

PIOTR MACIEJ SKORUPIŃSKI

Warszawa

WALIDACJA KONSTRUKTU

*Drogiemu Profesorowi Bolesławowi Niemierce –
z podziękowaniem za dotychczasowe
i w oczekiwaniu na kolejne
rozliczne inspiracje*

*Wysoką rangę naukową nadały trafności pomiaru prace Lee Cronbacha [...]¹. Zgoda – kluczowa rola Cronbacha w kształtowaniu pojęcia trafności teoretycznej sprowadzała się nie tylko do współautorstwa terminu i związanych z nim początkowo znaczeń. Cronbach także – można by rzec – popularyzował tę nową konwencję nazewniczą, wyjaśniając koncepcję walidacji teoretycznej w rozlicznych publikacjach. O pomnikowych projektach Amerykańskiego Towarzystwa Badań Edukacyjnych (AERA) oraz Narodowej Rady ds. Pomiarów Edukacyjnych (NCME) wspominałem już w *Genezie pojęcia trafności teoretycznej*². Również w klasycznych już *Podstawach psychologicznego testowania (Essentials of psychological testing)*, których kolejnych pięć edycji ukazało się w latach 1949–1990³, Cronbach poświęcił walidacji konstruktowi odrębny rozdział.*

W wydaniu II tego podręcznika – przygotowanym już po introdukcji terminu „trafność teoretyczna” – w problematykę trafności Cronbach wprowadził rozważaniami na temat „jakości” (ang. *quality*) testu, którą się

¹ Niemierko B., *Jaki pomiar dydaktyczny jest nam potrzebny?*, [w:] B. Niemierko, J. Syglicz, W. Walczak (red.), *Trafność pomiaru jako podstawa obiektywizacji egzaminów szkolnych*, Łódź 2003, s. 20.

² Skorupiński P.M., *Geneza pojęcia trafności teoretycznej*, [w:] B. Niemierko (red.), *Diagnostyka edukacyjna. Teoria i praktyka*, Kraków 2004, s. 32–33.

³ Cronbach L.J., *Essentials of psychological testing*, wyd. I: New York 1949; wyd. II: New York 1960; wyd. III: New York 1970; wyd. IV: New York 1984; wyd. V: New York 1990.

osądza (ang. *to judge*), ewentualnie ocenia (ang. *to evaluate*), i na którą się składa – m.in. – trafność, właściwość decydująca o wartości (ang. *value*) testu ze względu na określony cel jego zastosowania. W związku z powyższym – zdaniem Cronbacha – *nie możemy zadawać pytania ogólnego „Czy ten test jest trafny?”*. Pytanie, które należy zadać, brzmi: *„Jak trafny jest ten test dla decyzji, którą zamierzam podjąć?”* [...] ⁴. Z tego punktu widzenia Cronbach charakteryzuje przyjętą w *Technical recommendations for psychological tests and diagnostic techniques* typologię trafności (trafność prognostyczna, trafność diagnostyczna, trafność treściowa i trafność teoretyczna) ⁵, wiążąc ją z typologią decyzji, w procesie podejmowania których wykorzystuje się wyniki testów (decyzje związane z celami praktycznymi – selekcyjne, klasyfikacyjne i wynikające z ewaluacji programów edukacyjnych – lub naukowymi). W tym kontekście przedstawia problematykę trafności teoretycznej, twierdząc, iż kiedy test stosowany jest w celu opisu z użyciem pojęć teoretycznych – zwanych konstruktami – a jego wyniki interpretowane są w terminach tych właśnie konstruktów, to taką interpretację należy poddać walidacji teoretycznej (ang. *construct validation*) ⁶.

