

Mieczysława Kreutza
Zmienność rezultatów testów
– czytana raz jeszcze

Jerzy Marian Brzeziński*

*Szkoła Wyższa Psychologii Społecznej
Wydział Zamiejscowy we Wrocławiu*

A RE-READING OF MIECZYŚLAW KREUTZ'S
VARIABILITY OF TEST SCORES

Abstract. The article focuses on the monograph by Mieczysław Kreutz (1893-1971), one of the most eminent Polish psychologists, whose scientific activity spans the period from the inter-war years to the first two post-war decades. After several decades, the monograph is read anew. The author of the article concentrates on the analysis of Kreutz's original conception of factors influencing test survey, on the concept of "maximum score," and on recommendations concerning, as Kreutz phrased it, the necessary modification of the test theory. Moreover, the "echoes" of Kreutz's conception in the psychometric theories of Lord and Novick (1968) and Cronbach et al. (1972), though distant in time, are pointed out in the paper. In the author's view, it is worth remembering and returning to Mieczysław Kreutz's conception.

Przypomnieniu poglądów naukowych Mieczysława Kreutza, profesora trzech polskich uniwersytetów: Jana Kazimierza we Lwowie, Wrocławskiego i Warszawskiego, psychologa, twórcy oryginalnych koncepcji psychologicznych, poświęcona była, zorganizowana przez Wydział Zamiejscowy SWPS we Wrocławiu (7 listopada 2012 r.), konferencja naukowa. Niniejszy artykuł stanowi przeredagowaną wersję wygłoszonego wówczas referatu pt. *Mieczysława Kreutza „Zmienność rezultatów testów” (cz. I – 1927; cz. II – 1933) – odczytana na nowo.*

* Adres do korespondencji: Jerzy Marian Brzeziński, Szkoła Wyższa Psychologii Społecznej, Wydział Zamiejscowy we Wrocławiu, ul. Ostrowskiego 30, 53-238 Wrocław; e-mail: brzezuam@amu.edu.pl

Aby jednak w pełni zrozumieć to, co nowego chciał w swojej pracy odnoszącej się do kształtującej się nowej praktyki badawczej, odwołującej się do testów psychologicznych, przekazać psychologom M. Kreutz (1927, 1933), trzeba wpierw zapoznać się z konturami „mapy” ukazującej stan wiedzy psychometrycznej i praktyki związanej z konstruowaniem i stosowaniem testów psychologicznych. Na tym tle można będzie zrozumieć *novum* propozycji M. Kreutza.

Dopiero po dokonaniu tych rekonstrukcji możliwe będzie: (a) wskazanie tych poglądów, które wówczas nie znalazły zrozumienia wśród polskich psychologów i nie mogły go znaleźć poza granicami Polski (z powodów językowych – Kreutz nie publikował swoich prac w języku angielskim), oraz (b) wskazanie tego, co z myśli Kreutza można, jak mi nie mam, odnaleźć w późniejszych koncepcjach psychometrycznych, które pojawiły się w znanych w świecie pracach wybitnych psychologów pracujących na polu psychometrii; mówiąc inaczej: czy można się doszukać światowego oddźwięku myśli Kreutza.

DWA NURTY ROZWOJU PSYCHOLOGII – PRZED ROKIEM 1939

Za Lee J. Cronbachem (1957/2007) zwykło się wyróżniać dwa nurty (modele, podejścia) badawcze w psychologii: eksperymentalny i korelacyjny.

Jeśli chodzi o początki dwudziestowiecznej psychologii naukowej rozwijającej się w pierwszym nurcie, to najbardziej wyrazisty program metodologiczny, wprost wzorowany na praktyce badawczej nauk przyrodniczych i fizyce¹, przedstawili behawioryści (John B. Watson, Edwin G. Boring, Stanley S. Stevens, Burrhus F. Skinner, Edward Ch. Tolman). Te „nowinki” metodologiczne zostały podchwycone przez polskich badaczy. I tak, w niewielkim odstępie czasowym, powstały polskie pracownie psychologii eksperymentalnej. Zakładali je na początku XX wieku: Władysław Heinrich na Uniwersytecie Jagiellońskim, Kazimierz Twardowski na Uniwersytecie Jana Kazimierza we Lwowie oraz Edward Abramowski na Uniwersytecie Warszawskim.

Równie długą tradycję ma drugi nurt badań naukowych w psychologii – korelacyjny. Związany był z badaniami różnic indywidualnych (tu jako „ojca” można wskazać profesora Uniwersytetów we Wrocławiu i Hamburgu, Wil-

¹ Inspirujący psychologów eksperymentatorów program metodologiczny był przedstawiony w stanowisku operacjonizmu, wyłożonemu w sztandarowej pracy fizyka Percy W. Bridgmana (1927 – por. jego pogląd na definiowanie pojęć: „the concept is synonymous with a corresponding set of operations”). Jeden ze znamienitych zwolenników tego podejścia, Stanley S. Stevens (1935/2000), tak pisał: „[...] Doktryna operacjonizmu uznaje wyrażenie fakt, że pojęcie, czy zdanie, ma znaczenie empiryczne tylko wtedy, gdy może być obiektem kreślonych, konkretnych operacji, możliwych do wykonania przez zwykłych ludzi. Jest oczywiste, że trzeba podjąć z całą powagą przegląd – z perspektywy operacjonizmu – nagromadzonych przez psychologię pojęć, by uwolnić się od pojęć mglistych i nieokreślonych, które prowadzą do niekończących się dyskusji i sporów [...]” (s. 107).

liama Sterna, który jako pierwszy, w 1911 r., użył terminu: *Differential Psychology*). Jeśli chodzi o znaczące nazwiska ze światowej psychologii, to ograniczę się do wskazania kilku wybitnych postaci, których osiągnięcia wywarły wpływ na program badań naukowych prowadzonych w tym nurcie: Francis Galton, James McKeen Cattell, Karl Pearson, Charles Spearman, Alfred Binet, Lewis M. Terman, Luis L. Thurstone.

W tym dynamicznie rozwijającym się nurcie badań (i naukowych, i aplikacyjnych) szczególne znaczenie miały testy psychologiczne. Niektórzy psychologowie do tego stopnia zaczęli się z nimi identyfikować, że definiowali terminy psychologiczne w duchu prymitywnego operacjonizmu – jako wyniki operacji pomiarowych przeprowadzanych za pomocą testów psychologicznych.

W dynamicznie rozwijającej się wówczas metodzie testów da się, jak mnie mam, wyróżnić trzy fazy. Zatem krótko o nich.

