв'язки ланок теоретичної моделі психологічного феномену при розробці діагностичного інструмента можуть бути співвіднесені з
 - а) пунктами методики; б) надійністю методики; в) інтерпретацією результатів методики; г) шкалами методики.
6. Принцип формального аналізу, який вказує на співвіднесення елементів у ряду психологічних феноменів, – це
 - а) принцип ієрархії (субординації); б) принцип інваріантності; в) принцип координації та взаємодії; г) принцип індивідуальної єдності.
7. Принцип формального аналізу, який вказує на міру змінюваності психічних структур та елементів, – це
 - а) принцип ієрархії (субординації); б) принцип інваріантності; в) принцип координації та взаємодії; г) принцип індивідуальної єдності.
8. Принцип формального аналізу, який вказує на оптимальність і компенсація структурних елементів, – це
 - а) принцип ієрархії (субординації); б) принцип інваріантності; в) принцип координації та взаємодії; г) принцип індивідуальної єдності.
9. Принцип формального аналізу, який вказує на інтеграцію, адаптивність і результативність функціонування психічних якостей, – це
 - а) принцип ієрархії (субординації); б) принцип інваріантності; в) принцип координації та взаємодії; г) принцип індивідуальної єдності.

Лекція № 5
ФОРМАТ ПУНКТИВ
ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТАРІЮ

План

1. Інструкція психодіагностичного інструменту.
2. Формат завдань об'єктивного тесту.
3. Формат пунктів опитувальника.
4. Формат завдань проєктивної методики.
5. Створення пробної версії діагностичного інструменту.

Мета: ознайомити з основними форматами пунктів і завдань психодіагностичного інструментарію.

Основні поняття: інструкція, пункт, завдання, пробна версія методики, робоча версія методики, специфікація тестових завдань, L-дані, T-дані, Q-дані, фактор соціальної бажаності відповідей, установка на крайні відповіді, установка на середні відповіді.

§ 1. ІНСТРУКЦІЯ ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТУ

Перед тим, як укладати тестові завдання, необхідно здійснити специфікацію психодіагностичного інструмента й укласти інструкцію до нього. **Специфікація тестових завдань** полягає у встановленні формальних характеристик діагностичного інструмента. Зокрема, йдеться про:

- тип тестових завдань;
- формат пунктів;
- опис стимульного матеріалу (за потребою);
- склад і кількість шкал (субтестів);
- кількість пунктів для тесту в цілому та кожної зі шкал (субтестів);
- співвідношення змістовних елементів;
- процедура проведення дослідження;
- способи обробки результатів;
- тип стандартизованих показників;
- форма звіту.

Кожна складова специфікації є важливою і взаємопов'язаною з рештою складових. Тип тестових завдань та формат пунктів зводяться до того, яким чином ми операціоналізуємо психологічну феноменологію. Існують три основних групи (типи) даних, на основі яких можна робити висновки.

L-дані (*Life record data*). Цей тип даних отримується шляхом спостереження за реальною поведінкою суб'єкта та її фіксацією. Сюди належать щоденникові записи, звіти досліджуваного, експертні оцінки спостерігачів. Фіксується частота вияву феномену. Основним інструментальним викривленням цих даних є *ефект ореолу*. Його уникнення стає можливим у випадку збільшення кількості

джерел отримання інформації (наприклад, достатня кількість експертів, літературних джерел). Основна проблема полягає в операціоналізації даних різної модальності, упередженості спостерігача (експерта, свідка). Такі дані технічно складно перевести з якісних характеристик у кількісні показники, які можна зіставляти з рештою кількісних показників.

Q-дані (*Questionnaire data*). Цей тип даних отримується в результаті опитування та самооцінок суб'єкта. З цією метою розробляються особистісні опитувальники, які можуть вимірювати рівень агресивності, тривожності, вмотивованості тощо. Основним інструментальним викривленням є фактор *соціальної бажаності*, який може виявлятися як свідомо, так і несвідомо. Джерелом фактору соціальної бажаності стає сам суб'єкт. З цією метою в опитувальники часто вводять «шкалу неправди», щоби виміряти рівень щирості відповідей. Однак кількісні показники дозволяють порівняно легко скласти висновки.

T-дані (*Objective test data*). Цей тип даних є результатом проведених об'єктивних тестових процедур. Наприклад, йдеться про тести на встановлення типу нервової системи, тести інтелекту. Їхня основна перевага полягає у тому, що результати не викривлюються ані фактором соціальної бажаності (навмисно не покращуються), ані упередженістю дослідника, експерта чи свідка.

Зрозуміло, що вибір одного з типів даних має свої переваги та недоліки, а також ступінь релевантності дослідницькій меті. Приміром, T-дані неможливо використовувати для вимірювання особистісної типології суб'єкта, а Q-дані будуть нерелевантні дослідженню когнітивних здібностей.

Інструкція до виконання тестових завдань повинна формулюватися простими реченнями, бути когнітивно якомога простішою, містити однозначні та зрозумілі вказівки. Інструкції не мають бути перевантаженими науковими термінами. Зі змісту інструкцію суб'єкт не повинен здогадуватися про дослідницьку мету психодіагностики, надто щодо опитувальників і проєктивних методик, щоби уникнути фактору соціальної бажаності.

Інструкція спрямована пояснити досліджуваному суть психологічного випробування та запобігти уточнювальним запитанням досліджуваних. Якщо такі питання виникають, відповіді на них повинні бути теж простими і не наштовхувати досліджуваних на певні відповіді. В окремих випадках слід лише повторити текст інструкції, наголосити на окремих положеннях.

У випадку проведення об'єктивного тесту доцільно продемонструвати приклад виконання завдання або дозволити досліджуваним пройти тренування. Зрозуміло, що попередньо необхідно підготувати низку прикладів, які віддалено або за аналогією пояснюють суть виконуваного завдання (Kline, 2000). Інструкція може бути скоригована, так само як і завдання діагностичного інструменту, під час проведення пілотажних досліджень. У цьому випадку дослідник орієнтується на частоту питань певного змісту.

§ 2. ФОРМАТ ЗАВДАНЬ ОБ'ЄКТИВНОГО ТЕСТУ

Вибір релевантного формату пунктів діагностичного інструментарію повинен бути підкріплений змістовною коректністю сформульованих завдань. Передовсім діагностичні завдання (пункти) повинні бути сформульовані якомога *простіше* і бути зрозумілими респондентам із різним життєвим досвідом, академічною успішністю, тобто мають бути толерантними до індивідуальних особливостей виконавців. Це не означає, разом із тим, що пункти (завдання) мають бути когнітивно простими для виконання, що, наприклад, не відповідатиме вимогам валідності тестів інтелекту.

Зрозумілість завдань не повинна вказувати респонденту на мету дослідження. Це напрочуд важливо у випадку з опитувальниками, позаяк підвищуватиме рівень соціальної бажаності відповідей, що знижуватиме валідність інструмента.

За формою пункти різних типів діагностичного інструментарію матимуть істотні відмінності. Для об'єктивних тестів (до яких ми відносимо тести інтелекту, психофізіологічні тести, тести професійного відбору) доцільно давати завдання у вигляді певних задач, які необхідно вирішувати (інколи з урахуванням часу та кількості спроб). Важливим аспектом вимірювань є можливість варіювати складність завдань, успішність виконання яких вказуватиме на рівень розвитку тих чи інших когнітивних здібностей, інтелектуальних властивостей. Зокрема, релевантними будуть такі типи задач:

1. Завдання на **вибір правильного варіанту** відповіді з-поміж запропонованих, інколи з кількома правильними відповідями на вищих рівнях складності.

2. Завдання на **аналогії**. Повинна бути одна правильна відповідь з-поміж запропонованих. Йдеться про пошук еквіваленту.

3. Завдання на **співвідношення**. Часто в якості стимульного матеріалу пропонуються геометричні фігури, інші невербальні елементи.

4. Завдання на **відповідність**. Пропонуються два набори даних, які треба правильно між собою зіставити.

5. Завдання на **виключення зайвого**. Цей тип завдань обернений до першого типу: у переліку всі відповіді правильні, крім однієї.

6. Завдання на **послідовність**. У такий спосіб можна вивчати розуміння зв'язків, відношень елементів. Можливий варіант використання невербальних стимулів.

Часто-густо окремі тести використовують лише один тип завдань. Наприклад, Тест інтелекту Дж. Равена (шкала прогресивних матриць) використовує тільки завдання на послідовність (гомогенність завдань), причому на невербальному абстрактному матеріалі (Raven, 1989; Raven, 2000). Натомість Тест інтелекту Д. Векслера вивчає як загальний рівень інтелекту, так окремо досліджує вербальний та невербальний інтелект. Відповідно, завдання є гетерогенними (Векслер, 2018).

Сильною стороною гомогенних завдань, вочевидь, можна вважати зручність обробки результатів та їхнього зіставлення з показниками різних досліджуваних. Тоді як гетерогенні завдання вимагають більш тривалої процедури обробки, труднощів у порівнянні результатів.

Сильна сторона гетерогенних завдань – вивчення більшої кількості властивостей певного феномену. Тоді як гомогенні завдання часто-густо вивчають якусь одну здібність, властивість, яка дискретно відбиває розвиток феномену.

