

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ВНУТРІШНІХ СПРАВ**

*Факультет № 6  
Кафедра соціології та психології*

**ТЕКСТ ЛЕКЦІЇ**

з навчальної дисципліни «Психодіагностика»  
обов'язкових компонент  
освітньої програми першого (бакалаврського) рівня вищої освіти

*053 Психологія (практична психологія)*

**Тема № 7. Надійність та валідність**

**Харків 2023**

**ЗАТВЕРДЖЕНО**

Науково-методичною радою  
Харківського національного  
університету внутрішніх справ  
Протокол від 30.08.2023 № 7

**СХВАЛЕНО**

Вченою радою факультету № 6  
Протокол від 25.08.2023 № 7

**ПОГОДЖЕНО**

Секцією Науково-методичної  
ради ХНУВС гуманітарних та  
соціально- економічних дисциплін  
Протокол від 29.08.2023 № 7

Розглянуто на засіданні кафедри соціології та психології (протокол №8 від 15.08.2023)

**Розробник:**

Доцент кафедри соціології та психології, кандидат психологічних наук, доцент  
Твердохвалова Ю.Л.

**Рецензенти:**

1. Професор кафедри психології Харківського національного педагогічного університету імені Г.С. Сковороди, доктор психологічних наук, професор, Кузнєцов М.А.
2. Доцент кафедри соціології та психології факультету № 6 Харківського національного університету внутрішніх справ, кандидат психологічних наук, доцент Греса Н.В.

## План лекції

1. Надійність.
2. Валідність.
3. Адаптація методик.

### Рекомендована література (основна, допоміжна), інформаційні ресурси в Інтернеті

#### Основна:

1. Бурлачук Л. Ф. Основи психодіагностики для психологів / Л. Ф. Бурлачук // Психологія і особистість. - 2012. - № 1. - С. 199-200.
2. Галян І.М. Психодіагностика: навч. посібник - К.: Академвидав, 2009. –464 с.
3. Клайн. П. Справочное руководство по конструированию тестов. – Киев, 2000.
4. Корольчук М.С., Осьодло В.І. Психодіагностика. Навч. посіб для вузів . –К.: Ельга Ніка-Центр. –2004. –400 с.

#### Допоміжна:

5. Практикум з психодіагностики / Укл.: О.О. Музика, В.М. Остринька, В.В. Остринський. – 2-ге вид., випр. і допов. – Житомир, 2000.
6. Психодіагностика: навчальний посібник / уклад. Ю.В. Кушнір; відп.ред. В.П. Кушнір. - Донецьк: КиЦ, 2012. - 346 с.
7. Самошкіна Л. М. Психодіагностика: Навч. наоч. посіб. з мультимедійним курсом /Під ред.чл.-кор. АПН України, професора Е. Л. Носенко. –Д.: Вид-во ДНУ, 2018. –448 с.
8. Терлецька Л. Г. Основи психодіагностики. Навчальний посібник. –К.:Главник, 2016. –144 с.
9. Шевченко Л.О., Чепіга Л.П. Практична психологія: навч. посібник Харків : Константа, 2018. - 191 с.

### Текст лекції

Перш ніж психодіагностичні методики можуть бути використані для практичних цілей, вони повинні пройти перевірку по ряду формальних критеріїв, що доводять їх високу якість та ефективність. Ці вимоги у психодіагностиці склалися роками у процесі роботи над тестами та їх вдосконаленням. В результаті з'явилася можливість захистити психологію від всіляких безграмотних підробок, які претендують на те, щоб називатися діагностичними методиками.

До основних критеріїв оцінки психодіагностичних методик відносяться надійність і валідність. Великий внесок у розробку цих понять зробили зарубіжні психологи (А. Анастасі, Е. Гізеллі, Дж. Гілфорд, Л. Кронбах, Р. Торндайк та Е. Хаген та ін.). Ними були розроблені як формально-логічний, так і математико-статистичний апарат (перш за все кореляційний метод і факторний аналіз) обґрунтування ступеня відповідності методик зазначеним критеріям.

У психодіагностиці проблеми надійності та валідності методик тісно взаємопов'язані, проте існує традиція роздільного викладу цих найважливіших

характеристик. Дотримуючись її, почнемо з розгляду надійності методик.

### **Питання 1. Надійність**

У традиційній тестології термін «надійність» означає відносне сталість, стійкість, узгодженість результатів тесту при первинному і повторному його застосуванні на одних і тих же випробуваних. Як пише А. Анастасі, навряд чи можна з довірою ставитися до тесту інтелекту, якщо за ним на початку тижня дитина мала показник, що дорівнює 110, а до кінця — 80. По-друге застосування надійних методик дає подібні оцінки. При цьому певною мірою можуть збігатися як самі результати, так і порядкове місце (ранг), яке займається випробуваним у групі. І в тому і в іншому випадку при повторенні досвіду можливі деякі розбіжності, але важливо, щоб вони були незначними, в межах однієї групи. Таким чином, можна сказати, що надійність методики - це такий критерій, який говорить про точність психологічних вимірів, тобто дозволяє судити про те, наскільки вселяють довіру отримані результати.

Ступінь надійності методик залежить багатьох причин. Тому важливою проблемою практичної діагностики є з'ясування факторів, що знижують точність вимірів та їх класифікацію. Аналіз показує, що серед них найчастіше називають такі:

- 1) нестабільність діагностованої властивості;
- 2) недосконалість діагностичних методик (недбало складена інструкція, завдання за своїм характером різнорідні, нечітко сформульовані вказівки, як методику пред'являти випробуваним, і т. д.);
- 3) змінюється ситуація обстеження (різна пора дня, коли проводяться експерименти, різна освітленість приміщення, наявність чи відсутність сторонніх шумів тощо. буд.);
- 4) відмінності в манері поведінки експериментатора (від досвіду до досвіду по-різному пред'являє інструкції, по-різному стимулює виконання завдань тощо);
- 5) коливання у функціональному стані випробуваного (в одному експерименті відзначається гарне самопочуття, в іншому стомлення тощо. буд.);
- 6) елементи суб'єктивності у способах оцінки та інтерпретації результатів (коли ведеться протоколювання відповідей піддослідних, оцінюються відповіді за рівнем повноти, оригінальності тощо).

Якщо всі ці фактори мати на увазі і постаратися в кожному з них усунути умови, що знижують точність вимірювань, то можна домогтися прийнятного рівня надійності тесту. Одним з найважливіших засобів підвищення надійності психодіагностичної методики є однаковість процедури обстеження, його строга регламентація: однакові для обстежуваної вибірки випробуваних обстановка і умови роботи, однотипний характер інструкцій, однакові для всіх тимчасові обмеження, способи та особливості контакту з випробуваним. ними, порядок пред'явлення завдань і т. д. При такій стандартизації процедури дослідження можна істотно зменшити вплив сторонніх випадкових факторів на результати тесту і таким чином підвищити їх надійність.

На характеристику надійності методик великий вплив робить досліджувана вибірка. Вона може як знижувати, і завищувати цей показник.

