

АНАЛІТИЧНІ РЕВЮ

УДК 159.98

DOI <https://doi.org/10.32782/psy-2023-2-9>**Троценко Андрій***аспірант кафедри соціальної та практичної психології
Житомирського державного університету імені Івана Франка*

ОСОБЛИВОСТІ РОЗРОБКИ ТА ВАЛІДИЗАЦІЇ ТЕСТІВ ОСОБИСТОСТІ

Анотація. Метою статті є висвітлення основних принципів у розробці тестів особистості. У матеріалі зазначено, що радянська політика не сприяла розвитку психометрії в Україні. У результаті маємо дефіцит валідних і надійних методик. Запозичення закордонних тестів відбувалося без належної стандартизації. Сучасні дослідження та розвиток психології неможливі без об'єктивних інструментів. У зв'язку з цим автором розглянуті основні етапи розробки тесту. Наголошується на необхідності чіткого виокремлення досліджуваного явища та детально вивчення його у всіх можливих джерелах. Лише після цього можна починати з розробки самих тверджень. Наведений приклад таблиць у формі решітки полегшує формулювання запитань до тесту. Розкриті основні правила цього процесу. Зазначено, що необхідно формулювати зрозумілі й однозначні твердження, які будуть вимірювати конкретну поведінку, а не почуття. Вони не повинні мати подвійного заперечення. Від якості тверджень залежить валідність тесту. Після їхньої розробки проводять пілотажне дослідження. Отримані результати дозволяють визначити індекс ефективності кожного твердження та вилучити непотрібні. Для визначення кореляцій мінімальний розмір вибірки може становити 85 респондентів, якщо рівень статистичної значимості не більше 0,05. У статті представлені основні види валідності та способи їхнього встановлення. Валідний тест вимірює те, для чого призначений. Це один із критеріїв якості методики. Також показаний приклад обрахунку коефіцієнта кореляції Пірсона між шкалами двох різних методик. Зазвичай рекомендують досягти кілька видів валідності. Їхній вибір залежатиме від можливостей і цілей дослідження.

Ключові слова: тест, розробка, психометрія, поведінка, шкала, вибірка, коефіцієнт кореляції, валідність, особистість, рівень статистичної значимості, стандартне відхилення, критерій.

Trotsenko Andrii*Postgraduate Student of the Department of Social and Practical Psychology
of Zhytomyr Ivan Franko State University*

FEATURES OF DEVELOPMENT AND VALIDATION OF PERSONALITY TESTS

Abstract. The purpose of the article is to highlight the fundamental principles of developing personality tests. The material acknowledges that Soviet policy did not contribute to the development of psychometrics in Ukraine. As a result, a deficiency of valid and reliable methodologies can be observed. The foreign tests were borrowed without undergoing the process of standardization. Modern research and the development of psychology are impossible without applying objective instruments. Due to this, the author reviewed the key stages of test development. It is important to highlight the necessity of a clear distinguishing of the studied phenomenon and a detailed research of it from all possible sources. Only after this, the development of the very statements can be started. An example of the grid-form tables is provided to facilitate the formulation of questions for the test. The fundamental rules of this process are outlined. It is indicated that it is necessary to formulate clear and straightforward statements that measure specific behavior rather than feelings. They shouldn't contain double negatives. The quality of statements determines the validity of the test. After their development, a pilot study is conducted. The obtained results allow to determine the effectiveness index for each statement and to eliminate the unnecessary ones. A minimum sample size of 85 respondents is suggested to determine correlations if the level of statistical significance is not greater than 0.05. The article presents the main types of validity and methods for establishing them. A valid test measures what it is designed to measure. It's one of the criteria for the methodology's quality. An example of calculating the Pearson correlation coefficient between the scales of two different methods is also demonstrated. It is usually recommended to achieve multiple types of validity, with their selection depending on the research's possibilities and goals.