§ 3. ФОРМАТ ПУНКТИВ ОПИТУВАЛЬНИКА

Основна проблема при розробці опитувальників полягає у низькій валідності випробування. Зокрема, низьку валідність можуть обумовлювати (Kline, 2000):

- 1) установка **погоджуватися** (надто коли питання або твердження є неоднозначними);
- 2) **фактор соціальної бажаності** відповідей (бажання бути «правильним»);
- 3) установка на **середні відповіді** типу «не знаю»;
- 4) установка на **крайні відповіді**;
- 5) **очевидна валідність** відповідей.

Окрема проблема розробки опитувальників – формування адекватної вибірки досліджуваних і вибірки пунктів діагностичного інструмента. Відповідно, опитувальник вимагає створення *банку тверджень-питань*, достатнього для варіювання та заміни тверджень у процесі його розробки.

Опитувальники, які спрямовані на вимірювання різних характеристик особистості (емоцій, поведінки, мотивації, цінностей), можуть розроблятися як у форматі питань, так і тверджень. Респонденту належить дати відповідь з-поміж запропонованих або погодитися чи не погодитися з твердженням. У цьому випадку можливі кілька форматів відповіді. Найпростіший – «так» або «ні». Іншими варіантом такого оцінювання може бути «правда» або «брехня», «подобається» або «не подобається».

Можливим є такий варіант відповідей: «так» – «?» – «ні». Знак запитання у цьому випадку може трактуватися як «не знаю», «важко відповісти», «не певен». Однак в інструкції необхідно зазначити, щоби досліджувані по можливості уникали середньої відповіді, оскільки вона є «привабливою» у ситуації вагання. Втім, оптимальним способом нівелювання цієї проблеми є формулювання так званих трихотомічних завдань із такими варіантами відповідей: «зазвичай» – «інколи» – «ніколи». Іншим варіантом є пропонування розгорнутих відповідей, серед яких досліджуваний повинен обрати потрібну (Kline, 2000).

Більш складними вважаються відповіді «так», «скоріше так», «не знаю», «скоріше ні», «ні». Тобто йдеться про диференціацію відповідей за двома, трьома, п'ятьма або навіть сімома балами. Вища за 7-бальну градація відповідей вважається неефективною, надмірною. Відповіді також можуть надаватися у

вигляді цифрових балів, причому інколи їх можна зображати у континуумі, на якому респондент ставить відповідні позначки (так звані рейтингові шкали).

Важливо пам'ятати, що для кожного виду завдань існують чіткі специфічні правила їх конструювання, з якими можна ознайомитися у довідниках (наприклад, Furr, 2011; Dowing, 2006; Haladyna, 2004; Haladyna & Dowing, 1989; Kline, 2000). Основними взаємопов'язаними принципами конструювання завдань в опитувальниках повинні стати зрозумілість, конкретність, щирість, коректність, вичерпність, однозначність формулювань. Слід уникати двох питань-тверджень, вказівки на частоту дій («як часто...», «я часто...»). Рекомендується формулювати питання-твердження таким чином, щоби вони вказували на поведінкові аспекти психічної активності, а не на думки та почуття: поведінка має чіткий об'єктивний вияв, а думки та почуття часто-густо не об'єктивуються.

З метою виявлення соціально схвалюваних відповідей розробники опитувальників використовують шкалу «неправди». Однак адекватні результати відповідей на ці питання-твердження будуть зміщені у бік низьких показників, а відтак відхилятимуться від графіка нормального розподілу. Підставою для того, щоби не вводити шкалу «неправди», є підтверджена висока валідність опитувальника.

§ 4. ФОРМАТ ЗАВДАНЬ ПРОЄКТИВНОЇ МЕТОДИКИ

Завдання для проєктивних методик зазвичай зводяться до однієї дії: вибрати певний стимульний матеріал, намалювати, домалювати, розповісти, зобразити невербальним чином, встановити послідовність. В організаційно-методичному аспекті проєктивні методики виглядають простішими, однак в інтерпретаційному плані є значно складнішими за об'єктивні тести та опитувальники. Кольоровий тест Люшера вимагає від діагноста значного досвіду застосування методики (Станіславський, 2000).

Проєктивні методики можуть мати ключ до інтерпретації результатів (наприклад, малюнкові методики), але зміст тлумачення не є очевидним та однозначним навіть для фахівців. Досить часто існують альтернативні варіанти інтерпретації. Позаяк ключовим моментом є здатність фахівця представляти не окремі характеристики досліджуваного, а, радше, синтетичні, які інтегрують комплекс психологічних характеристик. Це, своєю чергою, вимагає високої теоретичної підготовки розробника проєктивної методики, зокрема у галузі глибинної психології. Прикладом розробки такого методу є комплекс тематичних малюнків, запропонований академіком НАПН України Т.С. Яценко (Яценко, 1996; Яценко, 2006; Яценко (Ред.), 2008).

§ 5. СТВОРЕННЯ ПРОБНОЇ ВЕРСІЇ ДІАГНОСТИЧНОГО ІНСТРУМЕНТУ

Пробна версія діагностичного інструменту передбачає, що розробник перевірятиме валідність всього банку пунктів (завдань). З цією метою проводяться пілотажні дослідження з використанням порівняно невеликої, проте репрезентативної вибірки респондентів. На цьому етапі відбувається відбір кращих зразків пунктів (завдань), які матимуть пряму сильну кореляцію (внутрішню узгодженість). Власне, саме це відрізняє пробну версію інструменту від робочої. **Робоча версія** – це вже затверджений реєстр пунктів і завдань, які перевіряються на валідність і надійність діагностичного інструменту.

Зазвичай кількість завдань для пробної версії інструменту *втричі* перевищує ту, яка планується на етапі специфікації. Але подача цих пунктів (завдань) має бути збалансованою. Тобто рекомендується чергувати порядок пунктів (завдань) різного типу, щоби досліджувані не звикали до одного алгоритму відповідей, що може спотворити результати.

Мета пробної версії психодіагностичної методики – запропонувати певну кількість пунктів, які можуть бути включені у робочу версію діагностичного інструмента.

Питання та завдання для самоконтролю

Схарактеризуйте значення кожної зі складових процедури специфікації психодіагностичного інструмента.

Який тип даних буде більш доречним для складання психологічної біографії видатної постаті?

На підставі якого типу даних можна робити висновки про розвиток здібностей дитини?

Запропонуйте приклади діагностичних завдань з вибором правильної відповіді з-поміж запропонованих для тесту професійного відбору.

Запропонуйте приклади діагностичних завдань за аналогією для тесту інтелекту.

Запропонуйте приклади діагностичних завдань на відповідність та послідовність.

Наведіть приклад питання або твердження, яке може актуалізувати установку на середні та крайні відповіді.

Сформулюйте кілька тверджень, які спрямовані виявляти ступінь щирості досліджуваного (у шкалі «неправди»).

Яку мету переслідує пробна версія діагностичного інструментарію?

Скільки рекомендується розробити завдань для пробної версії інструмента?

Яка мета робочої версії методики?

Тести для самоконтролю

- 1.** Цей тип даних отримується шляхом спостереження або самоспостереження досліджуваного:
а) L-дані; б) Q-дані; в) T-дані; U-дані.
- 2.** Цей тип даних отримується шляхом стандартизованого опитування суб'єкта:
а) L-дані; б) Q-дані; в) T-дані; U-дані.
- 3.** Цей тип даних отримується шляхом об'єктивного тестування:
а) L-дані; б) Q-дані; в) T-дані; U-дані.
- 4.** L-дані виявляються в результаті вивчення
а) експертних оцінок; б) самооцінок в опитувальнику; в) показників об'єктивного тесту; г) проєктивних малюнків дослідженого.
- 5.** Q-дані виявляються в результаті вивчення
а) експертних оцінок; б) самооцінок в опитувальнику; в) показників об'єктивного тесту; г) проєктивних малюнків дослідженого.
- 6.** T-дані виявляються в результаті вивчення
а) експертних оцінок; б) самооцінок в опитувальнику; в) показників об'єктивного тесту; г) щоденників суб'єкта.
- 7.** Ефект ореола стає викривленням
а) L-даних; б) Q-даних; в) T-даних; U-даних.
- 8.** Фактор соціальної бажаності є викривленням
а) L-даних; б) Q-даних; в) T-даних; U-даних.
- 9.** У діагностиці найважче викривити (штучно покращити)
а) L-дані; б) Q-дані; в) T-дані; U-дані.
- 10.** Встановлення формальних характеристик діагностичного інструмента – це
а) валідність завдань; б) надійність завдань; в) специфікація завдань; г) дискримінативність завдань.
- 11.** Правильною є одна відповідь з-поміж запропонованих – це завдання на
а) вибір; б) послідовність; в) відповідність; г) творчість.
- 12.** Зіставлення двох груп даних – це завдання на
а) вибір; б) послідовність; в) відповідність; г) виключення зайвого.
- 13.** Встановлення зв'язків між елементами вивчається за допомогою завдань на
а) вибір; б) послідовність; в) відповідність; г) творчість.
- 14.** Оберненим до завдання на вибір правильної відповіді з-поміж запропонованих варіантів є завдання на
а) виключення зайвого; б) послідовність; в) відповідність; г) творчість.
- 15.** Часті відповіді «не знаю», «не певен» вказують на
а) установку давати крайні відповіді; б) установку давати середні відповіді; в) фактор соціальної бажаності; г) установку погоджуватися.
- 16.** Бажання давати «правильні» відповіді є виявом
а) установки давати крайні відповіді; б) установки давати середні відповіді; в) фактору соціальної бажаності; г) установки погоджуватися.
- 17.** Часті відповіді «так» є проявом
а) установки давати крайні відповіді; б) установки давати середні відповіді; в) фактору соціальної бажаності; г) установки погоджуватися.
- 18.** Шкала «неправди» має на меті виявити відповіді, які обумовлені

а) установкою давати крайні відповіді; б) установкою давати середні відповіді; в) фактором соціальної бажаності; г) установкою погоджуватися.

