

ARTYKUŁY [Articles]

PSYCHOMETRIA I FUNKCJA POMIARU W PSYCHOLOGII

Bogdan Pietrulewicz

Instytut Psychologii

Uniwersytet Kazimierza Wielkiego, Bydgoszcz

Uniwersytet Laval, Quebec

PSYCHOMETRICS AND MEASUREMENT FUNCTION IN PSYCHOLOGY

Summary. The first attempts to measure various psychological functions in the context of professional selection were observed in ancient China. Many years have passed before Wolff in 1732 suggested creation of a new branch inside empirical psychology – psychometrics, which aimed at applying mathematical methods to examining different psychological characteristics. Prematurely born concept was nearly completely forgotten; the followers of the idea do not even mention its founder. The elements of the modern theory of measurement develop application of science model in psychology. The following paper gives a critical view and discusses the question, considering its fundamental methodological aspects.

Wprowadzenie

Od dawna uświadamiano sobie istnienie różnic międzyosobowych, a tylko w nielicznych przypadkach starano się je negować. Przykładem mogą być pierwsze próby stosowania systemu oceny przydatności zawodowej w starożytnej historii. Weiner (1987) podaje, że już 2200 lat przed Chrystusem urzędnicy państwowi w Chinach byli poddawani co 3 lata badaniom, które miały wykazać ich zdolność do dobrego funkcjonowania na danym stanowisku. Poczynając od 1115 roku przed Chrystusem, organizacja tego procesu została sformalizowana i kandydaci przechodzili przez różne próby, w zakresie muzyki, łucznictwa, jeździectwa, pisania tekstów, arytmetyki, ceremonialnych rytuałów w życiu publicznym i prywatnym. Do dzisiaj zachowały się zabytki architektoniczne będące miejscem tego rodzaju aktywności egzaminacyjnej.

Adres do korespondencji: Bogdan Pietrulewicz,
e-mail: bogdan.pietrulewicz@ulaval.ca

W kulturze chrześcijańskiej nie znajdujemy zbyt wielu przykładów pomiaru funkcjonowania ludzi: założenie o dualizmie ciała i duszy praktycznie wykluczało ewaluację świata duchowego w kategoriach materialnych lub doświadczalnych. Dopiero w 1732 roku Wolff, odkrywając psychologię empiryczną jako nową dziedzinę nauki dostrzega, że wrażenia, spostrzeżenia, myślenie i uczucia mogą mieć odniesienie do świata obiektywnego i rzeczywistości wymierzalnej. Założenia te pozwoliły mu sformułować postulat stworzenia nowej dyscypliny badań psychologicznych, włączającej zastosowania matematyczne do pomiaru różnych aspektów psychicznych przy pomocy konkretnych narzędzi, które należało wynaleźć. W tym interdyscyplinarnym kontekście pojawia się po raz pierwszy pojęcie „psychometria” (Wolff, 1732). Ta przedwczesna wizja naukowego podejścia do pomiaru różnych funkcji psychicznych, jako ilościowego kontinuum lub skalowania, uległa szybko zapomnieniu i to do tego stopnia, że nawet we współczesnych podręcznikach niemal całkowicie ignoruje się ją i jej twórcę. Oddzielenie przez Wolffa psychologii empirycznej od racjonalnej, stworzyło podstawy koncepcyjne „przejścia” od pierwszych prób z zakresu psychofizyki, połączonej potem z antropometrią Galtona, do pierwszego projektu laboratorium psychologii eksperymentalnej w Lipsku i wreszcie do testu Bineta-Simona. Postęp w statystyce dał narzędzia umożliwiające nowe spojrzenie na organizację danych empirycznych. W konsekwencji, przyczynił się do eksploracji wyższych procesów umysłowych, inteligencji, zdolności, zainteresowań i osobowości, rozszerzając jednocześnie znaczenie psychometrii.