Наприклад, надійність може бути штучно завищена, якщо у вибірці невеликий розкид результатів, тобто якщо результати за своїми значеннями близькі один до одного. В цьому випадку при повторному обстеженні нові результати також розташуються тісною групою. Можливі зміни рангових місць випробуваних будуть незначними, отже, надійність методики буде високою. Таке ж невиправдане завищення надійності може виникнути при аналізі результатів вибірки, що складається з групи, що має дуже високі результати, і групи з дуже низькими оцінками по тесту. Тоді ці далеко віддалені один від одного результати не перекриватимуться, навіть якщо і втрутитися в умови експерименту випадкові фактори. Тому в посібнику зазвичай робиться опис вибірки, на якій визначалася надійність методики.

В даний час надійність все частіше визначається на найбільш однорідних вибірках, тобто на вибірках, схожих за статтю, віком, рівнем освіти, професійної підготовки і т. п. Для кожної такої вибірки наводяться свої коефіцієнти цієї характеристики. Наведений показник надійності застосовний тільки до груп, подібних до тих, на яких він визначався. Якщо методика застосовується до вибірки, що відрізняється від тієї, де провірялася її надійність, то ця процедура має бути проведена заново.

Як підкреслюють багато авторів, різновидів надійності методик так само багато, як і умов, що впливають на результати діагностичних випробувань. Проте практичне застосування знаходять лише кілька її видів. Оскільки всі види надійності відбивають ступінь узгодженості двох незалежно отриманих рядів показників, то основний математико-статистичний прийом, з допомогою якого встановлюється надійність методик, — це кореляція (Пірсоном чи Спирмену). Надійність тим вища, що ближче отриманий коефіцієнт кореляції підходить до одиниці, і навпаки.

У цій книзі при описі видів надійності основний упор робиться на роботи К. М. Гуревича, який, провівши ретельний аналіз зарубіжної літератури з цієї проблеми, запропонував тлумачити надійність як:

- ☐ надійність самого вимірювального інструменту;
- ☐ стабільність досліджуваної ознаки;
- ☐ константність, тобто відносну незалежність результатів від особистості експериментатора.

Основні показники він запропонував позначати так:

- 1) показник, що характеризує вимірювальний інструмент, пропонується називати коефіцієнтом надійності;
- 2) показник, що характеризує стабільність вимірюваної властивості, - Коефіцієнтом стабільності;
- 3) показник оцінки впливу особистості експериментатора – коефіцієнтом константності.

Саме в такому порядку рекомендується здійснювати перевірку методики: доцільно спочатку перевірити інструмент вимірювання. Якщо отримані дані задовільні, можна переходити до встановлення міри стабільності вимірюваної властивості, а вже після цього за необхідності зайнятися критерієм константності.

Зупинимось на докладнішому розгляді цих показників, що характеризують з різних сторін надійність психодіагностичної методики.

#### Визначення надійності вимірювального інструменту

Від того, як складена методика, наскільки правильно підібрані завдання з точки зору їх взаємоузгодженості, наскільки вона однорідна, залежить точність, об'єктивність будь-якого психологічного виміру. Внутрішня однорідність методики показує, що її завдання актуалізують одну і ту ж властивість, ознаку.

Для перевірки надійності вимірювального інструменту, що говорить про його однорідність (або гомогенність), використовується так званий метод «розщеплення». Зазвичай завдання діляться на парні та непарні, окремо обробляються, а потім результати двох отриманих рядів корелюються між собою. Для застосування цього способу потрібно поставити досліджуваних у такі умови, щоб вони змогли встигнути вирішити (або спробувати вирішити) усі завдання. Якщо методика однорідна, то великої різниці в успішності рішення за такими половинками не буде, і, отже, коефіцієнт кореляції буде досить високим.

Можна ділити завдання та іншим шляхом. Наприклад, можна зіставити першу половину тесту з другої, першу і третю чверть з другої і четвертої тощо. Однак «розщеплення» на парні та непарні завдання видається найбільш доцільним, оскільки саме цей спосіб найбільш незалежний від впливу таких факторів, як вработуваність, тренування, втома та ін.

Методика визнається надійною, коли отриманий коефіцієнт не нижчий від 0,75–0,85. Найкращі за надійністю випробування дають коефіцієнти порядку 0,9 і більше. Слід звернути увагу на той факт, що на початковому етапі розробки діагностичної методики можна отримати невисокі коефіцієнти надійності, наприклад порядку 0,46–0,50. Це означає, що в методикі, що розробляється, присутня деяка кількість завдань, які в силу своєї специфічності ведуть до зниження коефіцієнта кореляції. Такі завдання необхідно спеціально проаналізувати і або переробити їх, або взагалі вилучити.

Щоб легше було встановити за рахунок яких завдань знижуються коефіцієнти кореляції, необхідно проаналізувати таблиці з виписаними даними, підготовленими для кореляцій. Слід зазначити, що будь-які зміни у змісті методики - вилучення завдань, їх перестановка, переформулювання питань або відповідей - вимагають заново вираховувати коефіцієнти надійності.

При ознайомленні з коефіцієнтами надійності слід забувати, що вони залежить лише від правильного підбору завдань з погляду їх взаємоузгодженості, а й від соціально-психологічної однорідності тієї вибірки, де перевірялася надійність вимірювального інструмента.

Насправді, у завданнях можуть зустрітися поняття, мало відомі одній частині піддослідних, але добре відомі іншій частині. Від того, як багато в методиці таких понять, залежатиме і коефіцієнт надійності; Завдання з такими поняттями можуть випадково розташуватися і в парній, і в непарній половині тесту. Очевидно, показник надійності не слід приписувати лише методиці як такої і не можна сподіватися на те, що він буде незмінним, з якою б вибіркою не проводилося тестування.

#### Визначення стабільності досліджуваної ознаки

Визначити надійність самої методики - це не означає вирішити всі питання, пов'язані з її застосуванням. Потрібно ще встановити, наскільки стійка, стабільна ознака, яку дослідник має намір вимірювати. Було б методологічною помилкою розраховувати на абсолютну стабільність психологічних ознак. У тому, що ознака, що вимірюється, з часом змінюється, немає нічого небезпечного для надійності. Вся справа в тому, в яких межах варіюються результати від досвіду до досвіду в одного і того ж випробуваного, чи не призводять ці коливання до того, що випробуваний з незрозумілих причин виявляється то на початку, то в середині, то в кінці вибірки. Зробити якісь конкретні висновки про рівень представленості ознаки, що вимірюється, у такого випробуваного не можна. Таким чином, коливання ознаки не повинні мати непередбачуваний характер. Якщо не зрозумілі причини, через які відбувається різке коливання, то така ознака не може бути використана в діагностичних цілях.

Для перевірки стабільності ознаки, що діагностується, властивості використовується прийом, відомий під назвою тест-ретест. Він полягає у повторному обстеженні піддослідних за допомогою тієї ж методики. Про стабільність ознаки судять за коефіцієнтом кореляції між результатами першого та другого обстежень. Він свідчить про збереження чи незбереження кожним випробуваним свого порядкового номера у вибірці.

На ступінь стійкості, стабільності діагностується властивості впливають різноманітні фактори. Число їх досить велике. Вище вже говорилося про те, наскільки важливо дотримуватися вимог одноманітності процедури проведення експерименту. Так, наприклад, якщо перше тестування проводилося в ранковий час, то і повторне повинно бути проведене вранці; якщо перший досвід супроводжувався попереднім показом завдань, то і при повторному випробуванні ця умова також повинна бути дотримана і т.д.