19. Вже затверджений реєстр пунктів (завдань) має

а) специфікація методики; б) робоча версія методики; в) план методики; г) пробна версія методики.

20. Перевірку на відповідність змісту пунктів пропонує

а) специфікація методики; б) робоча версія методики; в) план методики; г) пробна версія методики.

Лекція № 6
АНАЛІЗ ПРОБНОЇ ВЕРСІЇ
ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТУ

План

1. Емпірична апробація пробної версії.
2. Відбір пунктів і завдань.
3. Корекція пунктів і завдань.
4. Перевірка дискримінативності шкал.
5. Затвердження змісту та структури інструкції, пунктів та завдань.

Мета: ознайомити з процедурою емпіричної апробації пробної версії оригінального психодіагностичного інструментарію.

Основні поняття: апробація, дискримінативність шкал, класичний аналіз, факторний аналіз, критеріальний принцип відбору пунктів.

§ 1. ЕМПІРИЧНА АПРОБАЦІЯ ПРОБНОЇ ВЕРСІЇ

З емпіричної апробації пунктів розпочинається найбільш складна й клопітка частина розробки діагностичного інструментарію. Якщо до цього часу розробник перебував у межах власної концепції, то тепер зіставляє її з реальністю. Це, своєю чергою, вимагає не тільки скрупульозності підходу, але й чесності, готовності редагувати попередню роботу, подекуди суттєво її коригувати.

Мета апробації – змістовно наповнити методику таким матеріалом, який адекватно вимірюватиме психологічні феномени, заради чого діагностичний інструмент замислювався. Іншими словами, належить здійснити відсів пунктів, щоби ті, які залишаться, стали ефективними у вимірюванні. Такий відбір пунктів (завдань) здійснюється на основі статистичних показників. По суті це первинний відсів пунктів, які найбільш очевидно виявляють слабкість у виявленні психологічних характеристик. Усуваються з переліку ті пункти, які знижують рівень внутрішньої узгодженості психодіагностичної методики.

Найбільша увага приділяється розкиду даних. Чим менший такий розкид, тим гірше конкретний пункт відповідає діагностичним цілям. Адже мета дослідження полягає у встановленні відмінності. У випадку ж малого розкиду даних не задовольняється вимога діагностичного інструмента бути дискримінативним (таким, що успішно диференціює досліджуваних). Те саме стосується пунктів, на які досліджувані дають або лише крайні, або середні відповіді. Особливо важливо дотримуватися цих вказівок при розробці об'єктивних тестів, завдання яких, приміром, можуть бути або занадто легкими для конкретної вибірки, або занадто складними. Тож у великій нагоді стануть міри центральної тенденції: середнє арифметичне, мода, медіана (Климчук, 2009).

§ 2. ВІДБІР ПУНКТІВ І ЗАВДАНЬ

Методологи розробки психодіагностичного інструментарію пропонують чотири основні моделі відбору пунктів:

1. Відбір пунктів (завдань) на основі **класичного аналізу**. Застосовується у випадку розробки опитувальника з однією шкалою (наприклад, виявленням міри реактивної актуалізації конкретної емоції). Тоді зі шкали вилучаються ті пункти, які слабо або обернено корелюють із загальним показником шкали. Назагал таку ж процедуру можна застосувати у випадку наявності двох або навіть трьох шкал, якщо наявна достатня кількість пунктів. Тоді необхідно перевіряти кореляцію пунктів за показником шкали. Наприклад, у випадку вивчення реактивної (ситуаційної) та особистісної тривожності за методикою Ч.Д. Спілбергера та Ю.Л. Ханіна.

2. Відбір пунктів (завдань) за допомогою **факторного аналізу**. Застосовується у випадку існування кількох шкал, які, власне, будуть тотожними виокремленим факторам. Факторний аналіз дозволить видалити з реєстру ті пункти, які слабо корелюватимуть зі шкалами та іншими пунктами. При цьому слід подбати про достатню навантаженість всіх факторів. Разом із тим, існує небезпека формування дуже вузького змісту шкал, що суттєво збіднюватиме діапазон дослідження, знижуватиме дискримінативність психодіагностичного інструментарію. Втім, така ситуація є досить рідкісною, позаяк передбачає, що дослідник попередньо не здійснив належного теоретичного аналізу психологічного феномену і цілковито покладається на емпіричні показники.

3. Відбір пунктів (завдань) за **критеріальним принципом**. Застосовується у випадку здійснення відбору та прогнозування. Суть такого підходу полягає у встановленні кореляції конкретної шкали психодіагностичного інструмента з відповідним зовнішнім критерієм. Фактично вивчається, якою мірою декларування відповідає реальним, об'єктивним показникам. Основна мета такого підходу – *підвищити дискримінативність шкал*. Відповідно, слід відбирати пункти з високим показником дискримінативності. Проте ця модель відбору пунктів не вважається прийнятною для побудови класичних діагностичних інструментів (Kline, 2000), особливо для особистісних опитувальників. Позаяк складно й неекономно моделювати експериментальні ситуації, в яких досліджувані психологічні феномени можуть проявитися (об'єктивуватися) у поведінці.

4. Відбір пунктів (завдань) на основі «**Item response theory**». Застосовується при розробці тестів досягнень та здібностей. Для цього використовують спеціальні комп'ютерні програми (Embretson & Reise, 2000). Параметрами відбору пунктів стають одночасно показник дискримінативності, рівень труднощів та ймовірність вгадування відповідей. Це об'єктивується у побудові кривої на графіку. Важливою умовою застосування такої моделі відбору пунктів є високий початковий рівень їхньої узгодженості.

§ 3. КОРЕКЦІЯ ПУНКТИВ І ЗАВДАНЬ

Після відсіву менш прийнятних завдань (пунктів) методики, здійснюється подальший відбір за допомогою однієї з вище вказаних моделей. Після цього відбувається формування шкал або єдиної шкали, що перевіряються на відповідність внутрішній узгодженості та дискримінативності.

Тільки після такої процедури діагностичний інструмент остаточно затверджується. Втім, може відбуватися корекція пунктів та завдань. Основний напрямок корекції стосується виправлень занадто суб'єктивного тлумачення завдань досліджуваними. Окремі пункти можуть викривлено тлумачитися через неоднозначність формулювань. Пункти та завдання, з одного боку, повинні формулюватися досить широко, щоби бути зрозумілими для всіх респондентів, але з іншого – не повинні втрачати специфіки тестової ситуації, в якій може актуалізуватися досліджуваний феномен.

Пункти опитувальника повинні містити якомога менше слів, не бути перевантаженими складносурядними реченнями, термінологією, з якою респондент може бути не знайомий. У формулюваннях пунктів слід орієнтуватися на те, щоби респондент не перепитував діагноста щодо змісту того чи іншого пункту.

Кінцевий збір пунктів вимагає прискіпливого підходу до їхнього компонування. Зокрема, слід уникати накопичення у певній частині опитувальника тверджень, які є тільки прямими або тільки оберненими. Твердження повинні бути рівномірно й пропорційно розташовані в опитувальнику щодо приналежності до шкал, причому не бути в одному контексті. Наприклад, «Ви любите вирощувати квіти?» та «Вам подобаються намальовані квіти?». Порушення таких правил призводить до штучного підвищення або зниження показника кореляції.

§ 4. ПЕРЕВІРКА ДИСКРИМІНАТИВНОСТІ ШКАЛ

Подальші дослідження з уже скомпонованою формою діагностичного інструмента відбувається на новій вибірці досліджуваних, яка так само має відповідати вимогам репрезентативності. Основними показниками успішності розробки стають **внутрішня узгодженість** показників та **дискримінативність** шкал.

Коефіцієнт *внутрішньої узгодженості* діагностичного інструмента повинен бути від 0,9 до 0,7. Втім, оцінка узгодженості може достатньо сильно варіювати в залежності від типу діагностичного інструмента.

Дискримінативність тесту – це здатність інструменту виявляти відмінності між респондентами. Задовільна дискримінативність тесту об'єктивується, зокрема у нормальному розподілі даних. Дискримінативність вимірюється за допомогою дельти Фергюсона, де n – число завдань (пунктів), N – кількість досліджуваних, а f_i – частота кожного показника, який стрічається:

$$\delta = \frac{(n+1)(N_T^2 - \sum f_i^2)}{n \times N_T^2}$$

По суті показник дискримінативності вказує на рівень «чутливості» інструмента для вимірювання певного психологічного феномену, його здатність показувати весь діапазон оцінок – від максимальних до мінімальних. Показник дискримінативності може бути високим (вище 0,5), середній (від 0,1 до 0,5) та низький (менше як 0,1).

Загальний показник дискримінативності діагностичного інструмента залежить від дискримінативності всіх окремих завдань (пунктів) та шкал.

Після того, як показники внутрішньої узгодженості та дискримінативності стануть задовільними, можна продовжувати роботу над психодіагностичним інструментом. Однак розробник неодмінно стикається з тим, що показник дискримінативності методики буде суперечити показнику надійності. Позаяк високі показники дискримінативності автоматично зменшують показник кореляції (тісноти взаємозв'язку).

