

WSTĘP

Opis osoby bez odwołania się do innych może być fragmentem literatury, biografii lub powieści. Ale nauki? Nie.

Meyer, 1926, s. 271¹

Tematyka podejmowana w tej pracy jest odzwierciedleniem trwającej obecnie na świecie w środowisku psychometrów i psychologów dyskusji na temat jakości psychologii jako nauki (w 2018 roku powołano dwie organizacje zrzeszające dyskutujących o tym psychologów i psychometrów: Psychological Science Accelerator, Society for the Improvements of Psychological Science). Problem naukowości psychologii nie został rozwiązany przez stulecie jej istnienia, a ostatnio przybrał na sile w związku z tzw. kryzysem replikowalności psychologicznych badań naukowych (Turner, 2019). Spadek zaufania do wyników badań nie tylko wśród społeczeństwa, ale także wśród samych naukowców sprawił, że pojawiła się potrzeba uporządkowania problematyki prowadzenia badań i oceny ich jakości. W tej sytuacji przybierają na sile głosy krytyczne, które obecne są na każdym etapie rozwoju nauki, ale często były zagłuszane kolejnymi 'nowinkami' inkorporowanymi na grunt psychologii z ościennych dziedzin: socjologii, fizyki, matematyki czy ostatnio informatyki oraz neuronauki. Zachwyty nad nowymi metodami prowadzenia badań czy też analizą ich wyników sprawił, że psychologia 'szła do przodu, nie oglądając się za siebie'. Niestety problemy, które nie zostały dostatecznie zbadane i rozwiązane, wpływają teraz na spadek zaufania do psychologii jako dziedziny nauki.

Psychologia zawsze przejawiała aspiracje do bycia nauką opartą na badaniach empirycznych. Przedmiotem nauki empirycznej są obiekty o mierzalnych właściwościach. Dlatego podstawowym zmartwieniem psychologów badaczy jest udowodnienie, że zarówno mierzone obiekty psychiczne, jak i narzędzia pomiarowe używane w psychologicznej praktyce badawczej są godne zaufania. Problematykę empiryczności pomiaru psychologicznego można zatem sprowadzić

¹ „A description of one individual without reference to others may be a piece of literature, a biography or novel. But science? No”

do odpowiedzi na dwa pytania: 1) co specyficznego mierzy pomiar psychologiczny oraz 2) jak dobrze ten pomiar przebiega.

Podstawowym problemem dla psychologów odpowiadających na pytanie o przedmiot swoich zainteresowań badawczych jest udowodnienie, że obiekty psychiczne są mierzalne w jakikolwiek sposób. W latach trzydziestych ubiegłego wieku brytyjskie stowarzyszenie na rzecz rozwoju nauki (British Association for the Advancement of Science) powołało do życia Komitet pod przewodnictwem Allana Fergusona, by zbadał, jakie psychiczne atrybuty mogą być mierzone w sposób naukowy. Członkowie komitetu pod dużym wpływem Normana Roberta Campbella (zob. Michell, 1999, s. 144) stwierdzili, że ponieważ psychologiczne atrybuty nie poddają się konkatencji (nie można ich ilościowo porządkować), nie są mierzalne naukowo (Ferguson et al., 1940, za: Newman, 1974, s. 141). Niezgoda z tym oświadczeniem była głównym źródłem powstania operacyjnej teorii pomiaru stworzonej przez Stanleya S. Stevensa (1946)², w której zdefiniował on pomiar psychologiczny jako przypisywanie liczb do obserwacji zgodnie z różnymi funkcjami matematycznymi w zależności od charakteru mierzonej (wtedy jeszcze określanej jako *sensory events*) właściwości. Pozwoliło to wyróżnić cztery podstawowe skale pomiarowe, których definicja obowiązuje do dzisiaj i konstytuuje prawie wszystkie pomiary psychologiczne. Skoro zaś właściwości psychiczne okazały się mierzalne, to stało się możliwe zdefiniowanie, czym jest pomiar w psychologii. Wbrew obiegowym opiniom nie jest to jednak etap zakończony i definicja pomiaru w psychologii rozwijana jest nadal. Temu rozwojowi poświęcam rozdział 1 niniejszej pracy.