При визначенні стабільності ознаки велике значення має проміжок часу між першим та другим обстеженнями. Чим коротший термін від першого до другого випробування, тим (за інших рівних умов) більше шансів, що ознака, що діагностується, збереже рівень першого випробування. З збільшенням часового інтервалу стабільність ознаки має тенденцію знижуватися, тому що зростає кількість сторонніх факторів, що впливають на неї. Отже, напрошується висновок про те, що доцільно проводити повторне тестування через короткий термін після першого. Однак тут є свої складнощі: якщо термін між першим і другим дослідженнями невеликий, то деякі випробувані можуть відтворити свої колишні відповіді по пам'яті і таким чином відійдуть від сенсу виконання завдань. І тут результати двох пред'явлених методики не можна розглядати як незалежні.

Важко чітко відповісти на питання, який термін можна вважати оптимальним для повторного експерименту. Тільки дослідник, виходячи з психологічної сутності методики, умов, в яких вона проводиться, особливостей вибірки піддослідних, повинен визначити цей термін. У цьому такий вибір має бути науково обґрунтований. У тестологічній літературі найчастіше називаються часові інтервали кілька місяців (але не більше півроку). При обстеженні дітей молодшого віку, коли вікові зміни та розвиток відбуваються дуже швидко, ці

інтервали можуть бути близько декількох тижнів.

Важливо пам'ятати, що коефіцієнт стабільності не слід розглядати тільки з його вузько формального боку, за його абсолютними значеннями. Якщо методика досліджує властивість, яка в період діагностування знаходиться в процесі інтенсивного розвитку (наприклад, вміння проводити узагальнення), то коефіцієнт стабільності може виявитися невисоким, але це не слід тлумачити як недолік методики. Такий коефіцієнт стабільності повинен інтерпретуватися як показник певних змін, розвитку досліджуваного властивості. У цьому випадку, наприклад, К. М. Гуревич рекомендує розглянути частинами ту вибірку, на якій встановлювався коефіцієнт стабільності. При такому розгляді виділиться частина піддослідних, які проходять шлях розвитку в однаково рівному темпі; інша частина - де розвиток йшов особливо швидкими темпами; і частина вибірки, де розвиток у піддослідних майже зовсім не можна побачити. Кожна частина вибірки заслуговує на спеціальний аналіз і тлумачення. Отже недостатньо просто констатувати, що коефіцієнт стабільності низький, потрібно зрозуміти, від чого це залежить.

Зовсім інша вимога пред'являється до коефіцієнта стабільності, якщо автор методики вважає, що властивість, що вимірюється, вже сформовано і має бути досить стійким. Коефіцієнт стабільності у разі має бути досить високим (не нижче 0,80).

Отже, питання стабільності вимірюваного властивості вирішується який завжди однозначно. Рішення залежить від сутності самого діагностованого властивості.

Визначення константності, тобто відносної незалежності результатів від особистості експериментатора

Оскільки методика, розроблена для діагностичних цілей, не призначена для того, щоб завжди залишатися в руках своїх творців, дуже важливо знати, якою мірою її результати піддаються впливу особистості експериментатора. Хоча діагностична методика завжди забезпечується докладними інструкціями щодо її застосування, правилами і прикладами, що вказують, як проводити експеримент, проте регламентувати манеру поведінки експериментатора, швидкість його мови, тон голосу, паузи, вираз обличчя дуже важко. Випробуваний у своєму ставленні до досвіду завжди відобразить те, як сам експериментатор до цього досвіду відноситься (допускає недбалість або діє точно відповідно до вимог процедури, виявляє вимогливість, наполегливість або безконтрольність тощо).

Хоча в тестологічній практиці критерієм константності користуються нечасто, проте, на думку К. М. Гуревича, це може бути основою його недооцінки. Якщо в авторів методики виникають підозри щодо можливого впливу особистості експериментатора на результат діагностичної процедури, то доцільно перевірити методику з цього критерію. При цьому важливо мати на увазі наступний момент. Якщо під впливом нового експериментатора всі випробувані однаковою мірою стали працювати трохи краще або трохи гірше, то сам по собі цей факт (хоча і заслуговує на увагу) на надійність методики не вплине. Надійність зміниться лише тоді, коли вплив експериментатора на



випробуваних по-різному: одні стали працювати краще, інші гірше, а треті так само, як і при першому експериментаторі. Іншими словами, якщо піддослідні за нового експериментатора змінили свої порядкові місця у вибірці.

Коефіцієнт константності визначається шляхом кореляції результатів двох дослідів, проведених у відносно однакових умовах на одній і тій же вибірці випробуваних, але різними експериментаторами. Коефіцієнт кореляції не повинен бути нижчим за 0,80.

Вище були розглянуті три показники надійності психодіагностичних методик. Може виникнути питання, чи потрібно при створенні методик здійснювати перевірку кожного з них? У професійній літературі ще точиться дискусія з цього приводу. Одні дослідники вважають, що всі способи визначення надійності певною мірою ідентичні, і тому достатньо перевірити надійність методики якимсь одним з них.

Наприклад, автор неодноразово перевидавалася в США книги зі статистики для психологів і педагогів Г. Геррет не знаходить принципових відмінностей між способами перевірки надійності. На його думку, всі ці способи показують відтворюваність тестових показників. Іноді той, іноді іншою забезпечують кращий критерій. Інші дослідники дотримуються іншої точки зору. Так, автори Стандартних вимог до педагогічних і психологічних тестів у розділі «Надійність» відзначають, що коефіцієнт надійності — це родове поняття, що включає кілька видів, і кожен вид має свій особливий зміст. Поділяє цю думку і До. М. Гуревич. На його думку, коли говорять про різні способи визначення надійності, то мають справу не з кращою чи гіршою мірою, а з різними мірами по суті надійності. Справді, чого варта методика, якщо не зрозуміло, чи надійна вона сама по собі як вимірювальний інструмент чи не встановлена стабільність вимірюваної властивості? Чого варта діагностична методика, якщо не відомо, чи можуть змінюватись результати залежно від того, хто веде експеримент? Кожен окремо показник ніяк не замінить інших способів перевірки і, отже, не може розглядатися як необхідна і достатня характеристика надійності. Тільки методика, що має повну характеристику надійності, найбільш придатна для діагностико-практичного застосування.

## Питання 2. Валідність

Іншим після надійності ключовим критерієм оцінки якості методик є валідність. Питання про валідність методики вирішується лише після того, як встановлено достатню її надійність, оскільки ненадійна методика не може бути валідною. Але найнадійніша методика без знання її валідності є практично марною.

Слід зазначити, що питання про валідність досі є одним із найскладніших. Найбільш укоріненим визначенням цього поняття є те, яке наведено в книзі А. Анастасі: «Валідність тесту — поняття, що вказує нам, що тест вимірює і наскільки добре він це робить».

Валідність за своєю суттю - це комплексна характеристика, що включає, з одного боку, відомості про те, чи придатна методика для вимірювання того, для чого вона була створена, а з іншого боку, яка її дієвість, ефективність, практична корисність. З цієї причини не існує єдиного універсального підходу до

визначення валідності. Залежно від цього, яку бік валідності хоче розглянути дослідник, застосовуються й різні методи докази. Іншими словами, поняття валідності включає різні її види, що мають свій особливий зміст. Перевірка валідності методики називається валідизацією.