§ 5. ЗАТВЕРДЖЕННЯ ЗМІСТУ ТА СТРУКТУРИ ІНСТРУКЦІЇ, ПУНКТИВ ТА ЗАВДАНЬ

Затвердження змісту та структури психодіагностичного інструмента відбувається після того, як статистично підтверджуються показники валідності, надійності та дискримінативності. На цьому етапі ще відбувається уточнення окремих аспектів психодіагностики: час виконання завдань, послідовність дій, надання інструкції тощо. Методика повинна бути доведена до стану, коли її зможуть використовувати інші фахівці. Це означає, що має бути також розроблена інструкція до використання методики, інтерпретація та представлення її результатів, формулювання можливих рекомендацій. Тобто опис методики та її процесуальних аспектів має бути вичерпним, а сама методика максимально стандартизована, інакше це позначатиметься на її валідності. Наприклад, через зростання фактору соціальної бажаності або тенденції давати середні відповіді ми не матимемо цілісного та достовірного уявлення про поведінку суб'єкта, при експериментальному дослідженні фіксуватимемо розбіжності між деклараціями та реальною поведінкою.

Особливу увагу слід приділити внесенню змін до *інструкції*. Те, що розробнику часто-густо здається легким для розуміння, у досліджуваних може викликати запитання та нерозуміння. Для цього варто звертати увагу і під час пілотажного дослідження (які питання ставлять), і під час інтерв'ю, яке проводиться після дослідження (чи були труднощі у розумінні окремих завдань).

Варто також звернути увагу, яку установку дає інструкція: чи не є така установка шкідливою для валідності результатів?

Встановлюється формат реєстрації відповідей, алгоритм обробки результатів і приблизний час виконання завдань. В окремих випадках необхідно подбати про те, щоби час реєструвався з точністю до секунди. Наприклад, методика експрес-діагностики властивостей нервової системи за психомоторними показниками (теппінг-тест) Є.П. Ільїна. Час виконання завдання – шість випробувань по п'ять секунд – є основним параметром вимірювання (Мазяр, 2017).

Назагал часові обмеження у проходженні дослідження необхідні для того, щоби порівняти когнітивну ефективність (при виконанні тестових завдань) та уникнути надмірного обдумування відповідей з їх подальшим викривленням (при проходженні опитування). При цьому час подекуди стає одним із параметрів оцінки сформованості психологічного феномену.

Питання та завдання для самоконтролю

Яка основна мета апробації пунктів (завдань) психодіагностичного інструмента?

Яким чином можуть прислужитися міри центральної тенденції на етапі апробації методики?

Наведіть приклад методик, де релевантною процедурою відбору пунктів (завдань) буде класичний аналіз, факторний аналіз, критеріальний принцип.

Чому критеріальний принцип відбору пунктів психодіагностичного інструментарію майже не використовується для розробки особистісних опитувальників.

На що вказує показник дискримінативності шкал?

Тести для самоконтролю

1. Відсів неефективних пунктів психодіагностичного інструмента відбувається на етапі

а) планування методики; б) розробки методики; в) апробації методики; г) затвердження методики.

2. Відбір пунктів на основі класичного аналізу найбільш ефективний у випадку розробки

а) проєктивних малюнків; б) опитувальника з однією шкалою; в) тесту для вимірювання психофізіологічних властивостей; г) анкети з відкритими відповідями.

3. Відбір пунктів за допомогою факторного аналізу найбільш ефективний у випадку виокремлення

а) елементів проєктивного малювання; б) результатів психофізіологічного вимірювання; в) шкал особистісного опитувальника; г) частоти відкритих питань анкети.

4. Відбір пунктів та завдань за критеріальним принципом найбільш релевантне
а) тестам професійного відбору; б) проєктивним методикам; в) особистісним опитувальникам; г) мотиваційним опитувальникам.

5. Відбір завдань на основі «Item response theory» найбільш релевантне
а) тестам досягнень та здібностей; б) проєктивним методикам; в) особистісним опитувальникам; г) мотиваційним опитувальникам.

6. Високий показник дискримінативності шкали опитувальника:

а) 0; б) 0,1; в) 0,5; г) 0,9.

7. Середній показник дискримінативності шкали опитувальника:

а) 0; б) 0,1; в) 0,5; г) 0,9.

8. Низький показник дискримінативності шкали опитувальника:

а) 0; б) 0,1; в) 0,5; г) 0,9.

9. Здатність тесту вимірювати те, для чого він призначений –

а) валідність; б) надійність; в) дискримінативність; г) стандартизація.

10. Кількість пунктів у пробних версіях тесту має перевищувати заплановану у специфікації:

а) не менше, ніж у 4 рази; б) не менше, ніж у 3 рази; в) не менше, ніж у 2 рази; г) не менше, ніж у 6 разів.

Лекція № 7
ВАЛІДНІСТЬ ТА НАДІЙНІСТЬ
ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТАРІЮ

План

1. Поняття надійності психологічного вимірювання.
2. Надійність паралельних форм.
3. Основні чинники надійності.
4. Поняття валідності.
5. Методи оцінки валідності.
6. Чинники валідності.

Мета: ознайомити з умовами досягнення надійності та валідності психодіагностичного інструментарію.

Основні поняття: надійність методики, надійність паралельних форм, метод розщеплення, ретестова надійність, валідність, методи оцінки валідності, внутрішня узгодженість, коефіцієнт кореляції, дисперсія, внутрішня валідність, зовнішня валідність, конкурентна валідність, прогностична валідність, змістовна валідність, конструктна валідність.

§ 1. ПОНЯТТЯ НАДІЙНОСТІ ПСИХОЛОГІЧНОГО ВИМІРЮВАННЯ

Найбільш важливими властивостями діагностичного інструментарію є валідність, надійність та дискримінативність. Однак якщо дискримінативність методики була важлива на етапі укладання пунктів (завдань), то надійність та валідність як базові психометричні характеристики вимірюються на етапі затвердження робочої форми психодіагностичного інструменту.

Надійність психодіагностичного інструмента показує, наскільки точними є вимірювання, наскільки узгодженими та стабільними є результати дослідження у певних проміжках часу. Тобто надійний інструментарій забезпечує стійкі та відтворювані результати.

Надійність методики можна трактувати як

- 1) відношення дисперсії істинних балів до дисперсії балів, які спостерігаються;
- 2) квадратну кореляцію між балами, які спостерігаються, та істинними балами;
- 3) відсутність дисперсії помилки;
- 4) відсутність кореляції між балами, які спостерігаються, та помилкою.

Відповідно до кожного з таких підходів, розроблені формули обрахунку надійності (Furr, 2011).

Надійний психодіагностичний інструмент має два фундаментальних значення: 1) вказує на ступінь внутрішньої узгодженості; 2) вказує на ступінь повторюваності результатів (ретестова надійність) (Kline, 2000). Ці два

одночасних показники стають необхідною, але не достатньою умовою валідності психодіагностичного інструмента.

Вимога бути **внутрішньо узгодженим** означає, що окремі шкали та пункти психодіагностичного інструмента пов'язані із загальним результатом дослідження. Разом із тим, висока внутрішня узгодженість неодмінно звужуватиме спектр вивчення психологічного феномена до рівня встановлення окремих його характеристик (Cattell & Kline, 1977). Іншими словами, існує думка, що висока внутрішня узгодженість призводить до втрати валідності, адже у такому випадку феномен вивчається не цілісно, а дискретно.

Ретестова надійність показує, наскільки результати дослідження змінилися через певний проміжок часу. Задовільним вважається показник кореляції 0,7 та вище. Щоби не підвищити ретестову надійність штучним чином (і спотворити уявлення про діагностичну цінність інструмента), слід обрати репрезентативну вибірку й уникати близьких у часі повторних досліджень (Kline, 2000). Вочевидь, мова повинна йти про відстань у кілька місяців.

Розрахунок надійності тесту відбувається за формулою:

$$r_t = 1 - \frac{s_E^2}{s_X^2}$$

де s_E^2 – це дисперсія помилкового тестового балу, s_X^2 – дисперсія тестового балу, який спостерігається.

Своєю чергою, помилка вимірювання обраховується таким чином:

$$s_E = s_X \sqrt{1 - r_t}$$

§ 2. НАДІЙНІСТЬ ПАРАЛЕЛЬНИХ ФОРМ

Одна з форм вимірювання надійності психодіагностичного інструмента – перевірка **надійності паралельних форм**. З цією метою розробляється альтернативна форма методики, яка за змістом (для опитувальників) та складністю (для об'єктивних тестів) має бути еквівалентна вже розробленій. Як і у випадку з ретестовою надійністю, дослідження проводиться з тією самою вибіркою. Втім, додаткова трудність полягає саме в тому, щоби довести високий ступінь еквівалентності (взаємозамінності) завдань, який, зрозуміло, ніколи не може бути абсолютним. Класичним прикладом використання такого підходу є Шкільний тест розумового розвитку.

Перевага паралельних форм методики полягає у тому, що, по-перше, у такий спосіб можна скоротити час перевірки надійності, адже не треба чекати кілька місяців, щоби здійснити ретест. По-друге, додаткова форма психодіагностичного інструмента може використовуватися без небезпеки тренування досліджуваних та запам'ятовування відповідей. Це буває доречним, наприклад, у випадках надання психологічної допомоги, коли відбувається два незалежних і взаємозамінних вимірювання психологічних характеристик до початку

допомоги та після отримання послуг. Мало того, паралельна форма методики може застосовуватися під час надання допомоги для моніторингу психологічного стану клієнта.