Keywords: test, development, psychometrics, behaviour, scale, sample, correlation coefficient, validity, personality, level of statistical significance, standard deviation, criterion

Вступ. В Україні психометрія перебуває на етапі становлення. Це пов'язано з радянським минулим країни. У 1936 в СРСР була видана ухвала ЦК ВКП (б) «Про педологічні перекручення в системі Наркомпросу», яка обмежувала використання тестів. Радянська влада не сприяла розвитку психометрії. До 1970-х років тестологія переживала «період застою». Слово «тестування» радянська влада сприймала як приховану ідеологічну диверсію. Після здобуття Україною незалежності почали використовувати тести, зокрема в освіті, було впроваджено зовнішнє незалежне оцінювання. Науковці вважають, що вітчизняній психодіагностиці потрібно 15-20 років, щоби наздогнати рівень закордонної. У період від 1970-х до 1980-х років фахівці використовували переклади тестів без належної їхньої стандартизації. Наразі багато методик в Україні є застарілими, на відмінну від нових закордонних разом із сучасними дослідженнями [2; 7].

Інші країни активно розвивали психометрію та мають велику наукову базу та психометричні інструменти, які широко впроваджують у різних напрямках, наприклад, в освіті при відборі на роботу, у військовій сфері, психотерапії. Уряд США щорічно виділяє 7 млн. доларів на розробку нових методів психодіагностики для збройних сил [7]. Якщо відкрити базу даних Американської психологічної асоціації, можна побачити безліч наукових статей присвячених розробці різних методик із психодіагностики.

Розвиток вітчизняної психології залежить від сучасних об'єктивних досліджень і точних вимірювань, які можливі лише завдяки психометричним інструментам. Вивчення особистості є центральним у психології. Це дозволить нам краще інтегруватися в світову систему знань про людину.

Відомий англійський біолог Ф. Гальтон організував антропометричну лабораторію у Лондоні та проводив платні вимірювання. Його вважають засновником психометрії. Він першим використав анкету для збору інформації; самостійно розробляв тести для інших досліджень; застосував нові статистичні та математичні методи у психології. У праці «Природна спадковість» вперше згадано про коефіцієнт кореляції як міри зв'язку між ознаками. Згодом його учень К. Пірсон розробив відповідну математичну формулу [9].

Значущий внесок у розробку тестів особистості та психометрію зробили британські психологи П. Клайн та Р. Кеттел. За словами П. Клайна, оцінити поведінку й особистість за допомогою

опитувальників є набагато складнішим завданням, на відмінну виміру фізичних явищ [5].

Серед сучасних розробок особливу увагу варто звернути на розроблений «Скринінг потенційно травматичних подій та посттравматичного стресового розладу у дітей та підлітків CATS-2» (The Child and Adolescents Trauma Screen 2). Це нова версія оригінального CATS. Запитання скринінгу створені на основі опису посттравматичних симптомів в DSM 5 та МКХ 11. Він розроблений групою таких німецьких науковців як E. Redican, C. Sachser, E. Pfeiffer медичного центру Університету Ульма у співпраці з українським психіатром та психотерапевтом Д. Марценковським [19]. Дослідження проведено у 2022 році серед української вибірки, яка включає дітей і дорослих. Має три окремі варіанти: 1) для дітей та підлітків від 7 до 17 років; 2) для піклувальників дітей та підлітків від 7 до 17 років; 3) для піклувальників дітей від 3 до 6 років. Опитувальник дозволяє визначити симптоми ПТСР у дітей і підлітків, діагностувати травматичні події. CATS-2 має високі психометричні показники. Великою перевагою цього скринінгу є те, що він наявний у вільному доступі та безкоштовний для використання, на відмінну від оригінальної форми, яка потребує ліцензування для використання. Прикметно, що методика розроблена для українців. Розробники зазначають, що діагностика ПТСР не повинна спиратися лише на проведенні скринінгу CATS-2, а включати також оцінку клінічним спеціалістом, наприклад, лікарем-психіатром. Методику можна використовувати при роботі з травмованими дітьми під час психотерапії. У безпечному середовищі вона не призводить до повторної травматизації [4].