W ujęciu Cronbacha, *walidacja teoretyczna jest analizą znaczenia wyników testu w języku terminów psychologicznych* ⁷. Co więcej, przebiega analogicznie do procesu, podczas którego rozwijane są teorie naukowe – od wyjściowego „aktu wyobraźni” ⁸, konstytuującego wstępne znaczenie wyników, poprzez wyprowadzenie sprawdzalnych hipotez z teorii stanowiącej podstawę tego znaczenia oraz badania empiryczne, których celem jest weryfikacja wspomnianych hipotez. W trakcie walidacji sięga się chociażby po analizy zadań i związków pomiędzy nimi, analizy korelacji pomiędzy testem a innymi miarami tego samego konstruktów, wyniki procedur eksperymentalnych czy obserwacji (np. czynności wykonywanych podczas rozwiązywania zadań). Badanie takie powinno zostać przeprowadzone również ze względu na użytkownika testu, który – być może – chciałby wiedzieć, jak i z jaką ufnością można interpretować wyniki danego pomiaru, i któremu należy przedstawić wszelkie dostępne dowody, nie tylko te świadczące na korzyść sugerowanego znaczenia wyników ⁹.

⁴ Cronbach L.J., op. cit., wyd. II, s. 103.

⁵ Por. American Psychological Association, American Educational Research Association, National Council on Measurement used in Education, *Technical recommendations for psychological tests and diagnostic techniques*, „Psychological Bulletin” 1954, t. 51., nr 2 (supplement), s. 13–14.

⁶ Por. Cronbach L.J., op. cit., wyd. II, s. 104–105.

⁷ Ibidem, s. 120.

⁸ Ibidem, s. 121.

⁹ Por. ibidem, s. 123.

W ostatnim wydaniu *Podstaw psychologicznego testowania*, ukierunkowanym na kształtowanie umiejętności osądzania (ang. *to judge*) testów¹⁰, Cronbach modyfikuje swoją koncepcję walidacji. Twierdzi mianowicie, że *walidacja przekracza granice nauki w stronę argumentacji*¹¹. Dzieje się tak, ponieważ interpretacja wyników wiąże się z koniecznością jej perswazyjnej obrony na forum krytycznych użytkowników testu, których osąd przedstawionych im tez i argumentów konstruowany jest m.in. na podstawie cenionych przez nich wartości i poglądów. Argument walidacyjny (ang. *validation argument*) powinien zatem nie tylko uwzględniać empiryczne i logiczne dowody trafności treściowej, kryterialnej czy teoretycznej (a w wypadku tzw. silnej – ang. *strong* – walidacji także tzw. niewykluczone alternatywne hipotezy – ang. *plausible rival hypotheses* – podważające proponowaną interpretację wyników), lecz także obejmować odwołania do wartości, zdrowego rozsądku i innych czynników, uaktywnienie których czyni bardziej prawdopodobnym akceptację przyjętego rozumienia testu i jego wyników¹².

Po wtóre, Cronbach – określając walidację jako badanie solidności (ang. *soundness*) i odpowiedniości (ang. *relevance*) proponowanej interpretacji i podkreślając, że terminy dotychczas funkcjonujące jako nazwy tzw. typów trafności odnoszą się raczej do sposobów badania i rozumowania, których efekty integrowane są w procesie konstruowania argumentów walidacyjnych – twierdzi (być może – za Samuelem Messickiem), iż *pełne badanie przyniesie rozważania aksjologiczne: jakie są społeczne i osobiste konsekwencje, dobre i złe, zastosowania testu?*¹³.

Po trzecie, Cronbach uważa, iż interpretacja kryterialna, tłumacząca wyniki testu w języku ich korelacji z miarami zmiennych kryterialnych, oraz interpretacja kształtowana na podstawie badania związków pomiędzy zawartością testu a dziedziną testowania prowadzą do wyjaśniania wyników za pomocą – niekoniecznie wysublimowanych¹⁴ - pojęć teoretycznych (konstruktów) i pytań o trafność teoretyczną tych objaśnień. W tym kontekście Cronbach podkreśla, że *trafność testu* [rozumiana jako trafność

¹⁰ Por. Cronbach, op. cit., wyd. V, s. 145.