TRZY FAZY ROZWOJU METODY TESTÓW DO 1939 ROKU

Można wyróżnić trzy fazy, rzecz jasna, przyjmując rok 1939 za rok graniczny w rozwoju metody testów psychologicznych. Graniczny, gdyż wybuch II wojny światowej nie tylko zmienił mapę polityczną świata, ale w tej części Europy wprowadził, po jej zakończeniu, dominację ideologii komunistycznej z wszechobecną cenzurą. Ofiarą tej cenzury padła też metoda testów. Polska psychologia cofnęła się w swoim rozwoju, a tacy badacze, jak Kreutz byli, właśnie za swoje zainteresowania naukowe (Kreutz: introspekcja i metoda testów), napiętnowani (niestety także przez swoich kolegów-psychologów). Dopiero po 1956 roku można było wrócić na polskich uczelniach do badań (i kształcenia) w zakresie psychometrii. Pierwszym, po latach (niestety straconych i trudnych do odrobienia), był M. Choynowski (1959) ze swoim, dziś już klasycznym, artykułem: „Elementy teorii testów psychologicznych”.

Faza I. *Od Galtona do (wczesnego) Bineta – wnioskowanie z prostych reakcji psychomotorycznych człowieka o właściwościach jego złożonych struktur poznawczych.* Trzeba od razu powiedzieć, iż podejmowane przez wymienionych badaczy wysiłki nie zostały zwieńczone sukcesem. Ta droga wyprowadziła ich na naukowe manowce. Nie dało się też, co oczywiste, uzyskać przełożenia na działania praktyczne. Owe działania praktyczne związane były z poszukiwaniami „psychometrycznych” predyktorów sukcesów szkolnych uczniów. Największe zaś nadzieje wiązano z czasem reakcji. Francis Galton zamierzał wnioskować z prostych reakcji psychomotorycznych człowieka, na podstawie wyników *testów różnicowania zmysłowego*, o właściwościach jego złożonych struktur poznawczych. Zakończyły się one, jak wiadomo, niepowodzeniem. Nie znaleziono bowiem korelacji wyników tych prób z ocenami szkolnymi. Podobne niepowodzenie spotkało Jamesa McKeen Cattella, który posługiwał się takimi próbami psychometrycznymi, jak: *Sila mięśni, Szybkość ruchów, Wrażliwość na ból, Ostrość wzroku i słuchu, Różnicowanie ciężaru, Czas reakcji, Pamięć.* Gdyby wymienić najbardziej znaczących badaczy tego

nurtu, to niewątpliwie należy wskazać jeszcze na następujących badaczy: Emila Kraepelina, Hermana Ebbinghausa czy Alfreda Bineta (z wczesnego okresu jego twórczości).

Faza II. Od Alfreda Bineta i Williama Sterna do Lewisa M. Termana – pierwsze testy inteligencji. Istotny przełom, jeśli chodzi o testowy pomiar uzdolnień i inteligencji, dokonał się w 1905 r., gdy psycholog Alfred Binet i lekarz Theodore Simon skonstruowali pierwszą Skalę Inteligencji. Według Bineta „inteligentne myślenie” obejmuje trzy wymiary: wymiar ukierunkowania, wymiar adaptacji oraz wymiar krytycyzmu. Mówiąc inaczej, inteligentne myślenie polega na tym, że: dana osoba wie, „co” i „jak” trzeba zrobić, umie wybraną strategię działania dostosować do wykonywanego zadania, umie kontrolować przebieg realizacji zadania oraz adekwatnie oceniać własną aktywność. To był znaczący przełom w podejściu do pomiaru (za pomocą testów) takiego konstruktu, jak „inteligencja”. Alfred Binet i Theodore Simon w kolejnych latach (1908 i 1911) skonstruowali kolejne, ulepszone wersje swojej Skali Inteligencji. Od 1908 r. zaczęto posługiwać się terminem „wiek umysłowy” dla określenia poziomu rozwojowego badanej osoby, zaś William Stern w 1912 r. jako pierwszy wprowadził termin „iloraz inteligencji” (niem. *Intelligenzquotient*), określając go jako stosunek „wieku umysłowego” do „wieku chronologicznego” (w polskich przekładach posługiwano się też terminem „wiek życia”). Lewis M. Terman nadał mu natomiast (przez przemnożenie ułamka przez 100) ostateczny kształt. Zatem zrównoważony stosunek „wieku umysłowego” do „wieku chronologicznego” wyrażał się wynikiem $IQ = 100$. Odtąd psychologowie, i nie tylko oni, zaczęli posługiwać się (i to nie zawsze trafnie) skrótem IQ, który zrobił chyba największą karierę w świecie – nie tylko psychologicznym. To Henry H. Goddard, jako pierwszy, a po nim Terman, „przewieźli” Skalę Bineta i Simona przez Atlantyk. Terman ogłosił ją, w 1916 r. na Uniwersytecie Stanforda, jako Stanford-Binet Intelligence Scale (tzw. S-B 1), a przez kilkadziesiąt lat prowadził nad tą Skalą Inteligencji systematyczne badania. Z jego nazwiskiem świat zaczął wiązać testowe badanie inteligencji i IQ. Zauważmy jeszcze, że w 2003 r. ukazała się jej najnowsza, piąta wersja: S-B 5.

Faza III. Od Davida Wechslera – zmiana podejścia do konstrukcji IQ. Luis Leon Thurstone (1926) był chyba tym pierwszym psychologiem, który zwrócił krytyczną uwagę na termin „wiek umysłowy”, jakim posługiwano się w konstrukcji IQ. Konstrukcja wyniku testowego oparta była na jednym, bardzo istotnym założeniu o liniowej zależności „wieku umysłowego” od „wieku chronologicznego”. Jednakże daje się ono utrzymać jedynie dla przedziału „wieku chronologicznego” 4-16 lat. Aby jakoś rozwiązać ten kłopotliwy problem, Terman zalecał, by w przypadku osób, których wiek przekroczył 16 lat, wprowadzić do mianownika wzoru na IQ *wiek życia* = 16. Wnikliwa analiza powiązań *wieku umysłowego* i *wieku chronologicznego* doprowadziła Thurstone’a do zaniechania obliczania „wieku umysłowego” z danych surowych testu na rzecz wyniku standaryzowanego z , obliczanego na podstawie rzeczywistej wartości średniej i odchylenia standardowego rozkładu wyników surowych testu w reprezentatywnej grupie wiekowej, do której można przypi-

sać daną osobę badaną. To była istotnie nowa propozycja metodologiczna w podejściu do obliczania wyników testów inteligencji.