Валідність у її розумінні має відношення до самої методики, тобто. це валідність вимірювального інструменту. Така перевірка називається теоретичною валідизацією. Валідність у другому розумінні вже відноситься не так до методики, як до мети її використання. Це прагматична валідизація. Узагальнюючи, можна сказати:

- при теоретичній валідизації дослідника цікавить сама властивість, що вимірюється методикою. Це, сутнісно, означає, що проводиться власне психологічна валідизація;
- при прагматичній валідизації суть предмета виміру (психологічної властивості) виявляється поза увагою. Головний акцент зроблено на те, щоб довести, що «щось», що вимірюється методикою, має зв'язок з певними областями практики.

Якщо знову звернутися до історії розвитку тестології, то можна виділити такий період (20—30-ті рр.), коли науковий зміст тестів та їх теоретичний «багаж» цікавили меншою мірою. Важливо було, щоб тест працював, допомагав швидко відбирати найбільш підготовлених людей. Емпіричний критерій оцінки тестових завдань вважався єдино вірним орієнтиром у вирішенні наукових та прикладних завдань.

Тому на ранніх стадіях розвитку тестології, коли концепція валідності тільки складалася, існувало інтуїтивне уявлення про те, що саме даний тест вимірює:

- 1) методика називалася валідною, оскільки те, що вона вимірює, просто «очевидно»;
- 2) доказ валідності ґрунтувався на впевненості дослідника в тому, що його метод дозволяє «зрозуміти випробуваного»;
- 3) методика розглядалася як валідна (тобто приймалося твердження, що такий тест вимірює таку якість) тільки тому, що теорія, на основі якої будувалася методика, «дуже хороша».

Прийняття на віру голосливних тверджень про валідність методики не могло тривати тривалий час. Перші прояви дійсно наукової критики розвінчали такий підхід: почалися пошуки науково обґрунтованих доказів.

Використання діагностичних методик з суто емпіричним обґрунтуванням, без чіткої теоретичної бази нерідко призводило до псевдонаукових висновків, до невинуватених практичних рекомендацій. Не можна було назвати ті особливості, якості, які тести виявляли. Б. М. Теплов, аналізуючи тести того періоду, назвав їх «сліпими пробами».

Такий підхід до проблеми валідності тестів був характерний аж до початку 50-х років. у США, а й у інших країнах. Теоретична слабкість емпіричних методів валідизації не могла не викликати критики з боку вчених, які в розробці тестів закликали спиратися не тільки на «голу» емпірику та практику, а й на теоретичну концепцію. Практика без теорії, як відомо, сліпа, а теорія без

практики мертва. В даний час теоретико-прагматична оцінка валідності методик сприймається як найбільш продуктивна.

У поняття валідності входить велика кількість найрізноманітнішої інформації про тест. У цілому нині вона визначає сферу застосування методики і відбиває рівень обґрунтованості результатів виміру. Різні категорії цих відомостей та способи їх отримання утворюють різні типи валідності. Основними типами вважаються змістовна валідність, конструктна валідність і критеріальна валідність. Класифікація типів валідності в достатній мірі умовна, тому що нерідко для різних критеріїв валідності застосовуються загальні методи визначення, а з іншого боку, одні й ті самі вихідні дані можуть інтерпретуватися з погляду різних типів валідності. На рис. 2 наведена зразкова схема, що відображає види та зв'язки валідності.

Провести теоретичну валідизацію, на відміну прагматичної, часом виявляється значно важче. Не вдаючись поки в конкретні деталі, зупинимось в загальних рисах на тому, як перевіряється прагматична валідність: вибирається якийсь незалежний від методики зовнішній критерій, що визначає успіх у тій чи іншій діяльності (навчальній, професійній тощо), і з ним порівнюються результати діагностичної методики. Якщо зв'язок між ними визнається задовільним, то робиться висновок про практичну значущість, ефективність, дієвість діагностичної методики. Для визначення теоретичної валідності знайти який-небудь незалежний критерій, що лежить поза методикою, набагато складніше. Теоретична валідність складається зі змістовної та конструктної.



Мал. 2. Основні види валідності

Змістовна валідність - один з основних типів валідності методики, що характеризує ступінь репрезентативності змісту завдань тесту вимірюваної області психічних властивостей.

Валідність за змістом закладається в тест вже при доборі завдань майбутньої методики. У змістовній валідності реалізується синтетичний підхід в аналізі елементного складу змісту самого тесту, а не сукупності зовнішніх критеріїв валідизації. Першим етапом валідизації є визначення кола

досліджуваних властивостей та видів діяльності, розчленування складної здатності чи діяльності на елементи. На другому етапі розробляють власне модель тестової діяльності на основі найбільш важливих елементів реальної діяльності. Нарешті, на останньому етапі проводять аналіз ступеня відповідності розробленої моделі реальної діяльності, перевірку відповідності пропорцій представленості елементів у завданнях тесту та в реальній діяльності. Так, для тестів досягнень з окремих предметів розробки конкретного змісту тестових завдань передують повна систематична перевірка відповідних підручників і навчальних програм, а також консультації з фахівцями з даного предмета. На основі зібраної таким шляхом інформації складають специфікацію тесту, де вказують тестовані області змісту (теми), завдання (процеси) навчання, а також відносне значення кожної теми і процесу для досягнення цілей навчання на дані етапі. Конкретні завдання оцінюються експертами за принципом їхньої близькості до реальних вимог (логічна валідність). Експерти виносять судження про те, чи охоплює тест репрезентативну вибірку конкретних навичок та знань досліджуваної галузі навчання. Широке використання експертних оцінок зближує змістовну валідність з процедурою визначення валідності критеріальної. Однак істотною відмінністю між цими типами валідності є те, що експертні оцінки при аналізі змісту є критерієм самого тесту, у той час як при критеріальній валідації вони відносяться до випробуваних з вибірки стандартизації.

Поряд з тестами досягнень змістовна валідність є однією з найважливіших форм валідації тестів критеріально-орієнтованих, а також методик, призначених для профвідбору, аналізу успішності оволодіння професією. Для валідації опитувальників особистісних і тестів інтелекту критерії змістовної валідності мають обмежене застосування і використовуються лише на початкових стадіях складання тесту.

Очевидна валідність - уявлення про тест, сферу його застосування, результативності і прогностичної цінності, яке виникає у випробуваного або іншої особи, яка не має спеціальних відомостей про характер використання і цілі методики. Очевидна валідність не є компонентом валідності, що об'єктивно встановлюється. Водночас висока очевидна валідність у більшості випадків є дуже бажаною. Вона виступає як фактор, що спонукає випробуваних до обстеження, сприяє більш серйозному і відповідальному відношенню до роботи з виконання завдань тесту і до висновків, що формулюються психологом. Достатній рівень очевидної валідності є особливо значущим для методик обстеження дорослих.

Конструктна валідність - один з основних типів валідності, що відображає ступінь репрезентації досліджуваного психологічного конструкту в результатах тесту. Як конструкт можуть виступати практичний або вербальний інтелект, емоційна нестійкість, інтроверсія, розуміння мови, переключення уваги і т. д. Інакше кажучи, конструктна валідність визначає область теоретичної структури психологічних явищ, що вимірюються тестом.