¹¹ Ibidem, s. 145.

¹² Por. ibidem, s. 185–187.

¹³ Ibidem, s. 152.

¹⁴ Cronbach używa wyrażenia *relatywnie prymitywna teoria* dla określenia teorii wyjaśniającej wyniki pomiarów w naukach społecznych. Argumentuje, iż przecież *przrodnicy potrzebowali wieków, aby ukształtować takie pojęcia, jak atom czy siła grawitacji* (ibidem, s. 179).

interpretacji jego wyników – przyp. aut.] i trafność konstruktów są nierozłączne¹⁵.

Po czwarte wreszcie, Cronbach, wzbogacając listę technik walidacyjnych, przedstawioną już w II edycji podręcznika (analizy zadań, wewnętrznych interkorelacji, stabilności wyników, efektów różnicowania procedur testowych czy treningu, korelacji z innymi miarami itd.), o technikę – zaproponowaną wcześniej przez Donalda T. Campbella i Donalda W. Fiskego – wielu cech mierzonych za pomocą wielu metod¹⁶, zaznacza, że nie istnieje prosty czy idealny schemat badawczy, który mógłby być zrealizowany w trakcie walidacji teoretycznej¹⁷. I nie tylko dlatego, że jest ona niekończącym się procesem sprawdzania hipotez.

Zasygnalizowany wyżej kłopot metodologiczny w innym świetle przedstawił Cronbach podczas sympozjum, które odbyło się na Uniwersytecie Stanu Illinois w 1985 r. w celu uhonorowania przechodzącego na emeryturę profesora Lloyd G. Humphreysa. W swoim referacie – zatytułowanym *Trafność teoretyczna po trzydziestu latach (Construct validity after thirty years)* – Cronbach stwierdził wprost: *Tę koncepcję trudno było przelożyć na działania [...]. Z tego względu mówię o niedokończonym przedsięwzięciu*¹⁸.

Ta koncepcja to oczywiście koncepcja trafności i walidacji teoretycznej, u której początków znajdowało się pytanie – sformułowane przez Komisję ds. Standardów Testowych (The Committee on Test Standards) Amerykańskiego Towarzystwa Psychologicznego – dotyczące dowodów, które mogłyby uzasadniać psychologiczną interpretację wyników testu¹⁹. Odpowiedź, przygotowana wówczas przez Paula M. Meehla i Roberta C. Challmana²⁰, sprowadzała się – zdaniem Cronbacha – do stwierdzenia, iż *przeprowadzić walidację jakiegoś typu inferencji oznacza sprawdzić, czy inne dowody są zgodne z tą inferencją, w granicach właściwej próby*

¹⁵ Ibidem, s. 181.

¹⁶ Por. Campbell D.T., Fiske D.W., *Convergent and discriminant validation by the multitrait-multimethod matrix*, „Psychological Bulletin” 1959, t. 56., s. 81–105. Por. Niemierko B., *Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe*, Warszawa 1975, s. 213–215 (a zwłaszcza przykład zastosowania metody Campbella-Fiskego, nazywanej za M. Choynowskim macierzą wielu cech i wielu metod).

¹⁷ Por. Cronbach L.J., op. cit., wyd. V, s. 160.

¹⁸ Tenże, *Construct validity after thirty years*, [w:] R.L. Linn (red.), *Intelligence. Measurement, theory, and public policy. Proceedings of a symposium in honor of Lloyd G. Humphreys*, Urbana 1989, s. 147.

¹⁹ Por. *American Psychological Association* i in., op. cit., s. 0 (Foreword).

²⁰ Por. Cronbach L.J., Meehl P.M., *Construct validity in psychological tests*, „Psychological Bulletin” 1955, t. 52., nr 4, s. 281–282.