Tym tropem poszedł David Wechsler, psycholog pracujący (w latach 1932-1967) w Bellevue Psychiatric Hospital w Nowym Jorku, który zaproponował nową konstrukcję IQ, opartą na wyniku standaryzowanym. W zaprezentowanej w 1939 r. przez niego nowej Skali Inteligencji (w jakimś sensie „konkurencyjnej” względem S-B 2), pod nazwą: Wechsler-Bellevue Intelligence Scale, IQ obliczano nie tak, jak to się robiło dotychczas, ale poprzez przyrównanie wyniku danej osoby do średniej jej grupy wiekowej. Od tego momentu konstrukcja tych dwóch dużych Skali Inteligencji różniła się co do kształtu wyniku IQ. O ile Terman (i jego następcy) pozostał wierny klasycznej formule IQ, o tyle Wechsler (i jego następcy) w kolejnych wersjach ich Skali (z 1955 – WAIS, 1981 – WAIS-R, 1997 – WAIS-III, 2008 – WAIS-IV) konsekwentnie posługiwał się swoją, autorską formułą IQ. Jednakże już w wersji S-B 3 (z 1960 i 1972 r.) i w następnych (1986 – S-B 4 oraz 2003 – S-B 5) zmieniono klasyczną formułę IQ na formułę „wechslerowską”. Zatem dziś IQ – niezależnie od tego, czy został on pozyskany za pomocą Skali Inteligencji typu termanowskiego czy typu wechslerowskiego – obliczane jest w ten sam, właśnie „wechslerowski” sposób.

W interesującym nas okresie czasu nie tylko nastąpił znaczący rozwój statystyki (zauważmy, że bez jej narzędzi nie byłby możliwy rozwój metody testów psychologicznych). Psycholodzy mogli sięgnąć po stworzone przez statystyków współczynniki korelacji: r – Karla Pearsona (w 1884 r.) ρ – Charlesa Spearmana (w 1904 r.), ϕ – George’a U. Yule’a (w 1912 r.), testy istotności różnic: test t – Williama S. Gosseta („Studenta”, w 1908 r.), test χ^2 – Karla Pearsona, a także bardziej zaawansowane metody statystyczne: ANOVA – Ronald’a A. Fishera (w 1921 r. i w latach późniejszych), MANOVA (w latach trzydziestych XX w.), analizę czynnikową (modele zaproponowane przez Charlesa Spearmana w 1904 i 1927 r. i przez Luisa L. Thurstone’a w 1931 r.), modele wielokrotnej korelacji i regresji liniowej (lata trzydzieste XX w.).

Jeśli zaś chodzi o dokonania na polu psychometrii, to chciałbym zwrócić uwagę na wspomniane już modele analizy czynnikowej, prace z zakresu teorii rzetelności i trafności (zwłaszcza na tak do dziś popularne wzory 20. i 21. na obliczanie rzetelności testów, opracowane przez Frederica G. Kuder’a i Marion’a W. Richardson’a w 1937 r., a później rozwinięte przez Lee’a J. Cronbach’a (wzór K-R 20 znany jest nam dziś przede wszystkim jako „alfa-Cronbacha” z 1951 r.). Douglas N. Jackson w 1939, a Cyril Hoyt w 1941 r. zaproponowali użycie analizy wariancji (ANOVA) do badania rzetelności testu psychologicznego. Prace nad teorią testów psychologicznych zostały wstępnie podsumowane przez J. P. Guilford’a (1936, 1954). Całościowe ujęcie tej teorii, znane jako klasyczny model teorii testów czy też jako „teoria wyniku prawdziwego i błędu” (*theory of true and error scores*), zostało wyłożone w podstawowym dziele Harold’a Gulliksen’a (1950). Po nim opracowano kolejne modele psychometryczne. Najbliższe to autorskie ujęcie modelu klasycznego przez Frederica Lord’a i Melvin’a R. Novick’a (1968) oraz to późniejsze: L. Cronbach’a i współautorów (1972).

Na tak naszkicowanym tle psychologiczno-statystyczno-psychometrycznym spróbuję, w następnym punkcie, „umieścić” pracę Kreutza i wskazać na jej współczesne implikacje dla posługiwania się w praktyce społecznej metodą testów. I jeszcze jedno, gdyby Kreutz opublikował swoją pracę w jakimś uznanym wydawnictwie anglojęzycznym, to zapewne dziś moglibyśmy zapoznać się z pracami, których autorzy podjęli dialog z koncepcją Kreutza.

KREUTZA POGLĄDY NA WARTOŚĆ METODY TESTÓW

Zacznijmy od przytoczenia paru deklaracji teoretyczno-metodologicznych, które można znaleźć na kartach pracy Kreutza. Kreutz nie uległ wspomnianej wyżej modzie na prymitywny operacjonizm i – w szczególności – nie akceptował takich pseudonaukowych definicji pojęć, zgodnie z którymi można było definiować jakieś pojęcie poprzez wynik w teście psychologicznym „mierzącym” to pojęcie (np. inteligencja to jest to, co mierzy test inteligencji!). W pierwszej części pracy, poświęconej problematyce zmienności wyników testów psychologicznych, złożył mocno brzmiącą deklarację (Kreutz, I, s. 11)²:

Konieczność stworzenia teorii testów nie ulega wątpliwości; zadaniem zaś tej teorii byłoby takie przekształcenie testów, by umożliwiały one uzyskanie jak największej wiarygodności i ścisłości wyników, dorównującej o ile możliwości ścisłości badań w zakresie nauk przyrodniczych. Jedynie takie przekształcenie pozwoli na utrzymanie testów wśród metod naukowych psychologii teoretycznej; zarazem też psychologia stosowana, której wymagania co do ścisłości są mniejsze aniżeli wymagania nauk teoretycznych, odniosłaby wielką korzyść zarówno dla pewności swoich orzeczeń, jak też dla uproszczenia badania [wyróżnienie – J. M. B.].

I dalej:

Koniecznością coraz częściej podkreślaną staje się krytyczne, czysto naukowe, t.j. nie mające na celu żadnych potrzebą chwili bieżącej wywołanych zastosowań praktycznych, zbadanie testów od podstaw, rozpatrzenie zarzutów, przeciw nim podnoszonych, wprowadzenie odpowiednich poprawek – czyli praca nad teorią testów. Konieczność stworzenia teorii testów nie ulega wątpliwości; zadaniem zaś tej teorii byłoby takie przekształcenie testów, by umożliwiały one uzyskanie jak największej wiarygodności i ścisłości wyników, dorównującej o ile możliwości ścisłości badań w zakresie nauk przyrodniczych. Jedynie takie przekształcenie pozwoli na utrzymanie testów wśród metod naukowych psychologii teoretycznej; zarazem też psychologia stosowana, której wymagania co do ścisłości są mniejsze aniżeli wymagania nauk teoretycznych, odniosłaby wielką korzyść [...] (I, s. 10-11) [wyróżnienie – J. M. B.].

² Rzymską „I” będę oznaczał pierwszą część pracy Kreutza (Kreutz, 1927), a „II” – drugą część jego pracy (Kreutz, 1933).