До останніх досліджень можна віднести психодіагностичну методику «Парціальні позиції життєстійкості особистості» (І. Бринза, О. Кузнєцова). Цей тест розроблений у 2020 році. Він визначає рівень життєстійкості особистості під час перебування у складних життєвих обставинах. Розробниками використані такі методи перевірки валідності тесту: а) очевидна, б) конструктна, в) конвергентна, г) конкурентна (дискримінативна). Для перевірки надійності методики використано метод «ділення тесту навпіл». Для перевірки константності (незалежності результатів тесту) окремими психодіагностами проведено два вимірювання на одній вибірці та встановлений між ними відповідний статистичний зв'язок. Математичні обчислення валідності здійснені за допомогою факторного аналізу, коефіцієнту кореляції Пірсона, методу перевірки дискримінативності [1].

Мета статті полягає у висвітленні основних принципів у розробці та валідації особистісних тестів. Викладений матеріал покликаний полегшити роботу психологам при розробці тестових матеріалів.

Виклад основного матеріалу дослідження. Перед розробкою методики необхідно визначитися з її тематикою, сформулювати ціль та чітко виокремити психологічне явище, яке потрібно виміряти. Детально вивчити досліджуване явище в усіх можливих джерелах.

Тести можуть мати різні форми відповідей. Найбільш поширеними є дихотомічні шкали з двома відповідями «Так» – «Ні» та трихотомічні «Так» – «Складно відповісти» – «Ні». Також достатньо часто використовують популярну шкалу відповідей Лейкерта, яка має 7 градацій, де 1 означає мінімальну вираженість досліджуваної ознаки, 6 – середній показник, а 7 – максимальний. Така кількість градацій відповідей дозволяє обстежуваним адекватно виразити себе, а сам тест більш чутливий до індивідуальних відмінностей.

П. Клайн наводить такі правила розробки об'єктивних особистісних тестів:

1. Не розкривайте респондентам, яку саме рису особистості або ознаку досліджуєте, адже отримані дані можуть бути спотвореними. Сприйняття власного образу може відрізнитися від дійсності.

2. Формулюйте зрозумілі та недвозначні запитання або твердження.

3. Запитання мають відбивати конкретні ознаки поведінки.

4. У кожному завданні має бути одне запитання або твердження. Вони не повинні мати подвійного заперечення, тому що це їх ускладнює.

5. Уникайте слів, які описують частоту дій. Це дуже суб'єктивно. Кожен розуміє такі відповіді своєрідно. Наприклад, слова «часто» або «зрідка» краще замінити більш конкретними маркерами, такими як «двічі на тиждень», або «останні чотири дні».

6. Тест має складатися з опису конкретної поведінки, а не почуттів.

7. Установка в інструкції до тесту надавати відповідь, яка перша спадає думку. Якщо досліджувані починають думати над питанням, то це може свідчити про захисні механізми, що негативно впливає на результати тесту. Якщо запитання сформульоване в контексті конкретної поведінки, то відповідь респондента буде швидкою та точною [8].

Особливо важливе значення у розробці запитань до тесту має зрозумілість та конкретність. При ефективному завданні респонденти довго не замислюються. Самі запитання бажано формулювати такими, щоб слова, які зазначені, були актуальними та зрозумілими. П. Клайн наголошує, що однією з проблем опитувальників і тестів є установка на невизначені або середні відповіді, такі як «не знаю», «не впевнений», «інколи», «складно відповісти». Досліджувані можуть обирати їх як безпечний компроміс, що знижує валідність. Деякі досліджувані схильні обирати крайні відповіді, незалежно від змісту питання. Це можна нівелювати, якщо розробляти питання, де кількість прямих й обернених (із протилежним змістом) відповідей, які збігаються з ключем, буде приблизно однакова [5]. Також обстежувани можуть давати соціально бажані відповіді. Ця загроза валідності може бути незначущою, якщо питання складені належним чином і правильно надана інструкція до тесту [8].