Ze względu na specyfikę przedmiotu swoich badań, psychologia nie uniknęła krytyki, co do naukowego charakteru pomiaru i jego względności w tym obszarze. Nawiązane jednak wydają się argumenty, że pomiar aspektów fizycznych, jak wzrost, waga czy temperatura, jest bardziej precyzyjny. Zależy on także od dokładności narzędzi, przedmiotu pomiaru i reguł operacyjnych samego procesu mierzenia jako takiego. Z drugiej strony nietrudno wykazać, że troska o dokładność pomiaru różnych aspektów fizycznych człowieka może prowadzić do przesadnych skrajności.

Pomiar w psychologii odnosi się do charakterystyk złożonych, abstrakcyjnych, trudno dostępnych, co wymaga podejścia pośredniego, potencjalnie narażonego na błędy. W związku z tym, sytuację pomiaru własności psychicznych charakteryzuje niepewność, która wymaga odwołania się do konkretnego modelu teoretycznego. Zobowiązuje to psychologa do szczegółowej definicji przedmiotu i stosownej adaptacji metod pomiarowych. Oczywiście model taki może być poddawany krytyce w różnym kontekście, co do wartości oceny osób i różnych tego konsekwencji (Urbina, 2004). Przegląd elementów strukturalnych modelu wykazuje szeroki zakres problematyki i różnych relacji z tym związanych. W prezentowanym opracowaniu, wychodząc od skal pomiaru przejdziemy do natury zmiennych psychologicznych, znaczenia danych, rzetelności i trafności.

Skale pomiaru

Stevens (1946) zaproponował cztery poziomy hierarchii zmiennych lub skal pomiaru, wiążące liczby z przedmiotami, na których dokonujemy operacji numerycznych. Nazwał je skalami: nominalną, porządkową, interwałową i proporcjonalną.

Każda z kolejnych skal posiada swoje specyficzne charakterystyki oraz zawiera właściwości skali poprzedzającej.

Na poziomie nominalnym cyfry nie reprezentują ilości, służą jedynie grupowaniu oznaczonych tym samym symbolem przedmiotów, które posiadają wspólne charakterystyki. Na poziomie tym nie mamy do czynienia z prawdziwą miarą, a tylko z klasyfikacją, co oznacza, że jedyną możliwą operacją matematyczną jest zliczanie. Na poziomie porządkowym, przedmioty zajmują miejsce według rang, zgodnie z którymi posiadają one w większym lub mniejszym stopniu uwzględnione charakterystyki. Chodzi tu o najbardziej podstawową formę pomiaru, któremu nie dajemy innego znaczenia niż stopniowanie określonej cechy. Poziom przedziałowy pozwala zgrupować dwa poprzednie typy skal, jednakże bez zera absolutnego. Mamy tu do czynienia z równymi jednostkami. Na przykład, różnica 2 cm między 5 i 7 cm posiada taką samą wartość, jak między 15 i 17 cm, jednakże temperatura 10°C nie oznacza dwa razy wyższej niż 5°C. W skali stosunkowej natomiast jednostki są zarazem równe i proporcjonalne, mając także punkt zerowy: odcinek 10 cm jest dwa razy dłuższy niż odcinek 5 cm.

Pomiar w psychologii jest możliwy na poziomie nominalnym lub porządkowym, ale praktycznie postępujemy tak, jakbyśmy mieli do czynienia z pomiarem na poziomie przedziałowym, co pozwala wykorzystać zespół statystyk opisowych i inferencji statystycznej. Gdybyśmy nie wykonali operacji zmiany poziomu skali, nie moglibyśmy obliczyć nawet średniej czy odchylenia standardowego z danych nominalnych czy porządkowych. Z matematycznego punktu widzenia ta zmiana poziomu skali ma efekty minimalne, natomiast interpretacja psychologiczna wyników byłaby ryzykowna, jeśli nie wzięlibyśmy pod uwagę rzeczywistego poziomu pomiaru. Na przykład, różnica 5 punktów w ilorazach inteligencji (I.I.) 100 i 105 nie jest porównywalna z taką samą różnicą między ilorazami 55 i 60. Nie jest również prawdą, że I.I. 100 reprezentuje inteligencję dwukrotnie wyższą niż I.I. 50.