Będąc żarliwym zwolennikiem, a później obrońcą (już mało skutecznym) metody introspekcyjnej, Kreutz był zdecydowanym przeciwnikiem rozprzestrzeniania się testów, co prawda, skonstruowanych zgodnie z ówczesnymi standardami naukowymi wypracowanymi przez psychologów, ale stosowanych przez osoby do tego nieprzygotowane czy osoby przypadkowe, bez wykształcenia psychologicznego i tylko powierzchownie przyuczone do ich stosowania. Dość powiedzieć, że – niestety bezkrytycznie – psychologowie prowadzili wśród nauczycieli kursy szkolące ich w stosowaniu, jak pisano, amerykańskich testów inteligencji (np. na Uniwersytecie Poznańskim). Był też zdecydowanym przeciwnikiem odwoływania się do metody testów mechanicznie przenoszonych z jednego gruntu kulturowego na drugi. Tu jego poglądy współbrzmiały z poglądami Władysława Witwickiego (1928). Dlatego też, pisał, nie należy przywiązywać większej wagi do wyników pozyskiwanych za pomocą owych pseudotestów. I tak, po pierwsze, pisał:

[...] do dziś niema wcale naukowej teorii testów, niema definicji tej metody, niema ustalonych i dostatecznie uzasadnionych przepisów postępowania ani warunków, jakim dobry test powinien odpowiadać, niema schematu konstrukcji testu, nie znamy dokładnie źródeł błędów, którychby się należało wystrzegać [...] (I, s. 5).

Krytycznie też odnosił się do rozprzestrzeniania (dodam od siebie, iż dzięki niefrasobliwości psychologów i braku, takiego jak dziś, nacisku na etyczną stronę testowania) bezwartościowych testów. Ironicznie przywoływał w tym kontekście słowa Lipmanna, który takich psychologów „stosujących nieogłędnie metodę testów” pogardliwie nazywał: *Draufpsychologen*. Po trzecie, „[...] są nikle wyniki prac, przeprowadzonych metodą testów i to zarówno w psychologii teoretycznej, jak też i w praktyce życiowej” (I, s. 6).

Mogłoby się wydawać, że są to już mocno zdezaktualizowane zastrzeżenia, ale – poza pierwszym (brak teorii testów) – pozostałe dwa wcale nie odeszły w przeszłość. Umasowienie studiów psychologicznych, które (co należało przewidzieć) prowadzone są także w ośrodkach nie mających ani dobrze przygotowanej (teoretycznie i praktycznie) kadry, ani wystarczającej liczby materiałów ćwiczeniowych, nie sprzyja kształceniu studentów na odpowiednio wysokim poziomie psychometrycznym. Mało tego, przyspieszone kursy stosowania testów pobierają, na studiach podyplomowych, osoby, które nie mają przygotowania psychologicznego.

Kreutz nie był jednak „totalnym” krytykiem testów psychologicznych. Widział ich pozytywne strony. Warto to podkreślić, gdyż jako jeden z ostatnich w powojennej Polsce aktywny obrońca metody introspekcji (i znowu, w jakimś sensie, prekursor późniejszych prac nad metodą: *protocol analysis* czy *thinking aloud protocol*; na ten temat por. Ericsson, Simon, 1993, czy Tyszka, 1995) był postrzegany też jako krytyk podejścia psychometrycznego w badaniach psychologicznych. Tak jednak nie było. W ostatniej, wydanej za jego życia, pracy *Metody badań współczesnej psychologii* (por. Kreutz, 1962), w której zawarł przeredagowane wersje przytaczanych tu dwóch części jego przedwojennej monografii, nie zmienił swojego pozytywnego zdania na ten

temat. Aczkolwiek upominanie się o teorię testów nie mogło już mieć tego wydźwięku co przed laty. Wszak Harold Gulliksen opublikował w 1951 r. swoją fundamentalną monografię: *Theory of mental test scores*, a w 1955 ukazała się również fundamentalna praca Lee J. Cronbacha i Paula Meehla (1955) poświęcona wiązaniu testu z teorią psychologiczną, która go, w sensie psychologicznym, usensawnia – trafności teoretycznej (*construct validity*). Dziś jest to warunek *sine qua non*, który musi być spełniony przez każdego psychologa przystępującego do konstruowania testu czy do interpretacji wyników testu psychologicznego (por. AERA/APA/NCME, 2007).

Cofnijmy się jednak kilkadziesiąt lat w czasie. Kreutz wymienił cztery, jego zdaniem, najważniejsze zalety testów. Po pierwsze, „testy, chociaż nie wszystkie, dają możliwość liczbowego oznaczania stopnia zdolności psychicznych na miejsce nieścisłych, względnych określeń, jakimi poprzednio trzeba było się posługiwać” (I, s. 8). Po drugie, podkreśla „obiektywność metody testów”: „[...] bez testów musi się mówić o fantazji, zdolności kombinacyjnej i t. p. a wyrazy te są przeważnie nieścisłe i wieloznaczne” (I, s. 8). Po trzecie, zwraca uwagę na ścisłość wyników uzyskiwanych za pomocą testów: „[...] podobne wyniki, jakie osiąga się testami, można osiągnąć również przez obserwację zachowania się osób badanych. Jednakże obserwacja [...] nie da się ująć w ścisłe ramy i przepisy” (I, s. 8). Po czwarte, „[...] ustawiczna styczność badacza z wielką ilością ludzi, kontakt z żywym materiałem, a to zaznacza się bardzo dodatnio w porównaniu z badaniami, zwłaszcza dawniejszemi, posługującymi się wyłącznie metodą introspekcji i to spekulacyjnej, wykonywanymi przez badacza przy biurku, w odcięciu od ludzi, operującego przykładami przeważnie fikcyjnymi” (I, s. 8-9).

Kreutz zalecał przede wszystkim podejmowanie ścisłych, odwołujących się do metody naukowej, prac nad podstawami testów, nad czynnikami, których oddziaływanie w trakcie badania testowego ogranicza interpretację wyników testowych oraz nad – jak pisał – doskonaleniem metody testów. Tu prezentował własne, oryginalne poglądy.

WNIOSKI Z BADAŃ KREUTZA DLA DOSKONALENIA METODY TESTÓW

Czynniki zmieniające rezultat testu

Gdyby spojrzeć na część empiryczną projektu badawczego – zrealizowanego przez Kreutza – z punktu widzenia dzisiejszych standardów metodologicznych, odnoszących się do prac empirycznych w psychologii (pomijam to, że prowadzony był wedle paradygmatu introspekcyjnego), to jest on nader skromny i nie jest wolny od usterek. Znacznie ważniejsza i koncepcyjnie nośna – właśnie z punktu widzenia, jak pisał Kreutz, doskonalenia metody testów – jest wnikliwie potraktowana część wnioskowa.

Samo badanie miało charakter, jak byśmy dziś powiedzieli, *replikacji* (z pewnymi modyfikacjami) badań: (1) w części I były to badania Claparède’a (z 1919 r.) i Boga (z 1922 r.), (2) w części II były to badania Boga

(z 1922 r.); ta część badania poprzedzona była wnikliwą i krytyczną analizą ustaleń poczynionych przez Kruegera i Spearmana (z 1906 r.).