Оскільки прояви таких конструкцій, як, наприклад, інтелект у діяльності людини різноманітні і неоднозначні з погляду їх виділення, процедура

встановлення конструктивної валідності порівняно з валідністю критеріальною або валідністю змістовною більш складна.

Серед конкретних методів характеристики конструктивної валідності насамперед необхідно назвати зіставлення досліджуваного щодо конструктивної валідності тесту коїться з іншими методиками, конструктивне зміст яких відоме. Наявність кореляції між новим і аналогічним за конструктивним тестом вказує на те, що тест, що розробляється, «вимірює» приблизно ту ж сферу поведінки, здатність, особистісну якість, що і еталонна методика.

При аналізі конструктивної валідності методики зазвичай формують ряд гіпотез про те, як корелюватиме розроблюваний тест з широким колом інших тестів, спрямованих на конструктив, що знаходяться в теоретично відомому або передбачуваному зв'язку з досліджуваними. При цьому конструктивна валідність характеризується не тільки зв'язками тесту, що перевіряється з близькими показниками, але і з тими, де, виходячи з гіпотези, значущих зв'язків спостерігатися не повинно. Ці підходи визначаються як конвергентна (перевірка ступеня близькості прямого або зворотного зв'язку) і дискримінантна (встановлення відсутності зв'язку) валідації. Підтвердження сукупності теоретично очікуваних зв'язків становить важливе коло свідомостей конструктивної валідності. В англійській психодіагностиці таке операційне визначення конструктивної валідності позначається як «передбачувана валідність» (assumed validity).

Інкрементна валідність (англ. incremental - Зростання, прибуток) - один з компонентів критеріальної валідності, прогностичної валідності тесту, що відображає практичну цінність методики при проведенні відбору. Інкрементна валідність може бути кількісно виражена за допомогою коефіцієнта валідності.

Показник інкрементної валідності вказує на роль тесту у поліпшенні відбору осіб для реальної діяльності, ступінь поліпшення результативності процедури відбору в порівнянні з традиційною, заснованою на аналізі об'єктивних відомостей, документів, бесід, прийомі з випробувальним терміном. кому тощо. Пряме відношення до характеристики конструктивної валідності має факторний аналіз, що дозволяє суворо статистично проаналізувати структуру зв'язків показників досліджуваного тесту з іншими відомими і латентними факторами, виявити загальні та специфічні для групи зіставляваних тестів фактори, ступінь ленності в результатах, тобто визначити факторний склад і факторні навантаження результату тесту. Виняткова важливість такої процедури є підставою виділення її в особливий вид конструктивної валідності — факторну валідність.

Важливим аспектом конструктивної валідності є внутрішня узгодженість, що відображає те, наскільки певні пункти (завдання, питання), що становлять матеріал тесту, підпорядковані основному напрямку тесту як цілого, орієнтовані вивчення одних і тих же конструктивів. Аналіз внутрішньої узгодженості здійснюється шляхом корелювання відповідей на кожне завдання із загальним результатом тесту. Слід зазначити, що критерій внутрішньої узгодженості вказує лише на міру зв'язку всього змісту тесту з конструктивом, що вимірюється, даючи лише непряму інформацію про природу вимірюваної властивості.

При визначенні конструктної валідності важливе місце належить вивченню динаміки конструкту. При цьому ми можемо спиратися на гіпотези про його віковий розвиток, вплив тренувань, навчання, освоєння професії і т. д. Одним з таких підходів є застосування критерію вікової диференціації (валідність за віковою диференціацією). Характеристика валідності за конструктом тут полягає у визначенні відповідності результатів тесту теоретично очікуваним і практично спостережуваним віковим змінам даного конструкту або властивості. Найбільше значення валідності по віковій диференціації має для характеристики валідності тестів, спрямованих на вимірювання психологічних властивостей, функцій, що відрізняються відносно швидким зміною під впливом індивідуального досвіду, вираженою ієрархічністю ступенів розвитку (обізнаність, навички, інтелектуальні операції та ін.). Критерій валідності за віковою диференціацією зазвичай не знаходить широкого застосування при валідизації методик, призначених для психологічної діагностики функцій, властивостей, що не виявляють чіткої та однозначної тенденції до вікових змін. До них, зокрема, відносяться методи діагностики особистості.

У комплекс відомостей про конструктну валідність методики входять також дані, що належать до сфери критеріальної та змістовної валідності. Так, критерії, що використовуються при валідизації, несуть інформацію, що дозволяє розкрити область поведінки, якості, представлені в тесті у вигляді конструкту. Для характеристики конструктної валідності необхідні зв'язок з практичними формами діяльності, достовірність прогнозу реальної поведінки. Однак конструктна валідність є якісно більш високим і комплексним рівнем опису тесту, характеризуючи область вимірюваної поведінки в широких психологічних поняттях. Завдяки даним конструктної валідності ми можемо з психологічних позицій закономірно пояснити результати тесту та їх дисперсію, обґрунтувати діагноз, ввівши вимірювану властивість у систему психологічних категорій, прогнозувати поведінку в ширших межах, ніж це задається областю діяльності для якої визначалася змістовна валідність.

Таким чином, провести теоретичну валідизацію методики — це довести, що методика вимірює саме ту властивість, якість, яку вона за задумом дослідника має вимірювати. Для теоретичної валідизації кардинальною проблемою є відношення між психологічними явищами та їх показниками, за допомогою яких ці психологічні явища намагаються пізнати. Така перевірка показує, наскільки задум автора та результати методики збігаються.

Не так складно провести теоретичну валідизацію нової методики, якщо для вимірювання даної властивості вже є методика з доведеною валідністю. Наявність кореляції між новою та аналогічною, вже перевіреною методикою вказує на те, що розроблена методика вимірює ту ж психологічну якість, що і еталонна. Такий прийом особливо часто застосовується в диференціальній психофізіології при створенні методик діагностики основних властивостей нервової системи людини.

Набагато важче провести теоретичну валідизацію методики тоді, коли такий шлях перевірки неможливий. Найчастіше саме з такою ситуацією

стикається дослідник. У таких обставинах лише поступове накопичення різноманітної інформації про досліджувану властивість, аналіз теоретичних передумов і експериментальних даних, значний досвід роботи з методикою дозволяють розкрити її психологічний зміст.

Важливу роль розуміння те, що методика вимірює, грає зіставлення її показників із практичними формами діяльності. Але тут особливо важливо, щоб методика була ретельно опрацьована в теоретичному плані, тобто щоб була міцна, обґрунтована наукова база. Тоді при співставленні методики з взятим із повсякденної практики зовнішнім критерієм, що відповідає тому, що вона вимірює, може бути отримана інформація, що підкріплює теоретичні уявлення про її сутність.

Важливо пам'ятати, що якщо доведено теоретичну валідність, то інтерпретація отриманих показників стає більш ясною та однозначною, а назва методики відповідає сфері її застосування. Що стосується прагматичної валідизації, то вона передбачає перевірку методики з точки зору її практичної ефективності, значущості, корисності, оскільки діагностичною методикою має сенс користуватися тільки тоді, коли доведено, що вимірювана властивість проявляється у певних життєвих ситуаціях, у певних види діяльності. Їй надають великого значення особливо там, де постає питання відбору.