Таким чином, використання паралельних форм дозволяє одночасно перевіряти як надійність методики, так і ступінь взаємозамінності форм. Основний показник у цьому випадку – коефіцієнт кореляції результатів двох форм психодіагностичного інструмента. У цих двох формах має спостерігатися однакова дисперсія помилки, а значить будуть однакові середні та стандартні відхилення (Морозов, 1994).

Недоліком такого підходу стає те, що досягти високого рівня взаємозамінності двох форм вдається зрідка, надто у випадку використання опитувальників. Респонденти все одно можуть засвоювати принципи виконання завдань, зокрема тестових.

Альтернативою перевірки надійності паралельних форм може стати *метод розщеплення*. У цьому випадку тест ділиться на дві рівні частини. Принцип поділу може бути такий: парні – непарні завдання; перша половина – друга половина. Головне, щоби вони були структурно ідентичними (рівномірний розподіл пунктів шкал). Затим вимірюється кореляція між половинами методики. У цьому випадку може використовуватися формула Спірмена-Брауна:

$$r_t' = \frac{2r_t}{1+r_t}$$

де r_t' – виправлений коефіцієнт надійності; r_t – коефіцієнт надійності за половинами розщепленого психодіагностичного інструмента.

Також надійність може вираховуватися за альфою Кронбаха. У такий спосіб відбувається вимірювання розкиду кожного пункту із загальним розкидом даних. Якщо загальний показник менший, ніж розкид показників кожного пункту, значить цей пункт адекватно спрямований на вимірювання феномену. Результат, відповідно, варіюється від 0 до 1.

§ 3. Чинники надійності

Слід наголосити, що всі форми визначення надійності діагностичного інструмента не можуть вважатися самодостатніми чи взаємозамінними. Кожна з них відповідає за окрему сторону надійності методики, тож подекуди їхні показники можуть не збігатися. Зокрема, узгодженість внутрішніх форм передуює визначенню ретестової надійності. Мало того, друга форма надійності не вивчається без встановлення першої. Так, можна сказати, що ретестова надійність вказує на часовий фактор збереження внутрішньої узгодженості інструмента (Морозов, 1994).

Необхідно враховувати змістовні аспекти дослідження. Адже існують такі психологічні характеристики, які є більш динамічними утвореннями, а відтак за кілька місяців можуть статися певні зміни, що закономірно позначиться на

зменшенні показника ретестової надійності. Іншими словами, слід чітко розрізняти феномени диспозиційного та ситуаційного характеру (Мазяр, 2008; Мазяр, 2014a). Наприклад, до *ситуаційних феноменів*, а відтак більш динамічних, можна віднести настрої, емоційний стан, ціннісно-мотиваційні аспекти розвитку особистості. *Диспозиційними характеристиками* можна вважати екстраверсію та інтроверсію, темпераментальні характеристики суб'єкта.

До слова, у зв'язку з цим постає питання про часову придатність результатів дослідження, необхідність проводити повторне дослідження з метою моніторингу окремих психологічних характеристик. Зокрема, ця проблема постає при масових обстеженнях у закладах освіти. І в цьому аспекті проблема надійності може вступати у певну суперечність з валідністю діагностичного інструмента. Психодіагностична методика може мати низький рівень надійності у тривалих проміжках часу, але при цьому залишатися валідною. Хоча, як зазначалося вище, високий показник надійності є запорукою високого показника валідності.

Ще одним непомітним спершу чинником надійності методики є зміна обставин ретестового дослідження: рівень освітленості, втомленість респондента, наявність додаткових шумів, зміна дослідника, актуальний емоційний стан досліджуваного. Тому рекомендується проводити повторні дослідження у максимально стереотипних умовах. У зв'язку з цим можуть заповнюватися певні анкети, які додатково вказуватимуть на відповідність цих умов.

§ 4. ПОНЯТТЯ ВАЛІДНОСТІ

Валідність психодіагностичної методики вказує на те, що діагностичний інструмент вимірює те, що дослідник очікує виміряти. Йдеться про емпіричну відповідність (релевантність) теорії. Це означає, що валідність слід розглядати насамперед відповідно до авторського розуміння психологічного феномену. Однак питання, чи відповідає таке розуміння загальноприйнятому уявленню, лишається відкритим. Низька валідність інструменту автоматично знижує діагностичну та прогностичну цінність методики.

Слід зазначити, що переклад валідного діагностичного інструмента на іншу мову вимагає включати процедуру її повторної валідації. Адже технічний переклад методики не завжди забезпечує культурну релевантність завдань та питань. Наприклад, методика Д. Векслера в Україні 2018 року пройшла повторну валідацію, в якій взяли участь 1159 дітей (582 дівчини та 577 хлопців).

Внутрішня валідність методики вказує на те, як завдання (пункти) пов'язані з теоретичними положеннями, з яких виходить дослідник. Можна також зіставити результати психодіагностичного інструмента з результатами вже розроблених методик.

Зовнішня валідність по суті є очевидною, причому не тільки для розробника, але й для досліджуваних. Хоча в окремих випадках досягти цього ефекту складно чи навіть неможливо, що надто виражено у проєктивних методиках. Назагал методологія проєктивного дослідження збудована на тому, щоби виявляти несвідомі аспекти психіки, які не є очевидними для досліджуваних та навіть для неспеціалізованих фахівців.

Існують різні класифікації встановлення валідності.

§ 5. МЕТОДИ ОЦІНКИ ВАЛІДНОСТІ

Встановлення валідності може відбуватися різними шляхами. У психометрії заведено розрізняти такі види валідності:

1. Конкурентна валідність. Йдеться про встановлення кореляції між результатами досліджень різних тестових процедур. Зазвичай зіставляють показники з уже розробленими валідними методиками. Втім, це зіставлення не повною мірою є коректним. Адже якщо існує інша задовільна методика, тоді навіщо взагалі розробляти новий психодіагностичний інструментарій? Зрозуміло, що потреба виникає саме тому, що існують деякі відмінності у теоретичних положеннях. Відповідно, не можна вважати, що два психодіагностичних інструментарії призначені вивчати одне й те саме. Наприклад, якщо вивчатимемо феномен невротизації особистості, то можемо стикнутися з різними шкалами методик (дещо відмінною симптоматикою), дарма що загальний рівень невротизації за методиками може бути ідентичний. Певною альтернативою іншому психодіагностичному інструментарію можуть стати рейтингові шкали або експертні оцінки. Але й у цьому випадку слід досить виважено підходити до зіставлення результатів (Kline, 2000).

2. Прогностична валідність. Йдеться про кореляцію між показниками розробленого психодіагностичного інструмента та певним критерієм у майбутньому (за кілька років). Фактично вивчається правдивість прогнозів поведінки суб'єкта. Назагал прогностична валідність доводить ступінь ефективності оцінки у психодинамічному аспекті, для чого, власне, така оцінка здійснюється. Проте варто розуміти, що помилки у прогнозуванні можуть і не вказувати на відсутність валідності методики, адже суб'єкт розвивається у мінливому соціумі, який може гальмувати формування певних психологічних характеристик або стимулювати їх. У цьому випадку ми повертаємося до проблеми ситуаційної саморегуляції особистості, вивчення ємної проблеми взаємодії особистості та ситуації (Мазяр, 2014а).

Мало того, вибір зовнішній критерій повинен бути адекватний тому, що досліджує методика. Приміром, високі показники інтелектуального розвитку не обов'язково корелюють з «життєвим успіхом», матеріальним добробутом, сімейним благополуччям тощо. З іншого боку, показники нейротизму можуть

корелювати з психічними захворюваннями особистості; показники особистісної тривожності – з гіперопікою власних дітей, тривожним розладом тощо.

3. Змістовна валідність. Йдеться про охоплення завданнями методики всіх областей досліджуваного явища. Зазвичай застосовується тільки в об'єктивних тестах, де очевидним, природним є зв'язок між завданням та феноменом, який вивчається. Про наявність змістовної валідності можуть судити навіть респонденти або неспеціалісти.

4. Конструктна валідність. Цей тип валідності інтегрує властивості решти типів, оскільки йдеться про здатність методики повно й цілісно описати той конструкт, який досліджується. Можна сказати, що властивість методики емпірично підтверджувати чи спростовувати теоретичні положення, які висуваються розробниками. Конструктна валідність спрямована на те, щоб орієнтуватися не на один, а одразу на кілька критеріїв валідності методики. Йдеться про накопичення валідності з різних джерел (Cronbach, 1989).

Втім, центральним все-таки лишається питання пошуку критерію валідності психодіагностичного інструмента. Саме це показує як пов'язані між собою показники методики з об'єктивно наявними показниками. У цьому випадку постає питання про адекватний вибір зовнішнього критерію, який має бути співмірним із показниками розробленого психодіагностичного інструмента.

Можна виокремити три типи основних критеріїв визначення валідності методики:

- 1) *об'єктивний критерій* (кількість виконаної за певний час роботи);
- 2) *проективно-рефлексивний критерій* (суб'єктивна обмеженість відповідей власним досвідом досліджуваного);
- 3) *критерій фізіологічного впливу* (певні показники фізичного стану досліджуваного).