*badawczej*²¹. Na taki sposób uzasadnienia zastosowania testu (nazwany później walidacją konstruktów) składają się kolejne próby potwierdzenia interpretacji wyników testu; interpretacji wykorzystujących w większości wypadków nie rozwinięte i ugruntowane teorie, lecz raczej – wedle określenia Cronbacha – konstrukcje (ang. *constructions*)²². Jednak przyjęcie stanowiska, iż akumulowanie wyników tych prób służy confirmacji proponowanej inferencji, nie rozwiązywało problemu praktycznego – kłopotu z procedurami czy schematami walidacyjnymi.

Zdaniem Cronbacha, największy wkład w przełożenie koncepcji trafności teoretycznej na czynności badawcze należy przypisać Campbellowi i jego współpracownikom: D. W. Fiskemu, J. C. Stanleyowi i T. D. Cookowi²³. Jednakże i ich „recepta na badanie trafności teoretycznej”²⁴ – a w szczególności macierz wielu cech i wielu metod (*multitrait-multimethod matrix*) – realizowana była zazwyczaj w sposób bezmyślny i mechaniczny²⁵, natomiast rozumowania i badania wynikające z uwzględnienia niewykluczonych hipotez alternatywnych (ang. *plausible rival hypotheses*) stanowiły rzadkość w materiałach zdominowanych przez zbiory różnorodnych korelacji²⁶.

Cronbach wskazuje kilka prawdopodobnych przyczyn takiego stanu praktyki walidacyjnej – a więc przewagi nieukierunkowanego empiryzmu i nieusystematyzowanej, przedteoretycznej wiedzy – wśród których wymienia stopień rozwoju tej dziedziny nauk społecznych, w której zaproponowany i zaakceptowany został konstrukt trafności konstruktów. W sytuacji braku programów silnej walidacji teoretycznej Cronbach, nie tracąc wiary w ich (tudzież nauki) wartość, proponuje *realistyczną strategię dla wysiłków związanych z walidacją konstruktów*²⁷. Koncepcja ta bazuje na przeświadczeniu, że *uczni [...] są zobowiązani do pomocy im współczesnym w przemyśleniu ich problemów i ocenie proponowanych*

²¹ Ibidem, s. 148–149.

²² Cronbach: *Ten luźny zbiór konceptów i implikacji użytych w typowej interpretacji testu nazwałbym raczej „konstrukcją”, a nie „teorią”* (ibidem, s. 152).

²³ Por. Campbell D.T., *Factors relevant to the validity of experiments in social settings*, „Psychological Bulletin” 1957, t. 54, nr 4, s. 297–312; Campbell D.T., Fiske D.W., op. cit., s. 81–105. Campbell D.T., Stanley J.C., *Experimental and quasi-experimental designs for research on teaching*, [w:] N.L. Gage (red.), *Handbook of research on teaching*, Chicago 1963, s. 171–246; Cook T.D., Campbell D.T., *Quasi-experimentation. Design & analysis issues for field settings*, Boston 1979.

²⁴ Cronbach L.J., *Construct validity...*, op. cit., s. 153.

²⁵ Por. ibidem, s. 156.

²⁶ Por. ibidem, s. 155.

²⁷ Ibidem, s. 163.

rozwiązań. *Musimy to robić, używając obecnych konstrukcji*²⁸, mimo ich problematycznej trafności. Oznacza to, że **walidacja teoretyczna jest procesem** długotrwałym i nade wszystko **społecznym** – zachodzącym w nieustającej dyskusji uczonych, nie zaś jedynie w granicach przedsięwzięcia jednego badacza.