Do pomiaru potrzebne są mierzone podmioty. Z tej perspektywy przyglądam się pomiarowi w rozdziale 2. Poświęcony jest on problemowi pozyskiwania wyników badań psychologicznych głównie za pomocą metod samoopisowych. W dużym stopniu ufamy informacjom dostarczanym nam bezpośrednio przez osoby badane, które zanim informacji udzielią, konstruują ją w swoich umysłach (Kleka, 2017). Zastanawiam się, czy systematyczny błąd, jaki mogą popełniać badani przez nas ludzie, wystarczy uśredniać przez badanie grupowe i odwoływanie się do centralnego twierdzenia granicznego, czy też powinniśmy zacząć bardziej uwzględniać w naszych narzędziach pomiarowych to, że ludzie różnią się poziomem wiedzy o sobie i mają zróżnicowane ograniczenia w dostępie do tej wiedzy. Głębokość wglądu, poziom samowiedzy czy zdolność do samoopisu – to w moim odczuciu synonimy czynnika rzadko uwzględnianego w badaniach psychologicznych. W rozdziale 2 próbuję dokonać rozstrzygnięcia, czy ten immanentny czynnik badań psychologicznych jest nierozróżnialny od szumu tła (czynników losowych), czy też opierając się na przesłankach

² Zapewne nie bez znaczenia był też fakt, że komitet jako przykład potwierdzający tezę o niemierzalności atrybutów psychicznych podał skalę głośności opracowaną właśnie przez Stevensa.

teoretycznych i/lub empirycznych, jesteśmy w stanie go wyodrębnić i powinniśmy go kontrolować.

Dyskusja wokół odpowiedzi na drugie pytanie, czyli o jakość pomiaru, koncentruje się wokół jego rzetelności (precyzji) oraz trafności. Pojęcie rzetelności i towarzyszy problemowi pomiaru od zarania psychologii. Już Charles Spearman w swoich pracach na początku XX wieku (1904) zwrócił uwagę na obecność błędu pomiarowego i konieczność korygowania wyników badań psychologicznych o tę nieznaną wartość. Koncepcja rozwijana przez lata, ze znaczącymi pracami Williama Browna (1910), Joya P. Guilforda (1936, 1954), Frederica Kudera i Marion W. Richardson (1937), Philipa J. Rulona (1939), Cirila Hoyta (1941), Louisa Guttmana (1945), Harolda Guliksena (1950), Lee Cronbacha (1951), Frederica M. Lorda i Melvina R. Novicka (1968), Josa M.F. Ten-Berge i Fritsa E. Zegersa (1978), Williama Revelle'a (1979) oraz Rodericka P. McDonalda (1999), przeprowadziła rzetelność narzędzi badawczych w psychologii od alfy³ do omegi⁴ (szczegółowe porównanie współczynników rzetelności jako miary spójności testu por.: Zinbarg, Revelle, Yovel i Li, 2005; Dunn, Baguley i Brunnsden, 2014). Klasyczna teoria testu przybrała postać ugruntowanego modelu wyniku prawdziwego składającego się wraz z błędem pomiarowym na wynik obserwowany, rozszerzona o kontekst badania ma swoje uogólnienie w teorii uniwersalizacji (Cronbach, Gleser, Nanda i Rajaratnam, 1972). W obu teoriach rzetelność definiowana jest jako stosunek wariancji wyników prawdziwych do wyników obserwowanych.