Для проведення прагматичної валідизації методики, тобто для оцінки її ефективності, дієвості, практичної значущості, зазвичай використовується незалежний зовнішній критерій - показник прояву досліджуваної властивості в повсякденному житті. Такими критеріями можуть бути:

- 1) успішність (для тестів здібностей до навчання, тестів досягнень, тестів інтелекту);
- 2) виробничі досягнення (для методик професійної спрямованості);
- 3) ефективність реальної діяльності - малювання, моделювання і т.д. (Для тестів спеціальних здібностей);
- 4) суб'єктивні оцінки (для тестів особистості).

Американські дослідники Д. Тіфін і Е. Маккормік, провівши аналіз використовуваних для доказу валідності зовнішніх критеріїв, виділили чотири їх типи:

- 1) критерії виконання (до них можуть входити такі, як кількість виконаної роботи, успішність, час, витрачений на навчання, темп зростання кваліфікації тощо);
- 2) суб'єктивні критерії (вони включають різні види відповідей, які відображають ставлення людини до чогось або до кого-небудь, його думка, погляди, переваги; зазвичай суб'єктивні критерії отримують за допомогою інтерв'ю, опитувальників, анкет);
- 3) фізіологічні критерії (вони використовуються щодо впливу навколишнього середовища та інших ситуаційних змінних на організм і психіку людини; заміряється частота пульсу, тиск крові, електроопір шкіри, симптоми втоми і т. д.);

- 4) критерії випадковостей (застосовуються, коли мета дослідження стосується, наприклад, проблеми відбору для роботи таких осіб, які менш

схильні до нещасних випадків).

Зовнішній критерій повинен відповідати трьом основним вимогам:

- 1) він має бути релевантним;
- 2) вільним від перешкод (контамінації);
- 3) надійним.

Під релевантністю мається на увазі смислова відповідність діагностичного інструменту незалежному життєво важливому критерію. Іншими словами, має бути впевненість у тому, що в критерії задіяні саме ті особливості індивідуальної психіки, які вимірюються і діагностичною методикою. Зовнішній критерій і діагностична методика повинні знаходитися між собою у внутрішньому змістовному відповідності, бути якісно однорідними за психологічною сутністю.

Якщо, наприклад, тест вимірює індивідуальні особливості мислення, вміння виконувати логічні дії з певними об'єктами, поняттями, то й у критерії слід шукати прояв саме цих умінь. Це в рівній мірі відноситься і до професійної діяльності. Вона має не одну, а кілька цілей, завдань, кожна з яких специфічна і пред'являє свої умови до виконання. З цього випливає існування кількох критеріїв виконання професійної діяльності. Тому не слід проводити зіставлення успішності за діагностичними методиками з виробничою ефективністю в цілому. Необхідно знайти такий критерій, який за характером виконуваних операцій співвідносний з методикою.

Якщо щодо зовнішнього критерію невідомо, релевантний він вимірюється властивості чи ні, то зіставлення з ним результатів психодіагностичної методики стає практично марним. Воно не дозволяє дійти будь-яких висновків, які б дали оцінку валідності методики.

Вимоги волі від перешкод (контамінації) викликаються тим, що, наприклад, навчальна чи виробнича успішність залежить від двох змінних: від самої людини, її індивідуальних особливостей, що вимірюються методиками, та від ситуації, умов навчання, праці, які можуть привнести перешкоди, «забруднити» застосовуваний критерій. Щоб якоюсь мірою уникнути цього, слід відбирати для дослідження такі групи людей, які знаходяться в більш менш однакових умовах. Можна використовувати інший метод. Він перебуває у коригуванні впливу перешкод. Це коригування носить зазвичай статистичний характер. Так, наприклад, продуктивність слід брати не за абсолютними значеннями, а щодо середньої продуктивності робочих, що мають аналогічні умови праці. Коли кажуть, що критерій повинен мати статистично достовірну надійність, це означає, що він повинен відображати постійність та стійкість досліджуваної функції.

Пошуки адекватного критерію, що легко виявляється, відносяться до дуже важливих і складних завдань валідації. У західній тестології багато методик дискваліфіковано тільки тому, що не вдалося знайти відповідного критерію для їх перевірки. Зокрема, у більшій частини анкет дані щодо їхньої валідності сумнівні, тому що важко знайти адекватний зовнішній критерій, що відповідає тому, що вони вимірюють.

Існує кілька видів критеріальної валідності, обумовлених особливостями діагностичних методик, а також тимчасовим статусом зовнішнього критерію.



Однак найчастіше називаються такі:

1) Валідність за одночасністю (поточна валідність, або діагностична валідність) визначається за допомогою зовнішнього критерію, за яким інформація збирається одночасно з експериментами з перевіряемой методики. Інакше кажучи, збираються дані, які стосуються теперішнього часу: успішність під час випробування, продуктивність у цей період і т. буд. З ними зіставляють результати успішності по тесту.

2) «передбачувана» валідність (інша назва - прогностична валідність). Визначається також за зовнішнім критерієм, але інформація щодо нього збирається через деякий час після випробування. Зовнішнім критерієм зазвичай буває виражена в будь-яких оцінках здатність людини до того виду діяльності, на яку він оцінювався за результатами діагностичних випробувань. Хоча цей прийом найбільше відповідає задачі діагностичних методик — передбачення майбутньої успішності, — застосовувати його дуже важко. Точність діагнозу залежить від часу, заданого для такого прогнозування. Чим більше проходить час після виміру, тим більше факторів потрібно враховувати при оцінці прогностичної значущості методики. Проте врахувати всі чинники, що впливають на передбачення, практично неможливо.

3) Ретроспективна валідність. Вона визначається на основі критерію, що відображає події або стан якості в минулому. Може бути використана для швидкого отримання відомостей про передбачувальні можливості методики. Так, для перевірки того, якою мірою хороші результати тесту здібностей відповідають швидкому навчанню, можна зіставити попередні оцінки успішності, минулі експертні висновки і т. д. у осіб з високими і низькими на даний момент діагностичними показниками.

Оцінка валідності методик може мати кількісний і якісний характер.

Для обчислення кількісного показника - коефіцієнта валідності - зіставляються результати, отримані при застосуванні діагностичної методики, з даними, отриманими за зовнішнім критерієм тих же осіб. Використовуються різні види лінійної кореляції (за Спірменом, за Пірсоном).

Скільки випробуваних необхідно для розрахунку валідності? Практика показала, що їх не повинно бути менше 50, проте найкраще більше 200. Часто виникає питання, якою має бути величина коефіцієнта валідності, щоб вона вважалася прийнятною? Загалом зазначається, що достатньо того, щоб коефіцієнт валідності був статистично значущим. Низьким визнається коефіцієнт валідності порядку 0,2-0,3, середнім - 0,3-0,5 і високим - понад 0,6.

Але, як підкреслюють А. Анастасі, К. М. Гуревич та ін, не завжди для обчислення коефіцієнта валідності правомірно використовувати лінійну кореляцію. Цей прийом виправданий лише тоді, коли доведено, що успіх у якійсь діяльності прямо пропорційний до успіху у виконанні діагностичної проби. Позиція зарубіжних тестологів, особливо тих, хто займається профпридатністю та профвідбором, найчастіше зводиться до беззастережного визнання того, що для професії більше підійде той, хто виконав більше завдань у тесті. Але може бути і так, що для успіху в діяльності потрібно мати власністю на рівні 40% рішення тесту. Вищий показник у тесті вже немає значення для професії.