Тепер перейдемо безпосередньо до самої розробки питань тесту. Спершу потрібно перерахувати конкретні риси й особливості поведінки, які вивчаємо та описані у психологічній або художній літературі. Необхідно прочитати якомога більше інформації про сферу дослідження. Запитання мають бути простими і недвозначними. Слід уникати подвійних заперечень. Полегшити цей процес може таблиця у формі решітки. Наведемо приклад її наповнення для таких досліджуваних параметрів як «невротичність», «психотичність», «конформність» та «ізолюваність» (табл. 1). Оптимальна кількість комірок у таблиці може бути від 16 до 25. Це приблизно 4-5 рядків і 5 стовпців [3]. Таку решітку можна порівняти із сіткою для риболовлі: великі комірки затримують тільки «велику рибу», а «маленьку» пропускають. Тому, якщо така таблиця буде малою, тест не буде виявляти специфічні індивідуальні відмінності, а показуватиме загальні характеристики.

Далі зміст комірок необхідно переробити у запитання чи завдання. Для кожної ознаки потрібно визначити кількість питань, враховуючи важливість у порівнянні з іншими. Встановити значення досліджуваних ознак можна за допомогою приблизних відсоткових співвідношень. Для цього використаємо табл. 2. Варто зазначити, що відповідь на пряму не завжди може бути індикатором особистісної змінної, оскільки на неї можуть впливати інші фактори. Зв'язок із вказаною відповіддю виражається суто статистично [2].

Після розробки завдань можна залучити кількох експертів для оцінки. Це може бути 2-3

Таблиця 1

Структура тесту

Елементи	Ознаки поведінки або риси особистості				
Невротичність	Тривожність	Складно засинати	Тремтіння	Підвищене серцебиття	Невпевненість
Психотичність	Відсутність емпатії	Маніпуляції	Ризикована поведінка	Заподіяння шкоди іншим	Вчинення протиправних дій
Конформність	Пристосування	Уникнення конфліктів	Вчиняти як більшість	Відсутність власної думки	Стриманість
Ізольованість	Уникнення спілкування	Усамітнення	Інтроверто-ваність поведінки	Мале коло спілкування	Самотність

Таблиця 2

Визначення відсоткового значення ознак

Елементи	Ознаки поведінки або риси особистості					К-ть завдань
Невротичність	Тривожність 40%	Невпевненість 20%	Підвищене серцебиття 20%	Складно засинати 10%	Тремтіння 10%	20
	8	4	4	2	2	
Психотичність	Відсутність емпатії 40%	Маніпуляції 30%	Заподіяння шкоди іншим 10%	Ризикована поведінка 10%	Конфлікт із законом 10%	20
	8	6	2	2	2	
Конформність	Пристосування 40%	Уникнення конфліктів 20%	Вчиняти як більшість 20%	Стриманість 10%	Відсутність власної думки 10%	20
	8	4	4	2	2	
Ізольованість	Уникнення спілкування 40%	Інтроверто-ваність поведінки 20%	Усамітнення 20%	Мале коло спілкування 10%	Самотність 10%	20
	8	4	4	2	2	
К-ть завдань	32	18	14	8	8	80

фахівця, які безпосередньо пов'язані або дотичні до предмету дослідження [8]. П. Клайн зазначає, що ймовірність об'єктивної діагностики психометричних показників підвищується при збільшенні розміру тесту [5].

Наступний крок – проведення пілотажного дослідження. Вибірка повинна складатися з типових досліджуваних, для яких призначений складений тест. Таке дослідження дозволяє визначити недоліки методики, вибракувати неефективні питання. Кількість обстежуваних повинна бути вдвічі більшою за кількість завдань [3].