Inaczej jest jednakże z rozwojem koncepcji trafności narzędzi psychologicznych. Adekwatność w opisie rzeczywistości psychicznej i użyteczność wyników uzyskiwanych w badaniach psychologicznych opiera się w całości na ich trafności, sprawiając, że pojęcie to stało się fundamentem, na którym budowana jest psychologia jako nauka. Pojęcie trafności od odpowiedzi na proste pytanie: „Czy test mierzy to, do czego został stworzony?” wyewoluowało do skomplikowanego tworu⁵ obejmującego trzy, pięć lub według niektórych autorów sześć różnych aspektów testu i testowania⁶. Zagadnieniu, jak psychologia przy definiowaniu

³ Współczynnik rzetelności określający spójność wewnętrzną pozycji testowych, dominujący w badaniach psychologicznych od publikacji Cronbacha w 1951 roku, krytykowany przez różnych badaczy, także przez swego twórcę (por. Cronbach i Shavelson, 2004), głównie za bezrefleksyjne stosowanie przez większość badaczy i omijanie założeń stojących za tym współczynnikiem, np. tau-ekwiwalentności pozycji. Współczynnik alfa nie był pierwszym zaproponowanym, ale pierwszym, który został utożsamiony z rzetelnością.

⁴ Współczynnik rzetelności wewnętrznej testów oparty na dekompozycji wariancji w modelu hierarchicznym analizy czynnikowej – pozbawiony ograniczeń współczynnika alfa, ale wymagający bardziej skomplikowanych obliczeń matematycznych.

⁵ Specjalnie unikam sformułowania „konstrukt”, ponieważ będzie mi ono potrzebne w rozważaniach nad przedmiotem pomiaru w rozdziale 1.

⁶ Wyczerpujące zestawienie 151 różnych nazw/rodzajów trafności przedstawili w swojej pracy Newton i Shaw (2014, s. 8).

trafności przeszła od pytania o odwzorowanie treści do traktowania trafności jako właściwości interpretacji wyniku testowego (por. Messick, 1980), przyglądam się bliżej w rozdziale 3.

Rozdział 4 poświęcony jest praktycznemu sprawdzeniu odporności stosowanych powszechnie sposobów walidacji narzędzi badawczych na treść, a właściwie na jej brak w narzędziach będących przedmiotem sprawdzania. Uzyskane wyniki wskazują na raczej niski stopień zaufania do metod statystycznych odebranych od założeń teoretycznych. Przy dużej dostępności różnych, skomplikowanych metod analitycznych poprawność otrzymanych wyników nie ma żadnego związku lub ma związek niewielki z założeniem o trafności walidowanego narzędzia badawczego – algorytmy matematyczne nie odróżniają automatycznie testów wiarygodnych od niewiarygodnych, ale jedynie dostarczają wyników. A te poddawane są najczęściej subiektywnej interpretacji.

Zanim przejdę do omawiania poszczególnych tematów dotyczących pomiaru i problemów z nim związanych, chciałbym dla jasności wywodu poczynić kilka założeń definicyjnych. I tak terminu *t e s t* używam w znaczeniu narzędzia badawczego stosowanego do pośredniego mierzenia właściwości psychicznych człowieka. W tej szerokiej definicji mieszczą się kwestionariusze, skale i testy zadań, które są prezentowane osobom uczestniczącym w badaniach psychologicznych w celu gromadzenia danych odzwierciedlających obecność (a często również stopień) danej zdolności psychicznej, postawy lub zachowania. Zastrzegam sobie terminy „podtest”, „podskala” i „skala” dla oznaczenia podzbioru poszczególnych elementów testu. Zazwyczaj uważa się, że odzwierciedlają one aspekt danego atrybutu wyższego rzędu.

Psychologowie badają nie tylko jeden z najbardziej złożonych systemów naturalnych – człowieka, ale także procesy zachodzące w psychice, których jeszcze nikt nie zdążył z powodzeniem skonceptualizować. Co więcej, robią to, używając owych procesów. Fakty te powinny przynajmniej nasuwać pytanie, do jakiego stopnia ilościowe, eksperymentalne metody nauk przyrodniczych (fizyki, biologii itp.) mogą być sensownie stosowane w psychologii. Jak bardzo obiekt badania, jakim jest człowiek z perspektywy psychologicznej, sprawia, że samo badanie staje się specyficzne i wymaga specjalnych metod pomiaru? Czy opracowane do tej pory i stosowane powszechnie metody pomiaru są wystarczająco dobre? Czy potrafimy odróżnić wiarygodne metody pomiaru psychologicznego od niewiarygodnych? Postaram się odpowiedzieć na te pytania na kartach niniejszej książki.