§ 6. Чинники валідності

Додатковими чинниками валідності вже розробленого психодіагностичного інструментарію можуть стати:

1. *Кваліфікація та досвід спеціаліста*, який проводить дослідження. Надто це стосується проєктивних методик, які є достатньо відкритими для інтерпретації результатів, обговорення процесуальних питань. Найменш залежними від цього чинника валідності є об'єктивні тести.

2. *Ставлення досліджуваного до виконання завдань методики.* Найбільш вразливою при цьому є валідність опитувальників, позаяк це може об'єктивуватися у факторі соціальної бажаності відповідей. Навіть високий рівень мотивації не гарантує, а зумовлює дію цього фактору. Менш вразливими знову стають об'єктивні тести, які по суті є відкритими змаганнями. Високий рівень вмотивованості сприяє їхній валідності.

3. *Фізіологічний стан досліджуваного, рівень його стомлення, соматичного здоров'я.* Багато тестових завдань вимагають зосередженості, здатності рефлексувати, здійснювати продуктивну інтелектуальну діяльність.

Дослідник щоразу повинен враховувати ці чинники у комплексі, отримувати від досліджуваних адекватний зворотний зв'язок.

Питання та завдання для самоконтролю

На які властивості психодіагностичного інструмента вказує низький рівень надійності?

Яким чином вимірюється надійність методики?

Яке фундаментальне значення має надійність для розробки психодіагностичної методики?

З якою метою вимірюють ретестову валідність психодіагностичного інструмента?

Які переваги та недоліки вимірювання надійності паралельних форм методики?

Поясніть суть методу розщеплення тесту при вивченні надійності психодіагностичного інструмента.

Які чинники впливають на рівень надійності психодіагностичного інструмента?

Чим валідність тесту відрізняється від надійності?

У чому полягають відмінності між внутрішньою та зовнішньою валідністю психодіагностичної методики?

Яким чином вимірюється конкурентна валідність психодіагностичної методики?

У чому цінність прогностичної валідності психодіагностичної методики?

Що відрізняє змістовну та очевидну психодіагностичної валідність?

Охарактеризуйте основні чинники валідності психодіагностичної методики.

Тести для самоконтролю

1. Здатність тесту демонструвати точні та стабільні результати – а) валідність; б) надійність; в) дискримінативність; г) стандартизація.
2. Здатність тесту виявляти відмінності між випробовуваними – а) валідність; б) надійність; в) дискримінативність; г) стандартизація.
3. Ступінь внутрішньої узгодженості пунктів методики вказує на а) валідність; б) надійність; в) дискримінативність; г) стандартизацію.
4. Ступінь повторюваності результатів дослідження – це а) ретестова надійність; б) очевидна валідність; в) змістовна валідність; г) надійність паралельних форм.

- 5.** Ступінь відповідності результатів дослідження взаємозамінних варіантів методики – це
- а) ретестова надійність; б) очевидна валідність; в) змістовна валідність; г) надійність паралельних форм.
- 6.** Альтернативою вимірювання надійності паралельних форм психодіагностичної методики є
- а) метод фальсифікації тесту; б) метод дискримінації тесту; в) метод розщеплення тесту; г) метод подвоєння тесту.
- 7.** Ретестова надійність психодіагностичного інструмента вивчає його
- а) віковий фактор; б) ситуаційний фактор відповідей; в) часовий фактор внутрішньої узгодженості пунктів; г) статевий фактор.
- 8.** Зв'язок із теоретичними положеннями розробленої психодіагностичної методики вказує на
- а) внутрішню валідність; б) зовнішню валідність; в) надійність; г) дискримінативність.
- 9.** Очевидність застосування психодіагностичного інструмента вказує на
- а) внутрішню валідність; б) зовнішню валідність; в) надійність; г) дискримінативність.
- 10.** Показник кореляції різних тестових процедур вказує на
- а) конкурентну валідність; б) очевидну валідність; в) змістовну валідність; г) прогностичну валідність.
- 11.** Показник кореляції розробленої психодіагностичної методики та певним критерієм у майбутньому вказує на
- а) конкурентну валідність; б) очевидну валідність; в) змістовну валідність; г) прогностичну валідність.
- 12.** Найбільш інтегральним типом валідності є
- а) конкурентна валідність; б) очевидна валідність; в) змістовна валідність; г) конструктна валідність.
- 13.** Кваліфікація і досвід фахівця є найбільш вагомим фактором валідного використання психодіагностичного інструмента для
- а) тесту інтелекту; б) психофізіологічного тесту; в) особистісного опитувальника; г) проєктивної методики.
- 14.** Фактор соціальної бажаності справляє найбільший вплив на отримання валідних результатів
- а) тесту інтелекту; б) психофізіологічного тесту; в) особистісного опитувальника; г) проєктивної методики.
- 15.** Кількість продукції, виробленої за певний час, є
- а) об'єктивним критерієм валідності; б) проєктивно-рефлексивним критерієм валідності; в) критерієм фізіологічного впливу; г) критерієм випадковостей.
- 16.** Суб'єктивна обмеженість відповідей досліджуваного є
- а) об'єктивним критерієм валідності; б) проєктивно-рефлексивним критерієм валідності; в) критерієм фізіологічного впливу; г) критерієм випадковостей.
- 17.** Фізичний стан досліджуваного є

а) об'єктивним критерієм валідності; б) проєктивно-рефлексивним критерієм валідності; в) критерієм фізіологічного впливу; г) критерієм випадковостей.

18. Цей тип валідності визначається на основі співвіднесення результатів тесту із зовнішнім (щодо до тесту) критерієм:

а) критеріальна валідність; б) змістовна валідність; в) зовнішня валідність; г) конструктна валідність.

Лекція № 8
СТАНДАРТИЗАЦІЯ
ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТАРІЮ

План

1. Укладання остаточної версії психодіагностичного інструменту.
2. Стандартизація вибірки.
3. Нормалізація показників.
4. Комп'ютеризація тестування.

Мета: ознайомити з процедурою стандартизації психодіагностичного інструментарію.

Основні поняття: стандартизація вибірки, нормалізація показників, групові норми, предметно-орієнтовані норми, критеріальні норми, стени, станайни.

§ 1. УКЛАДАННЯ ОСТАТОЧНОЇ ВЕРСІЇ ПСИХОДІАГНОСТИЧНОГО ІНСТРУМЕНТУ

Остаточна, завершена версія психодіагностичного інструмента передбачає можливість стандартизації методики. Йдеться насамперед про отримання певних норм. Норми необхідні для того, щоби перевести отримані «сирі» бали у стандартні і порівнювати їх. Однак це робиться, якщо методику планують використовувати у прикладній діяльності; для суто наукових цілей стандартизація зазвичай не є обов'язковою (Furr, 2011; Kline, 2000). Назагал стандартизація дозволяє коректно порівнювати досліджуваних, тобто визначати їхнє унікальне місце на графіку нормального розподілу або відповідності певним соціокультурним нормам.

Стандартизація методики передбачає чітке визначення її мети. Зазвичай вона є суто практичною – визначення стану розвитку певного психологічного феномену, отримання рейтингового показника, прогнозування розвитку тощо. Встановлення норм допомагає адекватно оцінити отримані «сирі» показники та робити висновки, які статистично підтверджуються.

У психометричній літературі виокремлюють:

– *групові норми*, які показують ієрархічне (відносне) розташування дослідженого у певній стандартизованій вибірці за віковим, статевим, соціальним критерієм тощо;

– *предметно-орієнтовані норми*, які вказують на якісно-кількісні показники виконаних завдань у тестах досягнень (із залученням експертів);

– *критеріальні норми*, які вказують на очікуваний результат діяльності за певними критеріями (з розробкою таблиць очікувань).

Ці види норм можуть бути інтегровані в інтерпретацію результатів дослідження як самостійно, так і можуть бути представлені у комплексі.

Стандартизація діагностичного інструментарію проходить кілька послідовних етапів:

- I. Формування репрезентативної вибірки.
- II. Емпірична або експериментальна процедура встановлення стандартів.
- III. Вибір способу переходу до норм.

§ 2. СТАНДАРТИЗАЦІЯ ВИБІРКИ

Питання стандартизації вибірки у психометрії не вирішується однозначно. З метою зменшення стандартної погрішності рекомендується проводити дослідження не менше, як із 500 особами; щоби говорити про певну норму у загальній популяції (наприклад, у ситуації об'єктивного тестування), необхідна вибірка у 10 000 осіб (Kline, 2000). Слід зважати на склад конкретної вибірки, яка подекуди не може бути великою: наприклад, коли йдеться про нечисленні професійні групи, а самі дослідження є вузькоспеціалізованими. Тому заведено вважати, що репрезентативність вибірки важливіша за її величину, хоча й величина, безперечно, не може бути проігнорована. Європейська федерація психологічних асоціацій (EFPA) висуває вимогу щодо мінімальної вибірки стандартизації у 150 осіб; при цьому чіткої верхньої межі не існує.

Найскладніше розробляти психодіагностичний інструментарій для загальної популяції. У цьому випадку формування вибірки передбачає стратифікацію досліджуваних за певними категоріями (стать, вік, майновий статус, освіта тощо). Це, своєю чергою, вимагає знання пропорцій загальної популяції – відсоткової представленості тієї чи іншої категорії, перетини категорій. Останнє обумовлює суттєво зростання груп досліджуваних, що має відбиватися в анкетних даних, які передують дослідженню. Тому рекомендується зосередитися саме на тих ознаках (категоріях) стратифікації, які будуть найбільш значущими. Їх можна виокремити, спираючись на теоретичні положення та практичне спрямування психодіагностичного інструмента.