Osobami badanymi byli studenci Uniwersytetu Jana Kazimierza we Lwowie, uczestniczący w ćwiczeniach z zakresu psychologii eksperymentalnej, prowadzonych przez Kreutza, który jednocześnie prowadził te badania. Wiek osób badanych wynosił od 18 do 24 lat. Przebadano (uzyskując komplet wyników) 12 osób (9 mężczyzn i 3 kobiety). Osobom badanym odczytywano (czyli to sam Kreutz) trzykrotnie materiał testowy. Każdorazowo był to inny ciąg ośmiu liczb dwucyfrowych, składający się na test do badania pamięci bezpośredniej (zaczepnięty ze Skali Inteligencji Bineta-Simona). Osoby badane odtwarzały zapamiętane liczby, zapisując je na przygotowanych przez eksperymentatora kartkach. Eksperyment powtarzano osiem razy (co tydzień, zawsze o tej samej porze dnia – o godz. 19.00). Badania miały charakter zbiorowy. W końcowym efekcie Kreutz dysponował materiałem obejmującym $3 \times 8 = 24$ wyniki dla każdej z 12 osób badanych. Metodyka postępowania badawczego była zbliżona do tej, którą posłużyli się E. Claparède (w 1919 r.) i H. Bogen (w 1922 r.). To, co uderza swoją wysoką starannością i dbałością o każdy szczegół, to wnikliwa analiza wyników i staranne dokumentowanie każdego kroku wnioskowania.

Nie wchodząc w ów bardzo szczegółowy opis wnioskowania, możemy przejść od razu do tych najważniejszych wniosków. Zdaniem Kreutza (I, s. 70), zmienność rezultatów badania testowego jest pochodną zmienności interindywidualnej. Sama obserwowana zmienność może być potraktowana jedynie jako kryterium, jak pisze, wartości przeprowadzonych badań (Kreutz używa terminu „eksperyment”), a nie jako kryterium użyteczności testu. I: „zmienność rezultatów zaprzecza stałości zdolności lub istnieniu zależności prostej między zdolnością a rezultatem testu”.

Moim zdaniem warto powrócić do problemu czynników zmieniających rezultaty badań testowych. Podejmując ten problem Kreutz deklarował, iż: „omawiając poszczególne czynniki zmienne, będę się równocześnie zastanawiać nad sposobami ich usunięcia, gdyż ostatecznym celem pracy [...] jest ulepszenie metody testów” (II, s. 44).

Kreutz analizował zachowanie się osoby badanej w trakcie rozwiązywania testu (badanie testowe przeprowadził na sobie metodą introspekcji) – chodziło o dokładne poznanie testu oraz wymagań stawianych osobie badanej. Kreutz podaje protokoły z tych analiz, a czytelnik może sam sprawdzić, czy zgadza się z wnioskami badacza (tu: Kreutza). Tak przedstawione badania bez trudu można poddać replikacji; dziś powiedzielibyśmy, że Kreutz dopełnił wymogu *intersubiektywnej komunikowalności* (w sensie zasady racjonalności K. Ajdukiewicza, 1983).

Podstawowym dla Kreutza problemem było znalezienie tych warunków (nazywał je „fenomenologicznymi”), które wywołują zmianę wyników testu przy ponownym jego użyciu. Interesowało go też i to, czy możliwe jest ich zneutralizowanie. Owe warunki zmienności związane są wyłącznie z tymi czynnikami, które uległy zmianie przy przejściu od pierwszego badania do drugiego. Ponieważ nie każdy czynnik, który uległ zmianie po pierwszym

badaniu, musi być odpowiedzialny za wywołanie obserwowanej zmienności rezultatu testowego, więc należy, jak pisze Kreutz (II, s. 40) – odróżniać „czynniki zmienne” oraz „warunki zmienności”. Każdy warunek zmienności wyniku testowego ma status czynnika zmiennego. Oczywiście nie musi być odwrotnie; nie każdy czynnik zmienny jest warunkiem zmienności. W prowadzonych analizach interesowały go tylko te czynniki zmienne, które były możliwymi warunkami zmienności. Kreutz wprowadził kilka ograniczeń związanych z czasem trwania pierwszego i drugiego badania oraz z warunkami fizycznymi badania (ten sam pokój dla obu badań). Osoby badane poddane były badaniu przeprowadzonemu za pomocą „testu do badania pamięci liczb”. Badanie grupowe poprzedzone było badaniem przeprowadzonym na samym sobie.

Pominę tu omówienie bardzo wnikliwej analizy porównawczej materiału zebranego od grupy osób badanych i od samego siebie. Skupię się natomiast na wyłonionej przez Kreutza klasyfikacji warunków zmienności. Podstawą do jej sporządzenia była analiza danych eksperymentalnych (z samobadania Kreutza i badania grupy studentów). Kreutz uważał (II, s. 127), że tak drobiazgowej analizie powinny być poddane wszystkie wówczas dostępne testy, gdyż: „znajomość czynników zmiennych stanowi pierwszy i niezbędny krok do ulepszenia metody”.

Nawiasem mówiąc, zajmując się w latach osiemdziesiątych polską adaptacją testów wchodzących do Skali Inteligencji WAIS-R też (wzorem innych badaczy zajmujących się problemem podatności tych testów na różne czynniki mogące wpływać na wyniki osiągane przez osoby badane tymi testami) zestawialiśmy listę takich czynników. Przykładowo, autor (por. Brzeziński, 1998, s. 143) sporządził taką listę dla testu Wiadomości z WAIS-R. Znajomość tych czynników oraz ich uwzględnienie w diagnostycznej analizie wyników badania przeprowadzonego za pomocą WAIS-R nie tylko wpływa na trafność diagnozy, ale też ułatwia (w zastosowaniach klinicznych) projektowanie postępowania korekcyjnego. A, przyznaję (ze wstydem), klasyfikacja dokonana przez Kreutza nie była mi wcześniej znana.

Ale przejdźmy do sygnalizowanej klasyfikacji. Wpierw Kreutz dzieli czynniki zmienne na dwie klasy:

1. Czynniki wytwarzające: związane z – jak byśmy dziś powiedzieli – czynnościami osoby badanej rozwiązującej test. Pozostają one „w najściślejszym i bezpośrednim związku” z wykonywanym przez osobę badaną testem psychologicznym.

2. Czynniki modyfikujące: są one „[...] albo bardzo ogólnymi, ale dalszymi warunkami wytworu, jak np. samo istnienie danej osoby, powietrze konieczne do oddychania, zgoda na przeprowadzenie eksperymentu i. t. p. albo też działają, zmieniając w pewnej mierze wytwór, lecz same do jego powstania się nie przyczyniając, np. myśli obce, efekty i. t. p.” (II, s. 130-131).