Згідно з дослідженнями J. Cohen, мінімальна кількість вибірки для виявлення кореляційних зв'язків може становити 85 осіб при умові, що рівень статистичної значимості (level of significance) на рівні не більше 0,05 [13]. Цей показник означає, що є 5% ймовірності помилки при узагальненні отриманих даних, а виявлений зв'язок між ознаками є випадковим і недостовірним [6]. Прийнято вважати, що чим він менший,

тим краще. П. Клайн вважає, що кількість обстежуваних має бути не менше 200 осіб. У великих вибірках систематична помилка вибірки буде меншою. При такій кількості обстежуваних стандартні похибки кореляції є незначними [5]. Вибірка повинна бути репрезентативною, тобто включати представників різного віку, статі, рівня освіти, професій тощо. Найбільш простим методом відбору обстежуваних є «вибірка за можливістю». При цьому тестують усіх, хто відповідає цільовій групі.

На основі отриманих даних у пілотажному дослідженні проводимо аналіз завдань для відсіювання найкращих. Це дозволяє підвищити валідність тесту. Визначаємо *індекс ефективності завдання*. Для цього кількість обстежуваних, які надали правильну відповідь на завдання (відповідно до ключа тесту) ділимо на загальну кількість усіх учасників. Індекс для кожного питання визначаємо окремо. Отриманий показник допустимий в інтервалі від 0,25 до 0,75. Бажано,

щоби він наближався до середнього значення 0,50. Якщо індекс менше 0,25, це означає, що на нього правильно відповіло замало обстежуваних, більше 0,75 – забагато. Завдання, які виходять за дані межі інтервалу, є неефективними, тому їх потрібно вилучити з фінальної версії тесту. Вилучення таких завдань може зробити коротшу форму більш валідною [3]. У випадку відсутності у методиці ключа правильних відповідей можна визначити моду по кожному питанню. Наприклад, якщо це опитування, в якому респондент за кожну відповідь «Так» отримує 1 бал, а за відповідь «Ні» матиме 0 балів. Отриманий результат буде відповідати індексу ефективності. Одиниця буде знаходитися в інтервалі від 0,25 до 0,75.

Кожна розроблена психодіагностична методика повинна відповідати таким основним психометричним вимогам як валідність та надійність. Це своєрідні критерії якості тесту. Такі вимоги досить умовні, адже у кожного розробника методик вони можуть дещо відрізнятися. Комітет із психологічного тестування Американської психологічної асоціації рекомендує звернути увагу на чотири типи валідності: прогностична, конкурентна, змістовна та конструктивна [15]. У межах цієї статті розглянемо основні види валідності, які виділяють у науковій літературі.

А. Анастасів вважає, що *валідність* – це поняття, яке визначає те, що вимірює тест і наскільки якісно це здійснюється [12]. Тобто цей критерій показує, наскільки отримані результати відповідають меті методики. Валідний тест вимірює те, для чого він був призначений. Є декілька способів доведення валідності. Залежно від потреб і можливостей можна обрати один або кілька методів визначення даного критерію [8].

Очевидна і змістовна валідність. При очевидній валідності суб'єктивно оцінюють призначення тесту. Він має бути релевантний до умов, в яких буде застосований. Наприклад, тест для дітей дошкільного віку може бути не зовсім доречним для здобувачів вищої освіти. Очевидну валідність можна перевірити, опитавши кількох осіб, для яких він призначений. Вони можуть оцінити кожен елемент і повідомити, чи відноситься він до того, що тест вивчає. При змістовній валідності елементи тесту мають дійсно відноситися до того, що вивчаємо. Для того, щоб її визначити, можна провести опитування серед експертів із релевантної сфери дослідження. Вони повинні оцінити, чи відповідають розроблені елементи тому, що ми хочемо дослідити. Якщо в них виникають обґрунтовані сумніви з приводу чогось в тесті, то це варто

виключити. Експерти мають бути обрані із тієї популяції, серед якої проводиться дослідження. Змістовна валідизація не зовсім підходить для тестів особистості. Психологічні процеси у кожного можуть суттєво відрізнятися. Один і той самий тест може вимірювати різні ознаки. Тому перевіркою лише змісту неможливо встановити ці дані. Така валідність краще підходить для тестів досягнень і здібностей [5; 8; 12].