Mimo tych ograniczeń, skale pomiaru jako całość służą wyższej formie strukturyzacji danych poprzez przypisanie cechom wartości liczbowych. Są dostatecznie precyzyjne, aby pozwolić na zróżnicowanie indywidualne w zakresie danej cechy. Są one użyteczne w określonych granicach, jeżeli nie usiłujemy powiedzieć więcej niż pozwalają na to cyfry.

Natura zmiennych psychologicznych

Przedmiot pomiaru powstaje w bezpośredniej relacji ze sposobem jego pomiaru w psychologii. W efekcie różne charakterystyki przedstawione do pomiaru, zwane zmiennymi lub cechami, determinują w znacznej mierze wykorzystanie skal pomiarowych już w punkcie wyjścia.

W kategoriach pomiaru, niektóre właściwości są stałe, a inne nie. Zmienne mogą mieć charakter bardziej lub mniej konkretny lub przybierać postać konstrukcji pojęciowych o różnym stopniu abstrakcji. Mogą być proste lub jednowymiarowe albo zawierać różne koncepcje wielowymiarowości. Większość właściwości psychologicznych ma charakter abstrakcyjny, zmienny i złożony.

Zmienność obserwowana w aspekcie różnych atrybutów psychologicznych może być jakościowa lub ilościowa. W przypadku różnic jakościowych można mówić głów-

nie o pomiarze na poziomie opisowym, ponieważ jego zakres ogranicza się do klasyfikacji i skali nominalnej. Zmienne jakościowe mają charakter stopniowalny i możemy je rangować. Biorąc pod uwagę naturę konstruktów psychologicznych, dostrzegamy tendencję do definiowania pojęć w sposób ogólny, a następnie do określenia operacji, które najlepiej mierzą zmienną zależną i powiązane z nią zmienne niezależne. Inaczej mówiąc, zmienne psychologiczne nabierają znaczenia w kontekście operacji wykorzystywanych do ich pomiaru. Nie mając zamiaru podważenia całej koncepcji badań psychologicznych, w przypadku niektórych z badań można kwestionować zasadność wyboru określonych statystyk i interpretacji wyników, a więc znaczenia nadane zmien- nym. Na podstawie analizy danych empirycznych wyciąga się bowiem niekiedy zbyt daleko idące wnioski.

Tradycyjnie psychometria prawie zawsze traktowała cechy lub zmienne ze ści- śle strukturalnego punktu widzenia. Zachowanie najczęściej wyjaśnia się w systemie łączącym elementy struktury przyczynowo. Takie stanowisko wydaje się być nieak- ceptowane z behawioralnego punktu widzenia, a także przez klinicystów o orientacji humanistycznej. Z jednej i drugiej strony odrzuca się całkowicie psychometrię, nie biorąc pod uwagę tego, że elementy modelu mogą mieć zastosowanie w szerokim zakresie. Silva (1993) uważa, że krytycy psychometrii prawie zawsze wykazują brak wizji całościowej, przywołują cytaty spoza kontekstu i nie rozumieją podstawowych koncepcji. Proponowane przez nich rozwiązania są zbliżone do odrzuconych, tyle że sformułowane w innej terminologii.