Takie ujęcie problemu badania trafności teoretycznej – które Cronbach określa jako *rozwodnienie racjonalistycznego ideału*²⁹ – nie uchyla wszakże rekomendacji zawartych w tekstach wprowadzających do dyskusji koncepcję walidacji konstruktów. Bo – jak stwierdził Cronbach podczas konferencji *Trafność testu dla lat 90. i dalszych*, która odbyła się w siedzibie Służby Pomiaru Edukacyjnego (Educational Testing Service) w 1986 r. – *nie wszyscy będą zadowoleni z mojej niesubordynacji wobec tradycyjnego podejścia pozytywistycznego [...]. Nie jest jednak tak, iż stare zasady zostały obalone; lecz stare melodie nie brzmią tak samo w uszach podrażnionych przez [...] Kuhna i Toulmina*³⁰.

Stare zasady nie zostały obalone... W tym kontekście Cronbach podkreśla, że *każda zwykła interpretacja ma wiele implikacji. Ustalenie, które z nich są choć w pewnym stopniu niezawodne, i zdefiniowanie granic ich stosowalności mogłoby wymagać nieograniczonych czasu i środków*³¹. Z drugiej strony – zdaniem Cronbacha – człowiek jest skłonny do regulowania swoich spraw na drodze rozsądnego argumentowania opartego na niepełnej informacji. W związku z powyższym Cronbach proponuje rozważenie podejścia do testu i interpretacji jego wyników z wykorzystaniem argumentu trafności (ang. *validity argument*), który powinien wiązać konstrukty, dowody, społeczne i osobiste wartości i konsekwencje interpretacji oraz pełnić funkcję perswazyjną w stosunku do zróżnicowanego pod względem światopoglądowym odbiorcy testu, wyników pomiaru i ich interpretacji.

Pojęcie argumentu walidacyjnego³² Cronbach wykorzysta także – co zostało już wcześniej zasygnalizowane – w *Podstawach psychologicznego testowania*. Jednakże tylko w tym wystąpieniu scharakteryzuje te wąpli-

²⁸ Ibidem.

²⁹ Ibidem, s. 164.

³⁰ Cronbach L.J., *Five perspectives on validity argument*, [w:] H. Wainer, H.I. Braun (red.), *Test validity*, Hillsdale 1988, s. 3.

³¹ Ibidem, s. 4.

³² W innym – zaproponowanym przez Michała Daszkiewicza – przekładzie: dyskursu nad trafnością. Por. Daszkiewicz M., *Znaczenie i zawartość „argumentu” oraz jego rola w ustalaniu trafności i komunikowaniu wyników egzaminu*, [w:] B. Niemierko, J. Syglewicz, W. Walczak (red.), op. cit., s. 70–71.

wości, pytania i zarzuty, których nie można pominąć podczas przygotowywania argumentacji walidacyjnej.

Po pierwsze, osoby nieprzyjaźnie nastawione do interpretacji testu najprawdopodobniej wykorzystają alternatywne hipotezy jako amunicję³³. I dlatego wyrażenie „prawdopodobne alternatywne hipotezy” jest najlepszym drogowskazem podczas silnej (ang. strong) walidacji teoretycznej. Sygnałem do jej rozpoczęcia jest przełożenie wyników testu (czy szerzej: obserwacji, eksperymentu itd.) na słowa³⁴, opisujące zachowanie w terminach ogólnych. Koniecznym krokiem w trakcie badania trafności teoretycznej jest natomiast jak najbardziej precyzyjne sformułowanie zarówno proponowanego znaczenia wyników, jak koncepcji teoretycznej, stanowiącej podstawę tej interpretacji (mimo że – jak twierdzi Cronbach – w najbliższej perspektywie nie ma nadziei na rozwój „sieci nomologicznych”, które kiedyś wzniosłe przewidzieliśmy³⁵), a następnym – zaprojektowanie działań walidacyjnych (w wypadku słabego – ang. weak – programu celem jest zgromadzenie jak największej liczby danych empirycznych, czy to poprzez badanie różnorodnych korelacji, analizę czynnikową, obserwację pracy z testem, czy też studia przypadków, na przykład uczniów osiągających ekstremalne wyniki), które należy rozpocząć, mimo że program badania trafności teoretycznej – silny lub słaby – raczej nie osiągnie finału potrzebnego do obrony testu³⁶.