Te dwie główne klasy czynników zmiennych Kreutz dzieli na podklasy. Do pierwszej – czynniki wytwarzające – zalicza:

- 1.1. Czynności dozwolone: są one przewidziane przez konstruktora testu i ujęte w opisie technicznym testu (adresowanym do użytkownika testu)

oraz w instrukcji przekazywanej przez badacza osobie badanej testem. Do nich Kreutz zalicza:

1.1.1. Rodzaj czynności (w. źródło błędów)

1.1.2. Ilość czynności

1.1.3. Przypadkowość czynności. Do tej kategorii Kreutz zaliczył też zgadywanie.

1.2. Czynności niedozwolone: Kreutz utożsamia je z podejmowanymi przez osobę badaną czynnościami, które nie zostały przewidziane w instrukcji testowej. Podaje następującą ich definicję: „[...] dozwolona jest tylko ta czynność, którą eksperymentator chce badać, wszystkie zaś inne czynności nie są dozwolone” (II, s. 140). Także i one, podobnie jak czynności dozwolone, obejmują rodzaj, ilość i przypadkowość podejmowanych czynności. Niestety, „usuwanie ich natrafia na znaczne trudności, gdyż z samego rezultatu niepodobna już rozpoznać, czy został osiągnięty na drodze dozwolonej” (II, s. 142).

Z kolei do czynników modyfikujących zalicza:

2.1. Warunki psychiczne („warunki psychiczne, zachodzące w osobie badanej”)

2.1.1. Intelktualne

2.1.2. Emocjonalne

2.1.3. Wolicjonalne (2.1.3.1. zgoda osoby badanej; 2.1.3.2. „mniejsze lub większe staranie się o dobry wynik, co potocznie można wyrazić mówiąc, że osobie badanej zależy mniej lub więcej na wyniku, albo że chce lub że jej się nie chce pracować”, II, s. 153; Kreutz podnosi tu: „stopień” starania się o dobry wynik; problem motywów, jakimi kieruje się osoba badana, której zależy na tym, aby dobrze wypaść w badaniu; przyzwyczajenia i skłonności osoby badanej; stan zdrowia osoby badanej).

2.2. Warunki fizyczne

2.2.1. Zewnętrzne (poza osobą badaną); Problem uwarunkowań zewnętrznych wyników testów pojawił się w latach sześćdziesiątych i siedemdziesiątych XX wieku: w koncepcjach Lorda i Novicka (1968) oraz Cronbacha i współautorów (1972). Badacze pracujący nad tym zagadnieniem odwoływali się do planów ANOVA. Sam pisałem na ten temat prawie trzydzieści lat temu (Brzeziński, 1984).

2.2.2. Wewnętrzne (inaczej: fizjologiczne)

Poza warunkami zmiennymi Kreutz wymienia jeszcze warunki stałe: „należą do nich przede wszystkim dyspozycje psychiczne lub psychofizyczne, których stopień u poszczególnych osób chcemy właśnie przy pomocy metody testów mierzyć. Następnie wymienić należy pewne stałe cechy organizmu fizycznego osoby badanej i stałe składniki otoczenia, w którym ona przeprowadza eksperyment” (II, s. 172). Chciałem zwrócić uwagę na następującą wypowiedź: „[...] gdy chodzi o badanie dzieci szkolnych, zwłaszcza w eksperymentach jednostkowych, nauczyciel nie powinien być eksperymentatorem, gdyż z jego osobą kojarzą się stale myśli o szkole, notach, klasyfikacji, które mogą działać hamująco” (II, s. 172). To zalecenie wcale łatwo nie przebijало się do świadomości psychologów i pedagogów. Jeszcze dziś spotykamy badania nie respektujące tego ważnego zalecenia. Ten problem był w wielu badaniach

podnoszony w latach sześćdziesiątych (i późniejszych) pod „szyldem”: *The social psychology of psychological experiment*, a wiąże się to z interakcją „badacz–osoba badana” (także „nauczyciel–uczeń”) – por. m.in. prace R. Rosenthala, M. Rosenberga, M. T. Orne’a (w: Brzeziński, 2011; Brzeziński, Siuta, 1991; Trusz, 2010, 2013).

„Konieczna modyfikacja metody testów”

Wnikliwa (chyba najbardziej wnikliwa ze znanych mi analiz czynników mogących modyfikować wynik badania testowego) analiza – wyżej przedstawiona z konieczności w wielkim skrócie – czynników zmiennych jako warunków zmienności rezultatów testów doprowadziła Kreutza do sformułowania zaleceń dla konstruktorów i użytkowników testów. Te zalecenia Kreutz traktował jako konieczne do wprowadzenia w życie, aby można było sensownie (i odpowiedzialnie) posługiwać się testami psychologicznymi w pracy naukowej i diagnostycznej.

Czy warto się nad nimi „pochylić”? Czy nadal te przemyślenia mogą inspirować tych wszystkich, którzy chcą doskonalić metodę testów? Moim zdaniem – warto. Mało tego, pewne pomysły Kreutza znalazły się w koncepcjach kolejnej generacji psychometrów (lata sześćdziesiąte i siedemdziesiąte ubiegłego wieku). Mam na myśli przywoływanych już w tym artykule badaczy: Lorda i Novicka (1968), Cronbacha i współautorów (1972), a w Polsce: Nowakowską (1975) czy Brzezińskiego (1984).

Poświęcony idei naprawy metody testów VI rozdział II części rozprawy Kreutza zaczyna się od ważnej konstatacji, iż „[...] poszczególny rezultat zależy nie tylko od zdolności, lecz także od wielu innych warunków, częściowo stałych, w większości zaś części zmieniających się przy powtarzaniu eksperymentu i to z chwili na chwilę” (II, s. 174). Pomysł, którym dzieli się z czytelnikami, sprowadza się do tego, aby biorąc pod uwagę możliwie jak najpełniejszy zestaw warunków modyfikujących (por. poprzedni punkt), określić – w danej serii badań – dolną i górną granicę, jak pisze, „zmian dobroci rezultatu”. W serii badań (wyobraźmy sobie, że nieskończenie długiej) możliwe będzie wywołanie wszystkich możliwych rezultatów, w możliwych granicach, uzyskanych pod wpływem wszystkich możliwych kombinacji warunków zmiennych – i tych niekorzystnych, i tych sprzyjających rezultatowi badania testowego. Dolna granica wyniku jest praktycznie dostępna do osiągnięcia przez hipotetyczną osobę badaną. Kiedy? Ano wówczas, gdy osoba badana odmawia wykonania testu, gdy celowo doprowadza do „zerowego” wyniku albo gdy uzyskuje zerowy poziom mierzonej przez test cechy, np. osoba ze znacznym stopniem upośledzenia umysłowego może w Skali Inteligencji WAIS uzyskać w poszczególnych testach wyniki przeliczone równe 1. Problem stwarza górna granica, określająca jej faktyczne możliwości, których nie jest w stanie przekroczyć. Mimo to możemy sobie wyobrazić taki splot czynników, przy którym osoba badana może ujawnić swój wynik maksymalny. Problem praktyczny wiąże się z rozpoznaniem tych warunków i stworzeniem sytuacji badania przez ową optymalną dla danej osoby badanej kombinację warunków

dookreślonej. Kreutz uważa, że „[...] jako miara stopnia dyspozycji poszczególnych osobników może wchodzić w rachubę jedynie rezultat uzyskany przy najbardziej sprzyjających warunkach, pozwalających na jej pełne zaktualizowanie się, który nazwiemy rezultatem „maksymalnym” (II, s. 179) [wyróżnienie – J. M. B.]. Oczywiście należy z tak pomyślanego badania wyeliminować przypadek czy korzystanie z niedozwolonej pomocy czy ułatwień.