Наочний приклад з монографії К. М. Гуревича: листоноша повинен вміти читати, але чи читає він зі звичайною швидкістю або з дуже великою швидкістю - це вже не має професійного значення.

При такому співвідношенні показників методики та зовнішнього критерію найбільш адекватним способом встановлення валідності може бути критерій відмінностей. Можливий і інший випадок: більш високий рівень якості, ніж це вимагає професія, є на заваді професійному успіху. Так, ще на зорі ХХ ст. американський дослідник Ф. Тейлор знайшов, що найбільш розвинені робітники виробництва мають невисоку продуктивність праці, тобто високий рівень їх розумового розвитку заважав їм високопродуктивно трудитися. І тут для обчислення коефіцієнта валідності більше підійшли б дисперсійний аналіз чи обчислення кореляційних відносин.

Як показав досвід роботи зарубіжних тестологів, жодна статистична процедура не в змозі повністю відобразити різноманіття індивідуальних оцінок. Тому часто для доказу валідності методик використовують іншу модель – клінічні оцінки. Це не що інше, як якісне опис сутності досліджуваної властивості. У цьому випадку йдеться про використання прийомів, що не спираються на статистичну обробку.

У психологічній діагностиці валідність - обов'язкова і найважливіша частина відомостей про методику, що включає:

- 1) дані про ступінь узгодженості результатів тесту з іншими відомостями про досліджувану особу, отриманими з різних джерел (теоретичні очікування, спостереження, експертні оцінки, результати інших методик, достовірність яких встановлена і т. д.);
- 2) міркування про обґрунтованість прогнозу розвитку досліджуваної якості;
- 3) зв'язок досліджуваної області поведінки або особливості особистості з певними психологічними конструктами;
- 4) конкретну спрямованість методики (контингент піддослідних за віком, рівнем освіти, соціально-культурної належності тощо) і
- 5) ступінь обґрунтованості висновків у конкретних умовах використання тесту тощо.

У сукупності відомостей, що характеризують валідність тесту, міститься інформація про адекватність застосовуваної моделі діяльності з точки зору відображення в ній психології, що вивчається, про ступінь однорідності завдань (субтестів), включених у тест, їх сумісності при кількісній оцінці результатів тесту в цілому.

При наведенні даних про валідність розробленої методики важливо точно вказати, який вид валідності мають на увазі (за змістом, за одночасністю і т. д.). Бажано також повідомляти відомості про чисельність та особливості індивідів, на яких проводилася валідація. Така інформація дозволяє користувачам методики вирішити, наскільки валідним є цей прийом для тієї групи, до якої він збирається його застосовувати. Як і у випадку з надійністю, необхідно пам'ятати, що в одній вибірці методика може володіти високою валідністю, а в іншій - низькою. Тому, якщо дослідник планує використовувати методику на вибірці

піддослідних, що істотно відрізняється від тієї, на якій проводилася перевірка валідності, йому необхідно заново провести таку перевірку. Наведений у керівництві коефіцієнт валідності застосуємо тільки до груп піддослідних, подібних до тих, на яких він визначався.

Відомі два шляхи створення психодіагностичних методик: адаптація відомих методик (закордонних, застарілих, з іншими цілями) та розробка нових, оригінальних методик.

### **Питання 3. Адаптація методик**

Адаптація тесту (лат. *adaptatio* пристосування) - комплекс заходів, що забезпечують адекватність тесту в нових умовах його застосування. У вітчизняній психодіагностиці особливої актуальності набуває адаптація зарубіжних тестів.

Можна виділити такі основні етапи адаптації тестів:

- ☐ аналіз вихідних теоретичних положень автора тесту;
- ☐ переклад тесту та інструкцій до нього на мову користувача, що завершується експертною оцінкою відповідності оригіналу;
- ☐ перевірка валідності та надійності тесту, що здійснюється відповідно до психометричних вимог;
- ☐ стандартизація тесту на відповідних вибірках.

Особливі проблеми виникають у зв'язку з адаптацією різних видів опитувальників, а також вербальних субтестів, що входять до складу тестів інтелекту. Основні перешкоди для дослідника пов'язані з мовними та соціокультурними відмінностями між народами різних країн.

Лінгвістичний аспект адаптації тестів означає пристосування його лексики та граматики до вікової та освітньої структури контингентів населення, які плануються для обстеження, облік коннотативного (оцінного) значення мовних одиниць та категорій. Не менші складнощі пов'язані з соціо-культурними відмінностями, що відбиваються в мові особливостям культури того суспільства, в якому створювався тест, важко, а часом і неможливо знайти еквіваленти в іншій культурі.

Повна емпірична адаптація тестів у багатьох випадках не поступається складністю розробки оригінальної методики.

У 60-70-х роках адаптація тестів нашій країні розумілася спрощено, її нерідко зводили до перекладу тієї чи іншої зарубіжної методики, а кращому разі обмежувалися побудовою нормативного розподілу тестових показників. Теоретичні концепції авторів тестів не аналізувалися, дані про їхню надійність та валідність приймалися за справжні.

Останніми роками питання адаптації різних зарубіжних тестів дедалі частіше стають предметом обговорення радянських психологів. Розробляються відповідні рекомендації (Ю. Л. Ханін, 1985; А. Г. Шмельов та В. І. Похілько, 1985; Ю. М. Забродін з співавт., 1987 та ін.). На даному етапі розвитку психодіагностики вимоги, що пред'являються до адаптації тестів, передбачають високу професійну культуру психолога, широке використання сучасної обчислювальної техніки.

### **Критерії оцінки якості тестових процедур**

Вітчизняними психодіагностами сформульовані найбільш загальні вимоги, реалізація яких повинна, на думку авторів, забезпечити оптимізацію тестових методик за критеріями їх наукової достовірності та практичної ефективності.

Критеріями якості психологічних тестів є: надійність, валідність, об'єктивність, прогностична цінність, а також доступність та зручність практичного використання. Про надійність і валідність ми вже говорили, тому коротко розглянемо три останні критерії.

Критерій об'єктивності стосовно психологічних тестів має три значення. Два з них є спільними для всіх психологічних тестів. Це, по-перше, стандартність умов тестування щодо всіх випробуваних і, по-друге, суворе регламентація процедури оцінювання, що обумовлює виставлення однієї й тієї ж оцінки за даний результат будь-якою кількістю оцінювачів. Третє значення - специфічне для особистісних тестів - полягає в забезпеченні незалежності результатів тесту від можливих спроб їх свідомого спотворення піддослідним.

Під доступністю і зручністю практичного використання розуміються, перш за все, відносна простота і короткочасність процедури тестування та оцінювання, відсутність високих вимог до кваліфікації персоналу, що здійснює тестування, а також дешевизна стимульної частини тесту (апаратів, обладнання, буклетів і т. д.).