Конкурентна валідність – це кореляція результатів розробленого тесту з результатами одного або кількох інших, валідність яких вже встановлена. Обрана методика повинна оцінювати приблизно те саме, що розроблений тест. Важливо, щоби тест був диверсифікованим. Якщо він матиме дуже сильну кореляцію, це стане дублюванням. Кореляція – це статистична залежність між випадковими величинами, що носить ймовірнісний характер [11]. У таблиці 3 наведений приклад визначення кореляції Пірсона. Нехай «Тривожність» буде шкалою тесту «А», а «Невротизм» шкалою іншого валідного тесту «Б». У колонках x і y позначені індивідуальні відхилення показників від середнього арифметичного. Квадрати даних відхилень наведені в наступних колонках (x^2 та y^2), а дані останньої колонки (xy) призначені для визначення стандартних відхилень (Std. Deviation) за обома шкалами методик [12].

$$SD_x = \sqrt{\frac{126}{10}} = \sqrt{12,60} = 3,54 \quad SD_y = \sqrt{\frac{84}{10}} = \sqrt{8,40} = 2,89$$

$$r_{xy} = \frac{\sum xy}{N * SD_x * SD_y} = \frac{67}{10 * 3,54 * 2,89} = \frac{67}{102,30} = 0,65$$

У табл. 3 показано, що стандартне відхилення (SD) учасника №1 становить 2 за шкалою «Тривожність» та 0 за шкалою «Невротизм». Стандартне відхилення (SD) показує, наскільки показник відхиляється відносно середнього арифметичного. Чим більше цей показник відносно середнього, тим слабшою буде кореляція. Між двома наведеними шкалами коефіцієнт кореляції (r_{xy}) становить 0,65, що свідчить про сильний і прямий зв'язок. Це означає, що існує закономірність. Наприклад, при збільшенні показнику обстежуваного у шкалі «Тривожність» збільшується показник в шкалі «Невротизм», або навпаки. Кореляція також повинна мати статистичну значимість не більше 0,05. Це доводить конкурентну валідність розробленої шкали.

Розробники тестів зазначають, що розрахунки вручну дозволяють краще збагнути смисл коефіцієнта кореляції та суть розроблених питань [5].

Обчислення коефіцієнту кореляції Пірсона

Учасник	Тривожність (тест А)	Невротизм (тест Б)	x	y	x ²	y ²	xy
1.	11	10	2	0	4	0	0
2.	8	12	-1	2	1	4	-2
3.	7	8	-2	-2	4	4	4
4.	12	10	3	0	9	0	0
5.	17	15	8	5	64	25	40
6.	4	5	-5	-5	25	25	25
7.	6	10	-3	0	9	0	0
8.	9	14	0	4	0	16	0
9.	10	7	1	-3	1	9	-3
10.	6	9	-3	-1	9	1	3
Сума Σ	90	100	0	0	126	84	67
Середнє	9	10					

Критеріальну валідність можна отримати у результаті встановлення та порівняння зв'язку між наявними даними про обстежуваного з конкретного тесту. Критерієм для порівняння може бути будь-що. Наприклад, оцінки з дисциплін, рейтинги, результати професійної діяльності. Суб'єктивні оцінки також можна використовувати. Вони корисні для тестів особистості, тому що встановлення об'єктивних критеріїв має багато труднощів. Щоби виявити зв'язок між змінними при порівнянні двох груп, використовують *t*-критерій (критерій Стьюдента) для незалежних вибірок. Якщо потрібно порівняти більше двох груп, застосовують одномірний дисперсійний аналіз (ANOVA). Односторонній *t*-критерій буде доцільний у випадку, коли передбачаємо, що одна з груп зміниться у певному напрямку. У протилежному випадку використовуємо двосторонній *t*-критерій. Він є більш строгим, тому що перевіряє відмінності в обидві сторони. Кінцева версія розробленої шкали матиме показники залежної змінної [5; 8; 12].