Zauważmy, że w modyfikacji klasycznej teorii testów, przedstawionej przez Frederica M. Lorda i Melvina R. Novicka (1968), wynik uzyskany przez jakąś osobę badaną w jakimś teście traktowany jest: „jako realizacja zmiennej losowej [...], której wartościami są wszystkie potencjalnie możliwe u danej osoby wyniki tego testu, a rozkład prawdopodobieństwa tych wyników jest charakterystyczny dla tej osoby; ten rozkład prawdopodobieństwa to właśnie rozkład skłonności” (Nowakowska, 1975, s. 26). Czy nie można w tej definicji odnaleźć oddźwięku koncepcji Kreutza? Wszak w tym przedziale „możliwych” wyników mieści się i ten, nazwany przez Kreutza „maksymalnym”.

Po wnikliwej dyskusji nad warunkami uzyskiwania możliwego rezultatu „maksymalnego” Kreutz dochodzi do ostatecznego sformułowania swojej koncepcji modyfikacji badania testowego:

[...] poprawną miarą zdolności jest jedynie rezultat najlepszy z pewnej ilości rezultatów, uzyskanych metodą testów, przy wykluczeniu przypadku i pomocy niedozwolonych oraz przy zachowaniu niezmiennych czynności wytwarzających. Należy zatem brać za podstawę oceny stopnia zdolności wynik jednego tylko eksperymentu, inne zaś eksperymenty odrzucić jako nieudane (II, s. 184).

Ostatecznie Kreutz dochodzi do sformułowania trzech zasad „poprawnego stosowania testów” (II, s. 186-189):

1. „Przeprowadzenie badań tylko przy pomocy testów dokładnie zbadanych i ulepszonych”.

2. „Obserwacja zachowania się osób badanych w różnych warunkach. Celem dokonywania obserwacji należy przeprowadzić szereg eksperymentów i to najlepiej z osobami tej grupy społecznej, którą mamy zamiar później badać, to jednostkowo i zbiorowo, i obserwować przy współudziale pomocników ich czynności.

3. „Rozumowa analiza testu [...] zastanawiamy się mianowicie nad tem, jakimi środkami, przy pomocy jakich czynności możnaby dane zadanie rozwiązać, opierając się na doświadczeniu życiowym i na wyszukaniu wszelkich teoretycznych możliwości”.

„Takie są główne sposoby badania testu obok badań statystycznych, którym, jako zbyt powierzchownym, większej wartości nie przypisuję” [podkreślenie – J. M. B.].

Monografię zamyka – przeprowadzona na kilku stronach – dyskusja dotycząca różnych szczegółowych kwestii natury praktycznej, związanych ze stwarzaniem warunków, w których osoba badana może dojść do wyniku „maksymalnego”. Domyślam się, że Czytelnika nurtuje pytanie: A ile takich badań trzeba przeprowadzić, aby uzyskać ów mityczny wynik „maksymalny”?

Kreutz (II, s. 192) podaje, że badanie powinno się przeprowadzić na co najmniej... dwóch posiedzeniach, w różnych dniach (celem uniknięcia zmęczenia i usunięcia chwilowo niekorzystnych warunków). Aby zaś uniknąć zmęczenia i znużenia osoby badanej, należy skracać testy (współcześnie znane są różne sposoby przygotowywania tzw. *short form* testów wchodzących do Skali Inteligencji WAIS czy WISC; w Polsce problem skracania testów inteligencji i uzdolnień, z wykorzystaniem do tego celu najnowszego modelu psychometrycznego IRT, podjął Paweł Kleka, 2013). Kreutz dopuszczał też posługiwanie się testami równoległymi (choćby dlatego, aby wyeliminować wpływ czynnika wyuczalności wyniku).

Kreutz też „upierał się” przy tym, aby psycholog posługujący się testem psychologicznym nie tylko ograniczał się do analizy wyniku ogólnego (na ogół, w testach inteligencji i uzdolnień, przyjmującego postać sumy punktów za udzielone przez osobę badaną poprawne odpowiedzi na poszczególne zadania), ale aby też analizował odpowiedzi na poszczególne pozycje testu. Ku takiemu postępowaniu skłaniał się Rapaport (1945) w pierwszym rozdziale dwutomowej monografii (na temat psychologicznych metod diagnostycznych), poświęconym Skali Inteligencji Wechslera-Bellevue (wersja z 1939 r.). Sam także rekomenduje tak „głębokie” zejście w analizie wyników testu (por. Brzeziński, 2012).

Konieczność analizy procesu odpowiadania na poszczególne pozycje testowe została także zapisana w wymaganiach stawianych konstruktorom testu, którzy przystępują do analizy trafności testu. Wedle *Standardów dla testów stosowanych w psychologii i pedagogice* (AERA/APA/NCME, 2007) możliwe są następujące źródła danych walidacyjnych:

1. Dane, których źródłem jest test.
2. Dane oparte na analizie procesu udzielania odpowiedzi.
3. Dane wynikające z analizy struktury wewnętrznej testu:
 - 3.1. Dane zbieżne i różnicowe,
 - 3.2. Związek między testem a kryterium,
 - 3.3. Uogólnianie danych dotyczących trafności.
4. Dane oparte na konsekwencjach testowania.

Jak nietrudno zauważyć, punkt 2 jest spełnieniem postulatu Kreutza.

I jeszcze jeden oddźwięk postulatów Kreutza, które można odnaleźć w *teorii wyników generycznych i teorii uniwersalizacji*, będącej kolejną, po pracy Lorda i Novicka, próbą stworzenia modelu psychometrycznego (Cronbach i in., 1972). Psycholog ustala „oczekiwany wynik prawdziwy” (nie tak jak w *klasycznej teorii testów* Gulliksen), uzyskany przez jakąś osobę badaną w jakiejś losowej wersji („równoległej” w sensie wzoru Spearmana-Browna) testu. *Wynik generyczny* to średnia z wartości prawdziwych. Cronbach i współautorzy skupiają się na badaniu wpływu różnych czynników (mówił o nich Kreutz!) na wynik testowy przez odwołanie się do planów eksperymentalnych ANOVA. Tak poprowadzone badania tych czynników, traktowanych jako odrębne źródła wariancji wyniku testowego, są źródłem danych empirycznych dla formułowanej przez nich *teorii uniwersalizacji*. Ogólnie, analizuje się sche-

maty eksperymentalne: „wersje testu * warunki”. Brzeziński (1984) skonstruował wskaźnik DT (*dobroci testu*) odwołujący się do jednego z planów ANOVA.