Об'єктивні тести бувають досить чутливими до різних вибірок. Наприклад, тест інтелекту. Тестові норми у цьому випадку залежать від вікової специфіки, причому подекуди на рівні років та місяців досліджуваних (тест інтелекту Д. Векслера). Тому норми представлені за окремими групами досліджуваних, яких для стандартизації тесту потрібно набрати не менше, як 150 осіб. Відтак збільшення категорій та, відповідно, груп досліджуваних автоматично збільшує загальну вибірку.

У випадку, якщо не вдається охопити стандартизацією всі необхідні групи, розробники методики чітко зазначають, яка вибірка стала основою стандартизації. Це необхідно документально зазначити, щоби фахівці, які користатимуться психодіагностичною методикою, розуміли, для якої вибірки отримані результати стануть валідними та надійними.

§ 3. НОРМАЛІЗАЦІЯ ПОКАЗНИКІВ

Нормалізація даних відбувається у випадку, якщо є відхилення від графіка нормального розподілу даних або самі пункти таким чином сформульовані, що зумовлюють відхилення (часто в особистісних опитувальниках). Однак нормалізація показників психодіагностичного інструмента може виявитись складною задачею. Це пов'язано з багатьма чинниками. Наприклад, якщо шкали містять різну кількість пунктів (завдань), їх належить коректно порівнювати між собою. Наприклад, Карта спостереження Д. Стотта спрямована на вивчення емоційних та поведінкових аспектів шкільної дезадаптації. В ній містяться 16 симптомокомплексів, але кожен симптомокомплекс має різну кількість симптомів (від 3 до 15). Тому єдиного кількісного ключа до інтерпретації не існує, а кожна шкала інтерпретується окремо.

З метою коректного зіставлення шкал пропонуються різні математичні способи перетворення так званих сирих балів на нормативні показники. Найпростіший спосіб – підрахунок **процентилів** за формулою: $n(x \leq X) / N * 100$, де $n(x \leq X)$ – число досліджуваних, які отримали бал не менше числа X , X – кількість балів, який отримав досліджуваний, а N – число всіх досліджуваних. Для зручності проведення розрахунків результати досліджуваної вибірки ранжують. Однак при цьому варто пам'ятати, що у такому випадку дослідник отримає відносні показники та не зможе робити висновки про істинні відмінності між досліджуваними.

Більш інформативним є **стандартний** спосіб, коли відбувається вимірювання ступеня відхилення від середнього показника. При цьому в тестових процедурах слід орієнтуватися на збереження графіку нормального розподілу даних.

Нормалізація показників може відбуватися шляхом переведення «сирих» балів у **станайни** та **стени** за спеціальною математичною процедурою. Станайни ще називають «дев'ятибальною шкалою» або «стандартною дев'яткою», а стени – «десятибальною шкалою» або «стандартною десяткою».

§ 4. КОМП'ЮТЕРИЗАЦІЯ ТЕСТУВАННЯ

Комп'ютеризація психодіагностичного інструментарію дозволяє автоматизувати значну частину роботи (замінює роботу калькулятора). Це суттєво економить час, дозволяє уникнути механічних і математичних помилок під час обрахунку результатів та видруковувати якісну інтерпретацію результатів дослідження. Проте для цього необхідно переконатися, чи електронна форма тестування не суперечить певним нормам тестування як такого.

Наприклад, використання електронної версії кольорового тесту Люшера мало своєю проблемою точну передачу кольорів карток, які на моніторі зазнають

деякого викривлення. Разом із тим, комп'ютерний варіант дає можливість зробити велику кількість виборів, що робить результати більш точними.

Досить обережним слід бути з розробкою електронних варіантів психофізіологічних тестів. Наприклад, до таких можна віднести тепінг-тест Є.П. Ільїна. У цьому випадку виконання тесту за допомогою стукання олівцем по паперу і натискання на клавіші комп'ютерної миші не є повністю еквівалентними (Мазяр, 2017).

Назагал може поставати проблема монотонії досліджуваних під час виконання великих за обсягом психодіагностичних методик (наприклад, ММРІ). Інколи може виникати психологічний бар'єр у взаємодії з електронною машиною (Виноградов, 1997).

Серед очевидних переваг можна відмітити, що електронний варіант тестування дає можливість чітко фіксувати час виконання завдання й автоматично обмежувати час роботи, продукувати нові стимули, забезпечувати їхню варіативність, організовувати нову стимульну послідовність.

Питання та завдання для самоконтролю

Яка практична мета стандартизації діагностичного інструмента?

У чому суть виокремлення групових норм у тестуванні?

Яким чином здійснюється формування предметно-орієнтованих норм психодіагностичної методики?

Опишіть послідовність етапів стандартизації психодіагностичного інструментарію.

Схарактеризуйте основні вимоги до формування стандартизації вибірки.

Які страти можуть бути основними під час формування вибірка загальної популяції?

Які переваги та недоліки мають електронні варіанти психодіагностичного інструмента?

Тести для самоконтролю

1. Мінімальний прийнятний розмір вибірки стандартизації становить:
 - а) 300 респондентів; б) 100 респондентів; в) 150 респондентів; г) 250 респондентів.
2. Стандартизація об'єктивного тесту зорієнтована на
 - а) графік нормального розподілу даних; б) практичну значущість методики; в) достатню кількість завдань; г) вибір валідних завдань методики.
3. Ієрархічне розташування досліджуваного у загальній вибірці досліджуваних показують
 - а) предметно-орієнтовані норми; б) групові норми; в) критеріальні норми; г) норми академічної успішності.
4. Експерти допомагають визначити

а) предметно-орієнтовані норми; б) групові норми; в) критеріальні норми; г) норми академічної успішності.

5. Переведені «сирі» бали тесту у стандартні за допомогою дев'ятибальної шкали називають

а) станайнами; б) стенами; в) процентилями; г) медіаною.

6. Переведені «сирі» бали тесту у стандартні за допомогою десятибальної шкали називають

а) станайнами; б) стенами; в) процентилями; г) медіаною.

7. Перевагою електронного варіанту психодіагностичного інструмента є

а) стан монотонії; б) брак емоційної взаємодії з діагностом; в) неможливість фіксувати додаткові реакції досліджуваного під час виконання завдань; г) економія часу при кількісній обробці результатів.

8. Недоліком електронного варіанту психодіагностичного інструмента є

а) стан монотонії; б) брак помилок при кількісній обробці результатів; в) можливість обмежувати час виконання завдань; г) економія часу при кількісній обробці результатів.

Лекція № 9
ІНТЕРПРЕТАЦІЯ ТА ЗВІТНІСТЬ
ПСИХОДІАГНОСТИЧНОГО ДОСЛІДЖЕННЯ

План

1. Загальна схема аналізу результатів.
2. Алгоритм кількісної обробки даних.
3. Критерії складання висновків.
4. Розробка електронної програми обробки даних

Мета: ознайомити з особливостями проведення інтерпретації та складання звітності за результатами психологічного дослідження.

Основні поняття: інтерпретація, висновки, звіт, моделі звіту.

§ 1. ЗАГАЛЬНА СХЕМА АНАЛІЗУ РЕЗУЛЬТАТІВ

Схема інтерпретації отриманих результатів розробляється на етапі створення психодіагностичного інструмента. Під час практичного користування фахівці довіряють їй настільки, що зазвичай не утруднюють себе додатковими міркуваннями щодо перевірки. Особливо це очевидно у випадках, де схеми інтерпретацій є максимально прописаними і не вимагають від фахівця майстерності власного тлумачення. Наприклад, це представлено у кольоровому тесті Люшера.

З одного боку, відмова від власних інтерпретацій відповідає методологічним вимогам (надто що стосується збереження валідності дослідження). З іншого боку, в окремих випадках спеціалісти повинні виявляти достатню гнучкість, позаяк всі варіанти аналізу, особливо якісного, не можуть бути представлені інтерпретаційними схемами. Особливо це помітно під час інтерпретації результатів проєктивних методик або багатошкальних опитувальників і тестів, де важливим стає розуміння індивідуальної логіки їхнього комбінування. У цьому випадку фахівці набувають певного досвіду валідної інтерпретації, позаяк щоразу зіставляють власні тлумачення з досвідом психологічною практикою.

Загальна схема аналізу включає

- 1) розробку схем інтерпретацій;
- 2) підготовку текстів інтерпретацій;
- 3) опис моделей звітів.

При цьому основними принципами, що забезпечують якість інтерпретації є орієнтація на конкретні практичні цілі (відбір, консультування, прогноз), дотримання меж змісту та опора на емпіричні дані, які були отримані в результаті психометричної перевірки психодіагностичного інструмента. Втім, фахівець має виходити з єдиної, методологічно вивіреної стратегії проведення інтерпретації результатів, спиратися на ключові показники (визначення рівня, профілю, зіставлення) та комбінації шкал (якщо такі представлені), дотримуватись певної послідовності розгортання тлумачення отриманих показників.

§ 2. АЛГОРИТМ КІЛЬКІСНОЇ ОБРОБКИ ДАНИХ

Кількісна обробка даних для більшості методик становить перший етап психодіагностики. Фахівець вираховує «сирі» бали (відповідно до розробленого ключа), які за необхідності переводить у стандартні показники (стени або станайни). Це дозволяє надалі здійснювати коректне порівняння як самих шкал (для індивідуального дослідження), так й окремих респондентів (у груповому дослідженні). На підставі кількісних показників здійснюється якісне тлумачення результатів, які мають бути попередньо виписані.