Прогностичну валідність можна встановити шляхом вивчення кореляцій між показниками тесту та властивістю досліджуваного, яка буде проявлена через певний час. Ця властивість буде критерієм для порівняння. Наприклад, оцінка продуктивності працівника під час роботи. Можна визначити кореляцію між отриманими оцінками та показниками тесту. Якщо продуктивність включає дві категорії, наприклад, ті хто успішно виконав завдання і не успішно, то для визначення зв'язків використовують *t*-критерій. Залежною змінною будуть показники розробленого тесту. Якщо продуктивність має більше двох категорій, застосовуємо одномірний дисперсійний аналіз (ANOVA) [5; 8; 12].

Конструктивна валідність відбиває те, наскільки результати тесту можуть бути релевантні до розробленої теоретичної концепції. Вона включає різні типи валідності. Конструкт – припущена властивість особи, яку виявляють у результаті тестування. Для її встановлення необхідно детально описати змінну або властивість, яку тест буде вимірювати, розробити детальні гіпотези, які будуть підтверджені методикою. Тобто конструктивна валідність інтегрує не тільки конкретні методи дослідження, а й бачення самого розробника. Її вивчають, якщо відсутні чіткі критерії для виміру досліджуваної ознаки [8; 12]. Л. Кронбах вважає, що конструктивна валідність найважливіша та іноді необхідна для будь-якого типу психологічного тесту [15; 16]. Г. Айзенк та Д. Гілфорд вважають, що для перевірки гіпотези про конструкти доцільно використовувати факторний аналіз [17; 18].

Конвергентна (дискримінантна) валідність є частиною конструктивної, але має свою особливість, яку варто окремо зазначити. Для досягнення валідності визначають відсутність кореляції із параметрами, які нерелевантні до тесту, який розроблений. При цьому методі використовують тести, які виявляють зв'язок з ознаками, які діаметрально протилежні. Л. Кронбах стверджує, що навіть слабкі кореляції важливі та свідчать про орієнтовну вираженість ознаки у певний напрям [15; 12].

Від валідності безпосередньо залежить ефективність та якість методики. У США проводили дослідження, яке показало, що при більшій валідності тесту збільшувалась якість найму кращих працівників на роботу. Якщо коефіцієнт становив 0,80, то дохідність за рік коливалася від \$16,5 до \$97,2 млн., якщо він був 0,50, то прибуток був зменшувався в межах від \$5,6 до \$33,3 млн. [12].

Валідність тесту не може визначати лише один метод. Для її доведення потрібно використати кілька із зазначених способів. Переважно рекомендують досягти змістовної, очевидної валідності та один додатковий вид, залежно від можливостей та цілей.

Висновки. Прогрес психології як науки залежить від об'єктивного та точного вимірювання психометричних показників. Для того, щоби досягнути цього необхідно розробляти ефективні та валідні методики. Сьогодні маємо дефіцит вітчизняних тестів, а закордонні методики використовують без

належної стандартизації. Розробка тесту – тривалий і складний процес. Він починається з вибору теми методики, виокремлення психологічного явища, яке досліджуємо. Формулювання питань є одним із найважливіших етапів. Від їх змісту безпосередньо залежить валідність тесту. Вона є одним з основних критеріїв якості тесту. Це означає, що він вимірює те, для чого призначений. Проведення пілотажного дослідження дозволяє відсіяти неефективні питання методики та вилучити їх. Існують різні види валідності, які використовують залежно від цілей і можливостей дослідження.