Nadawanie znaczenia danym psychologicznym poprzez normalizację

W przeciwieństwie do nauk ścisłych, których rezultaty zazwyczaj mogą być pod- dane bezpośredniej interpretacji, wyniki badań psychologicznych, jako złożone i po- średnie, powinny być interpretowane w odniesieniu do norm. W rzeczywistości wynik surowy nie ma żadnego znaczenia bez odniesienia do narzędzia pomiaru jako takiego (podejście kryterialne), czy porównania z wynikami innych osób wykonujących ten sam test (podejście normatywne). W podejściu kryterialnym wskazujemy na to, co dana osoba jest w stanie aktualnie wykonać w zakresie konkretnego testu, jak inter- pretować na przykład iloraz inteligencji jako funkcję dobrze rozwiązanych prób. Mo- żemy także opisać badanego w takich samych terminach jak innych, którzy uzyskali porównywalny wynik surowy. Nelson i Hays (1979) uważają te strategie za szcze- gólnie obiecujące w klinice, gdzie konieczna jest koncentracja na jednostce. Należy nadmienić, że nie wszystkie testy mogą być wykorzystane według tego systemu, stąd na etapie konstrukcji narzędzia należy przewidzieć taką ewentualność. W praktyce rzadko posługujemy się testami według zasad kryterialnych, a podejście normatywne jest tym, które łatwiej utożsamia się z psychometrią.

Innym źródłem kontrowersji było przeciwstawienie podejścia nomotetycznego, z którym wiąże się psychometria klasyczna, badania grupowe i normalizacyjne, po- dejściu idiograficznemu, kojarzonemu z psychologią kliniczną i psychoterapią, kon- centrującymi się na jednostce. W rzeczywistości ten podział wydaje się sztuczny, po- nieważ nie można realizować zadań klinicznych bez odniesienia do szeroko pojętej normy. Cronbach (1990) podkreśla, że chodzi o dwa sposoby uprawiania psychologii,

raczej dopełniające się niż przeciwstawne. Jeżeli nawet cele diagnostyki psychologicznej są idiograficzne, sama ona opiera się na danych otrzymanych z dyscyplin nomotetycznych.

Z normatywnego punktu widzenia zakłada się, że charakterystyki psychologiczne posiadają rozkład normalny. Założenie to, jako uogólnienie obserwacji cech fizycznych, wydawało się logiczne. Dość długo utrzymywano przekonanie, że chodzi o prawo naturalne. Obecnie wiemy, że rozkład zmiennej jest funkcją wykorzystanej skali pomiarowej. W istocie łatwo jest otrzymać rozkład normalny, dobierając próbę osób badanych w odpowiedni sposób. Za rozkładem normalnym przemawiał również fakt, że krzywa Gaussa posiada pewne właściwości matematyczne, które czynią z niej łatwe narzędzie do pomiaru zmiennych psychologicznych. Testy implikują podporządkowanie się zasadom normy. Wyniki surowe muszą być przekształcone w standardowe, żeby stać się interpretowalnymi. Normy w swojej podstawowej funkcji pozwalają wskazać pozycję danej osoby w relacji do zbioru stanowiącego rozkład normatywny, dając konkretny punkt wyjścia do dalszych interpretacji psychologicznych.

Rzetelność

Z koncepcyjnego punktu widzenia, rzetelność należy rozpatrywać na dwóch poziomach: osoby ocenianej i instrumentu wykorzystywanego do jej oceny. Odnośnie instrumentu wiemy, że czynniki takie jak długość testu, układ prób (łatwe/trudne), rodzaj pytań (pytania typu dopełnienia tak – nie są łatwiejsze do sfalszowania niż wielokrotnego wyboru), charakterystyka grupy (im bardziej homogeniczna zbiorowość, tym mniejsza rzetelność) wpływają na wartość współczynnika rzetelności. Chodzi tu o uwagi techniczne, których wyjaśnienie szczegółowe znajdziemy w tekstach specjalistycznych (zob. Bernier, Pietrulewicz, 2001).