Po drugie, w polemikach podważana będzie funkcjonalna wartość pomiaru, np. jego dobroczynny wpływ na jakość pracy szkoły czy programów nauczania lub życiowe szanse uczniów nim objętych. Dlatego w programie walidacji należy uwzględnić sprawdzenie konsekwencji społecznych zastosowania testu³⁷, szczególnie wtedy, kiedy na podstawie wyników rozdzielane będą takie dobra, jak dostęp do określonego rodzaju edukacji czy zasobów.

Ponadto zarówno test, jak sposób interpretacji jego wyników mogą stać się przedmiotem polemicznej analizy przeprowadzonej przez ekspertów specjalizujących się w problematyce trafności treściowej (np. reprezentujących stanowisko operacjonalistyczne) czy kryterialnej (np. posiadających doświadczenie w stosowaniu psychometrycznych procedur doboru personelu), a także krytyki ze strony decydentów politycznych różnych szczebli (czy środowisk opozycyjnych), dysponujących swoistymi

³³ Cronbach L.J., *Five perspectives on validity...*, op. cit., s. 14.

³⁴ Por. ibidem, s. 13.

³⁵ Ibidem, s. 14.

³⁶ Ibidem.

³⁷ Por. ibidem, s. 6.

wyobrażeniami na temat sposobów konstruowania i stosowania testów oraz zmierzających do realizacji własnych celów.

Powyższe – i inne – wyzwania powodują, że – zdaniem Cronbacha – *walidacja „stawia na głowie cały twój umysł – a wszystkie chwytły dozwolone”*³⁸. Jednak to nie problem poglądów i oczekiwań członków społeczności, do których skierowany jest argument walidacyjny, stanowi zasadniczą trudność, z którą uporać się musi walidator. Podstawowym zadaniem pozostaje wciąż ustalenie trafności teoretycznej znaczenia wyników testu. W tej materii *walidatorzy powinni rzeczywiście wykonać to, co niezależny uczony – jak najlepiej potrafią, i zazwyczaj pod presją ograniczonego czasu*³⁹. W tym imperatywie jednak nie sposób nie usłyszeć nuty tragicznej, tym bardziej jeżeli uwzględnimy ówczesny pogląd Cronbacha na nauki społeczne.

Jak to zostało już wcześniej zasygnalizowane, w latach 80. Cronbach z dystansem podchodzi do wizji teorii jako „sieci nomologicznej”, przedstawionej w artykule wprowadzającym pojęcie trafności teoretycznej (ang. *construct validity*)⁴⁰. Co więcej, podczas konferencji *Możliwości dla wiedzy w naukach społecznych (Potentialities for Knowledge in Social Science)*, która odbyła się na Uniwersytecie Chicagowskim w 1983 r., Cronbach stwierdził, iż *w latach 50. podjęto poważne starania w celu opracowania rygorystycznych zasad oceny interpretacji testów psychologicznych [...]. Paul Meehl i ja opisaliśmy sposób, w jaki rygor ów mógłby być osiągnięty w świetle filozofii nauki [...]. Pojęcie trafności teoretycznej stało się popularne i te idee wciąż znajdują swoje zastosowanie w przypadku świeżo upieczonych badaczy. Chociaż formuły w jakiś sposób były i są wciąż użyteczne, ta tendencja przynosi represyjny efekt [...]. Program walidacji jakiegokolwiek konstruktów mógłby wymagać pracy przynajmniej całego życia, dlatego prawie nikt nie próbował go realizować*⁴¹.

Jak zatem wspomniany wyżej „niezależny uczony” miałby przeprowadzić badanie trafności teoretycznej wybranego pojęcia, skoro jednego życia za mało na realizację takiego projektu badawczego? Cronbach odpowiada: *Trwała kooperacja koncentrująca się na jakimś określonym aspekcie osobowości mogłaby tę pracę skrócić, ale taki wysiłek należy do rzadkości*⁴². W związku z tym *postęp wymaga, abyśmy respektowali ubogo*

³⁸ Ibidem, s. 14.