*

Wnikliwa, po latach, lektura pracy Mieczysława Kreutza doprowadza do dwóch refleksji. Pierwsza, nazwijmy ją merytoryczną, odnosi się do przekazu teoretycznego i metodologicznego. W pracy – co starałem się ukazać – znalazły się dziś jeszcze „żywe” koncepcje dotyczące idei wyniku testowego (jak powinno się go ujmować – „wynik maksymalny”), klasyfikacji czynników wywierających wpływ na wynik testowy, analizy odpowiedzi na poszczególne pozycje testu czy ogólniejsze zalecenia odnoszące się do procedury badania testowego. Dziś niewiele osób sięga do wcześniej opublikowanych prac psychologów. Uważa się, a przykład monografii Kreutza to pokazuje, niesłusznie, iż takie „stare” prace nie wniosą nic nowego.

Z kolei druga refleksja jest natury ogólniejszej i odnosi się do losu polskich, oryginalnych prac naukowych, które ogłaszane były drukiem w języku polskim. Można zapytać, dlaczego nie spotkały się one z należyтым rezonansem w światowym piśmiennictwie? Przecież Kreutz nie był odosobniony, jeśli chodzi o oryginalność i nośność prac badawczych. Weźmy chociażby monografię Jana Władysława Dawida (z 1911 r.) pt. *Inteligencja, wola i zdolność do pracy*. To w niej opisany został oryginalny „test obrazkowy”, pomysłu Jana W. Dawida, wyprzedzający test Decroly’ego, który został skonstruowany w 1913 r. Tylko Claparède wymienia nazwisko Dawida, gdy pisze o tej metodzie. Podobny los spotkał prace: A. Qualla (z 1938 r.), który zajął się wartością diagnostyczną badań kwestionariuszowych, czy F. Baumgarten-Tramerowej (z 1938 r.), która analizowała strategie odpowiedzi osób badanych na zadania testów inteligencji.

BIBLIOGRAFIA

- AERA/APA/NCME (American Educational Research Association, American Psychological Association, National Council on Measurement in Education) (2007). *Standardy dla testów stosowanych w psychologii i pedagogice*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Ajdukiewicz, K. (1983²). *Zagadnienia i kierunki filozofii. Teoria poznania. Metafizyka*. Warszawa: Czytelnik.
- Bridgman, P. W. (1927). *The logic of modern physics*. New York: MacMillan.
- Brzeziński, J. (1984). Badanie testu psychometrycznego metodą analizy wariancji. [W:] J. Brzeziński (red.), *Psychologiczne i psychometryczne problemy diagnostyki psychologicznej* (s. 9-45). Poznań: Wydawnictwo Naukowe UAM.
- Brzeziński, J. (1998²). Wiadomości. [W:] J. Brzeziński, E. Hornowska, E. (red.), *Skala Inteligencji Wechslera WAIS-R. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (s. 131-164). Warszawa: Wydawnictwo Naukowe PWN.

- Brzeziński, J. (2008). Stan świadomości psychometrycznej polskich psychologów przed rokiem 1939. [W:] M. Toeplitz-Winiewska, A. Sękowski (red.), *Polskie Towarzystwo Psychologiczne. Historia. Teraźniejszość. Przyszłość. 1907-2007. Księga Jubileuszowa* (s. 41-58). Warszawa: Polskie Towarzystwo Psychologiczne.
- Brzeziński, J. (2011⁵). *Metodologia badań psychologicznych*. Warszawa: Wydawnictwo Naukowe PWN.
- Brzeziński, J. (2012). Co to znaczy, że wyniki przeprowadzonych przez psychologów badań naukowych poddawane są analizie statystycznej? *Roczniki Psychologiczne*, 15, 3, 7-39.
- Brzeziński, J., Siuta, J. (red.) (1991). *Spółeczny kontekst badań psychologicznych i pedagogicznych. Wybór tekstów*. Poznań: Wydawnictwo Naukowe UAM.
- Choynowski, M. (1959). Elementy teorii testów psychologicznych. *Przegląd Psychologiczny*, 3, 151-170.
- Cronbach, L. J. (1957/2007). Dwa nurty psychologii naukowej. [W:] J. Brzeziński (red.), *Metodologia badań psychologicznych. Wybór tekstów* (s. 21-43). Warszawa: Wydawnictwo Naukowe PWN.
- Cronbach, L. J., Gleser, G. C., Nanda, H., Rajaratnam, N. (1972). *The dependability of behavioral measurement*. New York: Wiley.
- Cronbach, L. J., Meehl, P. (2005/1955). Trafność teoretyczna testów psychologicznych [Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302]. [W:] J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 403-430). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Ericsson, K. A., Simon, H. A. (1993). *Protocol analysis. Verbal reports as data* (rev. ed.). Cambridge, MA: A Bradford Book-The MIT Press.
- Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill (wyd. II – 1954).
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Kleka, P. (2013). *Zastosowanie teorii odpowiadania na pozycje testowe (IRT) do tworzenia skróconych wersji testów i kwestionariuszy psychologicznych (mps pracy doktorskiej, Instytut Psychologii, Uniwersytet im. Adama Mickiewicza w Poznaniu)*.
- Kreutz, M. (1927). *Zmienność rezultatów testów. Część I: Znaczenie zmienności rezultatów dla wartości testów*. Lwów: Polskie Towarzystwo Filozoficzne.
- Kreutz, M. (1933). *Zmienność rezultatów testów. Część II: Przyczyny zmienności rezultatów i konieczna modyfikacja metody testów*. Lwów: Polskie Towarzystwo Filozoficzne.
- Kreutz, M. (1962). *Metody badań współczesnej psychologii*. Warszawa: PWN.
- Lord, F. M., Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Nowakowska, M. (1975). *Psychologia ilościowa z elementami naukometrii*. Warszawa: PWN.
- Rapaport, D. (1945). *Diagnostic psychological testing* (t. 1). Chicago: Yearbook Publ.

- Stevens, S. S. (1935/2000). Operacyjne definiowanie pojęć psychologicznych. [W:] J. Siuta, K. Krzyżewski (red.), *Behawioryzm i psychologia świadomości* (s. 107-115). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Trusz, S. (2010). *Efekt oczekiwań interpersonalnych w edukacji*. Kraków: Wydawnictwo Naukowe Uniwersytetu Pedagogicznego.
- Trusz, S. (red.) (2013). *Efekty oczekiwań interpersonalnych. Wybór tekstów*. Warszawa: Wydawnictwo Naukowe Scholar.
- Thurstone, L. L. (1926). The mental age concept. *Psychological Review*, 33, 268-278.
- Tyszka, T. (1995). *Czy powrót do introspekcji*. Warszawa: Wydawnictwo Naukowe PWN.
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Witwicki, W. (1928). O narodowych testach amerykańskich do badania inteligencji, *Psychotechnika*, 2, 6, 23-32.