Щодо прогностичної цінності, то це — одна з основних вимог до тестів — виходить від зацікавлених практиків, і вже багато років є основною «областю вразливості» для критиків-теоретиків. Водночас у вітчизняній літературі є особлива думка про специфіку прогностичної цінності тестових вимірів. Виділяються два типи прогнозу: емпіричний, заснований на принципі простого «перенесення» даних у галузі «малих змін», і теоретичний, заснований на застосуванні теоретичних моделей явищ. Якщо емпіричний прогноз тісно пов'язаний з ідеєю «подібності» ситуацій, суб'єктів, рухів у часі і в просторі завдань, то теоретичний прогноз виявляється пов'язаним з проблемою інтерпретації і «силою і потужністю» теоретичної моделі. Ця проблема зачіпає проблеми «перехресної валідизації» і безпосереднє відношення до зв'язки: «Дані тестового вимірювання - математична модель обробки даних - психологічна модель явища (суб'єкта) - психічна регуляція поведінки - прогноз поведінки». Достовірність результатів психодіагностичного обстеження

Професійний розробник наукового психологічного тесту при його конструюванні слід особливому, науково обгрунтованому, технологічному алгоритму. Щодо кожного ключа висувається статистична гіпотеза. Вона перевіряється великий вибірці піддослідних. Ключ у тесті – результат виявленої експериментальним шляхом статистично значущої залежності. Ця залежність між відповіддю на конкретне питання та реальними подіями та фактами з життя випробуваного. У науковій психодіагностиці етап розробки тестів – це трудомісткий та суворий науковий кореляційний аналіз. Це психометричний експеримент, за допомогою якого розробники перевіряють, чи реально даний тест вимірює те, що він призначений вимірювати. Вибірка щонайменше 50-60 людина. Збирається також об'єктивна інформація про той чинник, який хочемо

виміряти. Багато разів "переробляється".

Алгоритм відомий всім фахівцям, тому вони завжди можуть перевірити обґрунтованість авторського ключа до тесту. Принципова відтворюваність алгоритму розробки тесту та перевірки його ключів (що пов'язують тестові завдання та латентні фактори) підпорядковує психодіагностику об'єктивному науковому критерію "незалежної відтворюваності".

Таким чином, критерії об'єктивності, достовірності наукових психодіагностичних методів закладені в принципах та методах їхньої розробки. Така справа з об'єктивністю у випадку з вимірювальними стандартизованими методиками (з тестами). А як забезпечується об'єктивність при застосуванні експертних методів? Дуже широко застосування в психології та "нестандартизованих" методів збору психологічної інформації, в яких результат безпосередньо залежить від компетентності тих, хто застосовує методику, - це експертні методи. приклад. Тест Роршаха.

Деякі експертні методики припускають підрахунок кількісних показників, але питання про те, чи підводиться чи не поводить відповідь піддослідного під якусь категорію, вирішує тут сам психолог, який застосовує методику. Проективні методики суворо не тести, а експертні методи. Для перевірки їхньої об'єктивності використовуються процедури статистичної перевірки узгодженості незалежних експертних оцінок (метод незалежних суддів).

То чи існують чи ні в психодіагностиці достовірні результати? Так! Але .... тільки, якщо їх використовує професійно підготовлений виконавець. Наприклад. Тест здібностей. При точному виконанні правил тестування ми можемо гарантувати об'єктивність. Це стосується і процедури: дилетант створив стрес, і методика вже вимірює не здібності, а стрес-стійкість особистості.

Користувач методики у разі об'єктивних тестів покликаний дотримуватися суворих методичних розпоряджень, констатувати дотримання умов коректного тестування, і тоді результат можна інтерпретувати відповідно до доданої до тесту інструкції з інтерпретації.

Прогноз же на підставі психологічного діагнозу за тестами на сьогодні можливий тільки з імовірнісною точністю. Це має місце в інших науках. Відмінність полягає у розмірі ймовірнісної помилки прогнозу, просто в психодіагностиці вона вища. Тому наукове обґрунтування тесту включає розрахунок і явне опублікування стандартної помилки вимірювання для даного тесту.

### **Самостійна розробка тестової методики**

Самостійна розробка тестової методики зазвичай має такі етапи:

1. Вибір предмета (яви) та об'єкта дослідження (контингенту).
2. Вибір виду тесту (об'єктивний, суб'єктивний, проективний), типу завдань (з запропонованими відповідями, з вільними відповідями) та шкал (числові, вербальні, графічні).
3. Підбір первинного банку завдань. Може здійснюватися двома шляхами: питання формулюються, з теоретичних уявлень про вимірюваному явищі (факторно-аналітичний принцип) чи питання підбираються відповідно до їх дискримінативності, тобто. здатністю відокремлювати випробуваних за

наявністю необхідної ознаки (критеріально-ключовий принцип). Другий принцип є ефективним у тестах відбору (наприклад, професійного або клінічного).

4. Оцінка завдань первинного банку (оцінка змістовної валідності тесту, тобто відповідності кожного із завдань вимірюваному явищу і повноти охоплення досліджуваного явища тестом в цілому). Проводиться за допомогою методу експертної оцінки.

5. Попереднє тестування, формування банку емпіричних даних.

6. Емпірична валідація тесту. Проводиться за допомогою кореляційного аналізу оцінок тесту і показників за зовнішнім параметром досліджуваного властивості (наприклад, шкільної успішності при валідації тесту інтелекту; лікарського діагнозу при валідації клінічних тестів; даних інших тестів, валідність яких відома, та ін).

7. Оцінка надійності тесту (стійкості результатів до дії випадкових факторів, зовнішніх та внутрішніх). Найчастіше оцінюється ретестова надійність, надійність частин тесту і надійність паралельних форм, якщо такі існують. Методика визнається надійною, якщо коефіцієнт кореляції результатів (первинного і повторного тестування; однієї та іншої частин тесту; однієї та іншої паралельних форм) становить не менше 0,75. При нижчому показнику надійності проводиться коригування завдань тесту, переформулювання питань, що знижують надійність.

8. Проводиться стандартизація тесту. Стандартизація оцінок передбачає перетворення нормальної чи штучно нормалізованої шкали первинних оцінок (емпіричних значень досліджуваного показника) на оцінки шкальні (відбивають місце у розподілі результатів вибірки піддослідних). Види шкальних оцінок: стіни (1-10), станайни (1-9), Т-оцінки (10-100) та ін.

9. По можливості визначається прогностична валідність.

#### **Алгоритм збору інформації про психодіагностичні методики**

Надійність і об'єктивність психологічного діагнозу визначається адекватністю методичних засобів, які застосовуються в ході діагностики, точністю процедури їх застосування і, звичайно, повнотою та глибиною інтерпретації отриманих результатів. Велику допомогу в цьому може надати дотримання алгоритму збору інформації про психодіагностичні засоби. Один із можливих алгоритмів наводиться нижче.

## ІНФОРМАЦІЙНА ПСИХОДІАГНОСТИЧНА КАРТКА

1. Назва методики

\_\_\_\_\_

2. Автор методики (або автор адаптації, модифікації)

\_\_\_\_\_

3. Джерело опису (із зазначенням прізвища та ініціалів автора збірки та редактора, місця, видавництва та року видання) \_\_\_\_\_

4. Призначення \_\_\_\_\_

\_\_\_\_\_

5. Стимульний матеріал та обладнання (у разі авторської розробки наводиться текст та стимульний матеріал) \_\_\_\_\_

6. Інструкція \_\_\_\_\_

\_\_\_\_\_

7. Процедура обстеження \_\_\_\_\_

\_\_\_\_\_

8. Спосіб обробки \_\_\_\_\_

\_\_\_\_\_

9. Реєстровані показники \_\_\_\_\_

10. Середньогрупові, шкільні або рівневі значення \_\_\_\_\_

\_\_\_\_\_

11. Інтерпретація показників \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_