Druga składowa rzetelności jest bardziej problematyczna i stanowi często przedmiot analiz krytycznych. Pojęcie rzetelności jest związane z różnicami indywidualnymi. Możemy badać tylko cechy względnie stałe, aby mierzyć różnice międzyosobowe. Nawet jeżeli uznajemy za normalne wahania wyników w pewnych granicach, wielkość zmienności zależy od natury badanego aspektu. Na przykład oczekujemy mniejszej zmienności w zakresie inteligencji czy zdolności, a więcej w osobowości. Przewidujemy więc z góry, że zmienne psychologiczne nie zachowują się w sposób stały i stąd nie oczekujemy, że rzetelność miary będzie doskonała, ale chcemy, by mieściła się w ramach uznanych granic. Podstawowy model psychometrii zakłada, że rezultat obserwowalny jest sumą wyniku otrzymanego i błędu pomiaru. Aktualna koncepcja rzetelności kładzie nacisk głównie na powtarzalność rezultatów w granicach błędu pomiaru (Aiken, Groth-Mamat, 2006). W efekcie, jeżeli mierzymy to, co zamierzeliśmy i jeżeli otrzymana charakterystyka posiada spójność logiczną i stałość, właściwości te powinny znaleźć się w pomiarze jako takim. Kiedy oceniamy rzetelność testu uwzględniamy naturę zmiennej, instrumentu w relacji do konkretnych kryteriów, porównując otrzymane współczynniki z wartościami charakteryzującymi podobne metody mierzące oczywiście podobną rzeczywistość psychologiczną.

Trafność

Kiedy mówimy o trafności, odnosimy to pojęcie do tego, co jest mierzone i w relacji do czego, odpowiadając tym samym na pytanie, na ile wyniki upoważniają do przewidywań (Kostrzewski, 1970; Brzeziński, 1984). Cechy psychologiczne (inteligencja, osobowość, zainteresowania) są konstrukcjami pojęciowymi i mogą być mierzone tylko pośrednio. Trzeba jednocześnie wykazać, że koncepcja teoretyczna jest właściwa i operacje wykorzystane do jej pomiaru są adekwatne. Test może być dobrze skonstruowany, dobrze wystandaryzowany i rzetelny, ale możemy nie wiedzieć, co mierzy on w rzeczywistości. Wielokrotnie wykazywano, że testy inteligencji są dobrymi predyktorami wyników szkolnych i wskaźnikami innych sukcesów. Można w tym przypadku postawić pytanie, czy mierzą one naprawdę inteligencję.

Często oceniamy trafność testu, odwołując się do związku pomiędzy jego rezultatami a jakimś kryterium zewnętrznym, za pomocą współczynnika korelacji – w przypadku testów inteligencji najczęściej są to wspomniane wyżej sukcesy szkolne. Weźmy pod uwagę typowy przykład, w którym współczynniki trafności wynoszą od 0,6 do 0,7. Jeżeli podniesiemy te wartości do kwadratu, możemy łatwo dojść do wniosku, że między 36% i 49% wariacji rezultatów szkolnych może być wyjaśniona poprzez odniesienie ich do wyników testów inteligencji. Z kolei, aby odpowiedzieć na pytanie, czy trafność 0,7 jest wysoka, musimy wziąć pod uwagę naturę wykorzystanego w tym przypadku kryterium. Logika, na której opiera się trafność przewidywania zakłada, że wyniki szkolne są funkcją inteligencji, a jeżeli test mierzy naprawdę inteligencję, musi istnieć istotna relacja między testem i kryterium zewnętrznym. Niewłaściwe jednak byłoby przekonanie, że rezultaty szkolne zależą wyłącznie od zdolności intelektualnych. Inne czynniki, takie jak motywacja, poziom nauczania, stymulacja środowiskowa, będą także miały wpływ na wyniki szkolne. Nie oczekujemy więc doskonałej zależności między testem i kryterium zewnętrznym. Należy określić czynniki wpływające na wyniki szkolne i ich istotność w relacji do inteligencji.

Współczynniki trafności są zazwyczaj wyższe w testach inteligencji i zdolności niż w testach osobowości. W żadnym przypadku nie upoważnia to jednak do twierdzenia, że jedne są lepsze od drugich. W przypadku testów osobowości jest możliwe, że proporcjonalnie większa ilość czynników ma wpływ na kryteria wybrane do walidacji. Należy także zaznaczyć, że współczynniki trafności wyrażają tylko stopień relacji linearnej między dwiema zmiennymi. Jest więc możliwe, że model liniowy bardziej odpowiada rzeczywistości psychologicznej testów inteligencji niż osobowości. Obecnie nie uznaje się już ogólnego znaczenia trafności, ponieważ zazwyczaj sprawdzamy wartość interpretacji danych otrzymanych w konkretnej sytuacji, na konkretnej grupie i przy pomocy konkretnego instrumentu.