³⁹ Ibidem.

⁴⁰ Por. Cronbach L.J., Meehl P.M., op. cit., s. 290.

⁴¹ Cronbach L.J., *Social inquiry by and for earthlings*, [w:] D.W. Fiske, R.A. Shweder (red.), *Metatheory in social science. Pluralisms and subjectivities*, Chicago 1986, s. 85–86.

⁴² Ibidem, s. 86.

uformowane, a nawet „nietestowalne” idee⁴³. Ponadto – skoro (zdaniem Cronbacha) *jakikolwiek badanie jest przede wszystkim studium przypadku*⁴⁴, które na obecnym etapie rozwoju nauk społecznych może przynieść bądź opis historyczny, bądź charakterystykę pojęcia (konstrukt), a nie zdania ogólne (które ewentualnie złożyłyby się w przyszłości na teorię) – walidacja teoretyczna może dostarczyć jedynie w miarę pełnej relacji ze zdarzeń zaobserwowanych w określonym czasie, specyficznym miejscu i swoistym kontekście w trakcie do pewnego stopnia powtarzalnych procedur badawczych. A zatem badanie trafności teoretycznej interpretacji wyników testu – interpretacji, która (w opinii Cronbacha) choć powinna być zgodna z danymi, *musi opierać się na tym, co Vico nazwał „fantasia” i co Niemcy nazwali „Verstehen”*⁴⁵, i która w akcie komunikacji narażona jest na wynikające z natury języka przekształcenia semantyczne (np. na niezamierzone uogólnienia związane z zastosowaniem w wypowiedzi eliptycznych środków wyrazu) – to w najlepszym razie sekwencja prowadzonych z metodologicznym pietyzmem studiów precyzyjnie zdefiniowanych przypadków. Efekty takich czynności badawczych złożą się oczywiście na postęp naukowy – jednakże rozumiany nie jako mozolne konstruowanie doskonałej teorii, lecz raczej jako wzbogacanie dziedzictwa kultury⁴⁶.

Tej tragicznej perspektywy oglądu epistemologicznej kondycji człowieka (stanowiącej – być może – symptom kryzysu legitymizacji wiedzy) nie znajdziemy we wcześniejszych – powiedzmy – instruktażowych publikacjach Cronbacha, chociażby w przywołanej wcześniej II edycji *Podstaw psychologicznego testowania czy Walidacji miar edukacyjnych (Validation of educational measures)*⁴⁷. Popularyzowany w nich dyskurs stanowił,

⁴³ Ibidem.

⁴⁴ Ibidem, s. 92. Cronbach, wyróżniając trzy strategie badawcze (agregowanie różnorodności w studium statystycznym, studium homogenicznej podklasy sytuacji, studium wielu typów), podkreśla, że *czysto statystyczne studium jest niedosłym studium przypadku* (ibidem, s. 99). W tym kontekście należy przywołać stanowisko Cronbacha w sprawie koncepcji testu, który „mierzy tę samą rzecz” – *ekstremalna homogeniczność nie wydaje mi się ani realistyczna, ani rozsądna [...]*. *Jakakolwiek reakcja afektywna czy czynność intelektualna jest najprawdopodobniej wielokrotnie zdeterminowana* – Cronbach L.J., *Methodological studies – a personal retrospective*, [w:] R.E. Snow, D.E. Wiley (red.), *Improving inquiry in social science. A volume in honor of Lee J. Cronbach*, Hillsdale 1991, s. 387.

⁴⁵ Cronbach, *Social inquiry...*, op. cit., s. 98.

⁴⁶ Por. ibidem, s. 104.