Окремо слід виокремити дослідження, які передбачають факторизацію даних. У цьому випадку неможливо передбачати кількість та зміст виокремлених факторів, а відтак інтерпретація залежить від здатності фахівця тлумачити логіку утворених факторів, вміння їх зіставляти, пояснювати взаємозв'язок окремих елементів, які найпотужніше корелюють між собою (Климчук, 2009). Разом із тим, саме факторна модель дозволяє аналізувати унікальність структури свідомості.

Кількісні дані дають можливість унаочнити отримані результати: побудувати відповідний профіль особистості, діаграму, графік. З одного боку, це дозволяє цілісно охопити досліджуваний феномен; з іншого – аналізувати окремі елементи, зіставляти їх один з одним. При складанні висновку рекомендується користатися стандартними даними для пояснення інтра- й інтеріндивідуальних відмінностей.

§ 3. КРИТЕРІЇ СКЛАДАННЯ ВИСНОВКІВ

Висновки становлять собою текст якісного аналізу результатів дослідження, який складається на підставі отриманих кількісних емпіричних показників. У випадку проєктивного дослідження кількісні показники можуть бути відсутніми. Висновки складають фахівці, які мають базову психологічну освіту, отримали спеціальну професійну підготовку, пройшли навчання та набули досвід використання психодіагностичного інструмента.

Формат висновку має бути однаковим для всіх досліджуваних, щоб уникнути апеляцій щодо ступеня його розгортання та повноти для окремих досліджуваних.

Основними принципами складання висновку повинна стати коректність інформації, лаконічність, однозначність, корисність, доступність сприймання. Слід уникати змістовних повторів, необхідно бути послідовним, наголошувати на ключових моментах. Може існувати *кілька варіантів висновків*: розгорнутий висновок, ознайомлювальний висновок (для певної категорії осіб) та короткий висновок (резюме). Кожен із них має свою мету і рівень складності подання

інформації в залежності від того, для кого висновок призначається. З цією метою висновок може ілюструватися графіками, діаграмами, рисунками тощо.

Зазвичай висновок містить систему рекомендацій або узагальнену, лаконічно сформульовану рекомендацію. Основна їхня задача – допомогти фахівцю та/або досліджуваному прийняти зважене рішення.

§ 4. РОЗРОБКА ЕЛЕКТРОННОЇ ПРОГРАМИ ОБРОБКИ ДАНИХ

З метою уникнення механічних та математичних помилок під час обрахунків рекомендується максимально автоматизувати складання висновку. Для цього користуються спеціально розробленими програмами або використовують можливості програм Excel, Statistica та інші. Рекомендується всі підрахунки здійснювати у цих програмах (коефіцієнт кореляції, факторний аналіз тощо). Вони також допомагають генерувати візуалізацію отриманих результатів (графіки, діаграми).

На сьогодні існують методики психодіагностики, які повністю автоматизовані. Причому не тільки у плані проведення та кількісної обробки результату, але й якісної (з відповідним складанням якісного тлумачення та формулюванням рекомендацій). Однак досягнення такого результату вимагає залучення до розробки психодіагностичного інструментарію фахівців у галузі ІТ-технологій, а також постійну технічну підтримку такої програми.

Питання та завдання для самоконтролю

Опишіть загальну схему аналізу результатів дослідження.

Які основні принципи якісної інтерпретації?

Який загальний алгоритм кількісної обробки даних?

Які існують варіанти висновків за результатами дослідження?

Які переваги електронної програми обробки даних?

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- Бочелюк, В. Й., & Бочелюк, В. В. (2008). Методика та організація досліджень із психології : навч. пос. для студ. Київ : Центр учбової літератури.
- Векслер, Д. (2018). WISC-IV. Шкала інтелекту для дітей. URL: <https://giuntipsy.com.ua/clinical/wisc-iv/>
- Виноградов, О. Г. (1997). Дослідження надійності комп'ютерного варіанту особистісного опитувальника 16PF. *Вісник Київського університету. Психологія*, 3.
- Горбунова, В. В. (2005). Експериментальна психологія в схемах і таблицях : навчальний посібник. Житомир : [б. в.].
- Гриньова Н. В. (Уклад). (2012). Історія психології : курс лекцій : навчальний посібник. Умань : Візаві.
- Кайко, В. І., & Мацієвська, С. М. (2018). Щодо експертної характеристики індивідуальних психологічних особливостей особистості за її почерком. *Вісник університету Альфреда Нобеля. Серія «Педагогіка і психологія». Педагогічні науки*, 2 (16), 58–65.
- Киричук, О. В., & Роменець, В. А. (Ред.) (2006). Основи психології : підручник. Київ : Либідь. [вид. 6-те, стереотип.].
- Климчук, В. О. (2009). Математичні методи в психології : навчальний посібник. Для студентів психологічних спеціальностей. Київ : Освіта України.
- Коханова О. П. (Авт-уклад.) (2016). Історія психології: від античності до початку ХХ століття : навч. посібник. Київ : НВП «Інтерсервіс».
- Мазяр, О. В. (2008). Особливості суб'єктної поведінки в екстремальних ситуаціях. *Актуальні проблеми психології : збірник наукових праць Інституту психології ім. Г. С. Костюка АПН України / за ред. С. Д. Максименка*. Київ : «Логос», 7, 15, 195–198.
- Мазяр, О. В. (2013). Концептуалізація проблеми обдарованості у працях І. О. Сікорського : автореф. дис. ... канд. псих. наук : 19.00.01 / Інститут психології імені Г. С. Костюка НАПН України. Київ.
- Мазяр, О. В. (2014а). Особливості ситуаційної саморегуляції творчо обдарованої особистості. *Наука і освіта*, 5, 237–242.
- Мазяр, О. В. (2014б). Сікорський в історії психології та історія психології про Сікорського. *Практична психологія та соціальна робота*, 2, 75–78.
- Мазяр, О. В. (2015). Психологічний експеримент І. О. Сікорського. *Прикладні аспекти психології особистісного зростання : збірник наукових праць студентів та викладачів*. Випуск 1 (С. 33–35) / за ред. Л. П. Журавльової, Н. Ф. Портницької, Ю. Ю. Дем'янчук. Житомир : Вид-во ЖДУ імені Івана Франка.
- Мазяр, О. В. (2017). Диференціальна психологія: модульний курс : навчальний посібник. Київ : Видавничий дім «Кондор».
- Морозов, С. М. (1994). Засоби контролю діагностичних якостей психологічних тестів : навч. посібник. Київ : ІСДО.

Музика, О. Л. (2007). Курсові роботи з психології : навчальний посібник для студентів вищих навчальних закладів / 2-е вид., перероб. і доп. Житомир : Вид-во ЖДУ ім. І. Франка.

Станіславський, С. В. (2000). Діагностика захисних механізмів особистості на основі тесту вибору кольору М. Люшера. *Практична психологія та соціальна робота*, 8.

Тітов, І. (2018). Елементи практичної психометрики. Київ : Кондор.

Яценко Т. С. (Ред.) (2008). Концептуальні засади і методика глибинної психокорекції : Підготовка психолога-практика : навчальний посібник. Київ : Вища школа.

Яценко, Т. С. (1996). Психологічні основи групової психокорекції : навчальний посібник. Київ : Либідь.

Яценко, Т. С. (2006). Основи глибинної психокорекції: феноменологія, теорія і практика : навчальний посібник. Київ : Вища школа.

Bain, S. K., & Allin, J. D. (2005). Book review: Stanford-Binet intelligence scales, fifth edition. *Journal of Psychoeducational Assessment*, 23, 87–95.

Cattell, R. B., & Kline, P. (1977). *The Scientific Analysis of Personality and Motivation*. New York : Academic Press.

Cronbach, L. J. (1960). *Essentials of psychological testing*. New York : Harper & Row.

Cronbach, L. J. (1989). Construct validation after thirty years. *Intelligence: measurement, theory, and public policy* / ed. by R. E. Linn (p. 147–171). Urbana : University of Illinois Press.

Dowing, S. M. (2006). Twelve steps for effective test development. *Handbook of test development* / ed. by S. M. Dowing, T. M. Haladyna, 3–25.

Embretson, S. E. & Reise, S. R. (2000). *Item response theory for psychologists*. Mahwah, NJ: ed. Lawrence Erlbaum Associates.

Fechner, G. Th. (1882). *Revision der hauptpuncte der psychophysik*. Leipzig, Breitkopf und Härtel.

Furr, M. (2011). *Scale Construction and Psychometrics for Social and Personality Psychology*. SAGE Publications Ltd.

Galton, F. (1889). Correlations and their measurement chiefly from anthropometric data. *Nature*, 39, 239.

Galton, F. (1902). The most suitable proportion between the values of first and second prizes. *Biometrika*, 1, 385–399.

Galton, F. (1907). Grades and deviates. *Biometrika*, 5, 400–406.

Haladyna, T. M. & Downing, S. M. (1989). A taxonomy of multiple-choice items-writing rules. *Applied Measurement in education*, 1, 37–50.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Hillsdale, NJ : Lawrence Erlbaum associates.

Kline, P. (2000). *The Handbook of Psychological Testing* / second ed. London and New York : Routledge.

Raven, J. (1989). The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation Within the United States. *Journal of Educational Measurement*, 26, 1, 1–16.

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41, 1, 1–48.