Trafność testu jest jednym ze składników wartości diagnostyki psychologicznej i musimy zawsze rozważać ją w kontekście innych elementów, gwarantujących precyzję pomiaru.

Wnioski

Próby wyeliminowania psychometrii, nawet w warunkach ściśle klinicznych, nie doprowadziły nigdy do satysfakcjonujących rozwiązań. Powiązanie ścisłego podejścia ilościowego z otwarciem na włączenie informacji jakościowych wydaje się najbardziej efektywnym w działaniu praktycznym. Taka postawa nie wyklucza czystego modelu ilościowego, sugerowanego przez twórcę psychometrii Wolffa, jako psychologicznych zastosowań matematyki, a ograniczenia mogą wy pływać tylko z charakteru zmiennych.

Test psychometryczny nie jest niczym więcej niż narzędziem zbierania danych, które odpowiada zespołowi reguł, często pierwotnie zapożyczonych z nauk ścisłych w formie zmodyfikowanej i zaadaptowanej do nauk humanistycznych. Cykl modyfikacji dokonał się w trakcie dyskusji pomiędzy teoretykami a praktykami z wyłączeniem opinii podmiotów badanych. Niektóre elementy teorii testów są dyskusyjne, ale model jako całość jest logiczny i dobrze dostosowany do przedmiotu swoich badań. Oczywiście model nie implikuje idealnej zgodności z rzeczywistością. Znajomość jego silnych i słabych punktów i ograniczeń daje możliwość poprawnej interpretacji informacji, których nam dostarcza w aspekcie różnych zmiennych psychologicznych.

Psychometria daje psychologii zalety nauki ścisłej, biorąc pod uwagę jej metody i odwoływanie się do ujęcia ilościowego. Próby eliminowania jej w przeszłości nie proponowały nic w zamian. Dodatkowo, wprowadzenie informatyki pomaga w uproszczeniu skomplikowanych procedur technicznych obliczania wyników badań, redukuje prawdopodobieństwo błędów i usprawnia prace psychologa w tym zakresie.

Literatura cytowana

- Aiken, L. R., Groth-Mamat, G. (2006). *Psychological testing and assessment*. Boston, MA: Allyn and Bacon.
- Bernier, J. J., Pietrullewicz, B. (2001). *La psychométrie traité de mesure appliquée*. Montréal-Paris-Casablanca: Gaetan Morin Éditeur.
- Brzeziński, Z. (red.) (1984). *Wybrane zagadnienia z psychometrii i diagnostyki psychologicznej*. Poznań: Uniwersytet im. Adama Mickiewicza.
- Cronbach, L. J. (1990). *Essentials of psychological testing*. Cambridge, MA: Harper & Row.
- Kostrzewski, J. (1970). *Wprowadzenie do Skali Inteligencji D. Weschlera dla dzieci od 5 do 15 lat*. Warszawa-Łódź: Polskie Towarzystwo Higieny Psychicznej.
- Nelson, R. O., Hays, S. C. (1979). The nature of behavioral assessment: A commentary. *Journal of Applied Behavioral Analysis*, 12, 491-500.
- Silva, F. (1993). *Psychometric foundations of behavioural assessment*. Newbury Park, CA: Sage Publications.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, N. Y.: John Wiley & Sons.

Wainer, H. (1987). *The first four millennia of mental testing: From ancient China to the computer age. Research Report*. Princeton, NJ: Educational Testing Service.

Wolff, C. (1732). *Psychologia Empirica, Methodo Scientifica Pertractata Qua Ea Quae De Anima Humana Indubia Experientiae Fide Constant, Continentur Et At Solidam Universae Philosophiae Practicae Ac Theologiae Naturalis Tractationem Via Sternitur*. Francofurti et Lipsiae: Prostat In Officina Libraria Rengeriana.