⁴⁷ Cronbach L.J., *Validation of educational measures*, [w:] G.H. Bracht, K.D. Hopkins, J.C. Stanley (red.), *Perspectives in educational and psychological measurement*, Englewood Cliffs 1972, s. 88–102 (pierwodruk: Cronbach L.J., *Validation of educational measures*, [w:] Educational Testing Service, *Proceedings of the 1969 invitational conference on testing problems*, Princeton 1970, s. 35–52).

iż w dziedzinie pomiaru edukacyjnego bada się trafność decyzji podjętej na podstawie wyników testu (np. w procedurach eksperymentalnych czy studiach trafności kryterialnej) oraz trafność opisowych interpretacji wyników testu – zarówno tych opartych na porównaniu czynności wykonywanych podczas rozwiązywania zadań z operacyjnie ujętą dziedziną testowania (określanie trafności treściowej), jak i tych, w których wykorzystywane są konstrukty (badanie trafności teoretycznej). Z tego punktu widzenia, walidacja teoretyczna rozpoczyna się od hipotezy, iż test stanowi miarę określonego konstruktów, który w następnej fazie staje się przedmiotem opisu (lokującego dane pojęcie w sieci relacji, również z terminami empirycznymi), z którego wynikają specyficzne implikacje (i szczególnie użyteczne alternatywne hipotezy dotyczące znaczenia wyników), sprawdzane później w badaniach korelacyjnych, eksperymentalnych i logicznych⁴⁸. *Od badacza oczekuje się, iż po zakończeniu zbierania danych zintegruje on hipotezy i wyniki badań oraz zaoferuje końcowe wnioski na temat solidności interpretacji z użyciem konstruktów [...]*⁴⁹ – konkluduje Cronbach.

Przewartościowanie przywołanych powyżej sądów na temat walidacji teoretycznej – omówione przez Cronbacha m.in. podczas konferencji *Możliwości dla wiedzy w naukach społecznych oraz Trafność testu dla lat 90. i dalszych* – stawia – bez wątpienia – w nowym świetle teorię i praktykę pomiaru edukacyjnego (również tę spod znaku behawiorystycznego operacjonizmu, batalię przeciwko przeżytkom którego prowadził z uporem Cronbach również po wprowadzeniu do dyskursu pomiarowego pojęcia trafności konstruktów⁵⁰). Z tego punktu widzenia – ocena jakości narzędzi pomiarowych i skutków ich zastosowania, np. w celu sprawdzenia tzw. osiągnięć szkolnych lub różnicowania kandydatów w procesie rekrutacji do szkół średnich czy wyższych, staje się zadaniem wręcz tragicznym: koniecznym oraz karkołomnym. Bo skoro **jednego życia za mało** na przeprowadzenie walidacji konstruktów, to co faktycznie stanowi przedmiot pomiaru dydaktycznego⁵¹ – w jego konkretnym zastosowaniu?

⁴⁸ Por. ibidem, s. 98.

⁴⁹ Ibidem, s. 101.

⁵⁰ Por. Niemierko B., *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania*, Warszawa 1990, s. 203.

⁵¹ We współczesnym dyskursie pomiarowym funkcjonują chociażby następujące konstrukty: osiągnięcia szkolne (por. Okoń W., Dembska J., Niemierko B., *Pięcioletni słownik terminów pedagogicznych*, Wrocław 1990, s. 40), wyniki nauczania (por. Niemierko B., *Drogi i bezdroża ewaluacji*, „Kwartalnik pedagogiczny” 1996, nr 3–4, s. 5), czynność (por. ibidem). Ponadprzedmiotowy charakter tych pojęć – tudzież teorii pomiaru dydaktycznego – stanowi jedno z istotnych ich ograniczeń (por. Niemierko B., *Założenia metodologiczne ewaluacji Programu Nowa Matura*, [w:] B. Niemierko, E. Kowalik (red.), *Perspektywy diagnostyki edukacyjnej*, Gdańsk 1998, s. 104).