Список використаних джерел

1. Бринза І., Кузнецова О. Психодіагностична методика "Парціальні позиції життєстійкості особистості": результати розробки та апробації. *Вісник Київського національного університету імені Тараса Шевченка. Психологія*. 2020. № 11. Т. 1. С. 14–22. DOI: [https://doi.org/10.17721/BSP.2020.1\(11\).3](https://doi.org/10.17721/BSP.2020.1(11).3).
2. Бурлачук Л. В. Психодіагностика особистості: понятійний апарат та методи дослідження. *Психологія і суспільство*. 2014. № 4. С. 85–102.
3. Галян І. Психодіагностика : навчальний посібник. Київ : Академвидав, 2009. 465 с.
4. Заксер С. Скринінг потенційно травматичних подій та посттравматичного стресового розладу у дітей та підлітків CATS-2. *Психічне здоров'я України*. URL: <https://youtu.be/c1CtOIHcBMw?si=WILRroq0Hj1DEoTh>.
5. Клайн П. Довідковий посібник із конструювання тестів. Введення у психометричне проектування. Київ: ПАН ЛТД, 1994. 276 с.
6. Климчук В. О. Математичні методи у психології : навчальний посібник для студентів психологічних спеціальностей. Київ : Освіта України, 2009. 288 с.
7. Кухар Л. О., Сергієнко В. П. Конструювання тестів : навчальний посібник. Луцьк : 2010. 182 с.
8. Ловенталь К. М., Льюїс К. А. Введення в психологічні тести і шкали. Київ : *Науковий світ*, 2022. 210 с.
9. Мазяр О. В. Основи психометрики : курс лекцій. Житомир : Вид-во ЖДУ, 2022. 108 с.
10. Мазяр О. В., Кириченко В. В. Шкала особистісної динаміки: перевірка надійності паралельних форм. *Габітус*. 2023. № 46. С. 176–180.
11. Руденко В. М. Математична статистика : навчальний посібник. Київ : Центр учбової літератури, 2021. 304 с.
12. Anastasi A., Urbina S. Psychological testing. Upper Saddle River : Prentice Hall, 1997. 736 p.
13. Cohen J. A power primer. *Psychological Bulletin*. 1992. Vol. 1(112). P. 155–159. DOI: <https://doi.org/10.1037/0033-2909.112.1.155>.
14. Cohen J. Things I have learned (so far). *American Psychologist*. 1990. Vol. 12(45). P. 1304–1312. DOI: <https://doi.org/10.1037/0003-066X.45.12.1304>.
15. Cronbach L. J., Meehl P. E. Construct validity in psychological tests. *Psychological Bulletin*. 1955. Vol. 52(4). P. 281–302. DOI: <https://doi.org/10.1037/h0040957>.
16. Cronbach L. J. A validation design for qualitative studies of personality. *Journal of Consulting Psychology*. 1948. Vol. 12(6). P. 365–374. DOI: <https://doi.org/10.1037/h0059005>.
17. Guilford J. P. Factor analysis in a test-development program. *Psychological Review*. 1948. Vol 55(2). P. 79–94. DOI: <https://doi.org/10.1037/h0056564>.
18. Eysenck H. J. Criterion analysis – an application of the hypothetico-deductive method in factor analysis. *Psychological Review*. 1950. Vol. 57(1). P. 38–53. DOI: <https://psycnet.apa.org/doi/10.1037/h0057657>.
19. Redican E., Sachser C., Pfeiffer E., Martsenkovskiy D., Hyland P., Karatzias T., Shevlin M. Validation of the Ukrainian caregiver-report version of the Child and Adolescent Trauma Screen (CATS) in children and adolescents in Ukraine. *Psychological Trauma: Theory, Research, Practice, and Policy*. 2023. Vol. 15. DOI: <https://doi.org/10.1037/tra0001570>.