

Piotr Maciej Skorupiński

Warszawa

Geneza pojęcia trafności teoretycznej

Esej ten poświęcam pamięci Paula Meehla,
który odszedł,
pozostawiwszy bez odpowiedzi list pełen pytań...

Edward L. Thorndike (1864–1949) – uczeń jednego z pionierów-założycieli Amerykańskiego Towarzystwa Psychologicznego (por. Fernberger, 1932: 2): Jamesa M. Cattella (1860–1944), który studiował i przygotowywał rozprawę doktorską pod kierunkiem samego Wilhelma Wundta (1832–1920) w Lipsku (por. O'Donnel, 1985: 31) – w *An introduction to the theory of mental and social measurements* (Thorndike, 1904), pierwszym podręczniku teorii pomiaru w naukach społecznych, mimo że nie operuje jeszcze takimi terminami, jak „trafność” czy „walidacja”, w rozdziale *Jednorodności i skali* objaśnia, jakimi właściwościami charakteryzuje się trafna skala (*valid scale*). Jest ona mianowicie: obiektywna – w takim stopniu, w jakim wszyscy kompetentni obserwatorzy zgadzają się co do jej znaczenia (por. Thorndike, 1913: 11–12); jednorodna – w sensie, iż „użyte jako skala serie faktów powinny być zróżnicowane pod względem ilości tej samej rzeczy lub jakości” (tamże: 13); precyzyjna – a więc konstytuowana przez perfekcyjnie zdefiniowane fakty, ilości i różnice pomiędzy nimi; aplikowalna – czyli przystająca (przykładalna) do przedmiotów pomiaru; a także odniesiona do punktu zerowego rozumianego jako fakt tak właśnie oznaczony. O jakości każdej z powyższych właściwości skali orzekają kompetentni obserwatorzy, stopień trafności skali zatem stanowi funkcję ich liczby, kompetencji i zgodności. Tego wątku jednakże Thorndike nie rozwija, ograniczając się do wskazania dwóch punktów granicznych na kontinuum obiektywizm–subiektywizm skali i przechodząc do technik statystycznych, użytecznych w trakcie analizy zbioru wyników jej zastosowania.

„Trafna skala”... – właściwie stosujemy się do tej konwencji nazewniczej od niedawna. Jeszcze pod koniec lat 50. ubiegłego stulecia Maria Grzywak-Kaczyńska powtórzyła za Józefem Pieterem (1904–1989): „symptomatyczna skala” (Grzywak-Kaczyńska, 1960: 26; Pieter, 1934: 26), a dopiero na początku lat 60. Mieczysław

Choynowski – jako pierwszy – powiedziałby o tej skali, iż jest trafna (Choynowski, 1966: 144, 181–182). I mimo że żadna z tych propozycji terminologicznych nie wydaje się w pełni trafna, jeżeli weźmiemy pod uwagę bogactwo znaczeń wywodzącego się z łacińskiego *validus* (na przykład: zdrowy; mocny, silny; ważny w obliczu prawa; potężny; wpływowy; bogaty; skuteczny – jako określenie lekarstwa; silny, obronny – w odniesieniu do twierdzy; męski; pożywny; uprawniony do czegoś) (Sondel, 1997: 976) angielskiego *valid* (na przykład: prawomocny; ważny – na przykład chrzest; dobrze ugruntowany, uzasadniony, poważny – o teoriach, argumentach, wnioskach, koncepcjach, zasadach; efektywny, trafny – test; silny, potężny; zdrowy) (Gove, 1993: 2529–2530), w latach 70. staropolska „trafność” (na przykład: strzału, sądu, spostrzeżenia, oceny) (Doroszewski, 1967, t. IX: 215) dystansuje inne propozycje terminologiczne w języku rozwijanej w Polsce teorii pomiaru w naukach społecznych (por. Niemierko, 1975: 169–216; Nowakowska, 1975: 46–48; Brzeziński, 1978: 190–196).

W latach 70. ubiegłego wieku w tym języku zadomawia się również inny nowy termin – „trafność teoretyczna” (por. Niemierko, 1975: 212–216; Brzeziński, 1978: 191). I choć jeszcze Maria Nowakowska sięga po takie formy, jak „test trafny względem/dla konstruktów” czy „trafność z konstruktami” (Nowakowska, 1975: 46), także i w tym wypadku przyjmuje się wyrażenie zaproponowane przez Choynowskiego ponad 10 lat wcześniej (Choynowski, 1966) – jako tłumaczenie *construct validity*, terminu kultywowanej w kręgu kultury amerykańskiej teorii pomiaru psychologicznego i edukacyjnego.

Angielski odpowiednik „trafności teoretycznej” wprowadzony został do języka psychometrii w pierwszej połowie lat 50. XX wieku. Wtedy bowiem w „*Psychological Bulletin*” – periodyku Amerykańskiego Towarzystwa Psychologicznego (*American Psychological Association*) – opublikowane zostały dwa kluczowe teksty: w 1954 r. dzieło zbiorowe pt. *Technical recommendations for psychological tests and diagnostic techniques* (*American Psychological...*, 1954: 1–38), a w roku 1955 *Construct validity in psychological tests* (Cronbach, Meehl, 1955: 281–302) – artykuł Paula M. Meehla (1920–2003) i Lee J. Cronbacha (1916–2001).

Technical recommendations... to – jak się później okazało – pierwsza z serii publikacji, znanej polskiemu czytelnikowi dzięki – między innymi – przygotowanemu przez Elżbietę Hornowską i Jerzego Brzezińskiego jednemu przekładowi amerykańskich *Standardów dla testów stosowanych w psychologii i pedagogice* (1985) z roku 1974. Stanowią one zbiór 165 opatrzonego komentarzem stwierdzeń – wcześniej zaproponowanych przez siedmioosobową komisję (*Committee on Test Standards*) Amerykańskiego Towarzystwa Psychologicznego, której przewodniczył Cronbach, a następnie zmodyfikowanych najpierw we współpracy z trzyosobowymi komisjami reprezentującymi Amerykańskie Towarzystwo Badań Edukacyjnych (*American Educational Research Association*) i Narodową Radę ds. Pomiarów Edukacyjnych (*National Council on Measurements Used in Education*), a potem w trakcie konsultacji społecznej – rekomendujących określone metodologiczne rozwiązania w dziedzinie informacji towarzyszących szeroko rozumianemu testowi, zarówno psychologicznemu, jak i edukacyjnemu. W dziale poświęconym informacjom dotyczącym „stopnia, w jakim test

umożliwia osiągnięcie określonych celów” (American Psychological..., 1954: 13), obok trzech znanych wcześniej aspektów trafności (*aspects of validity*) – trafności treściowej (*content validity*), trafności prognostycznej (*predictive validity*) i trafności diagnostycznej (*concurrent validity*) – wyodrębniony został typ (*type*) nowy: trafność teoretyczna (*construct validity*) (por. tamże).

W świetle *Technical recommendations*, w podręczniku (*manual*), który powinien towarzyszyć publikacji testu (rekomendacja A.1. – por. tamże: 8), wśród informacji dotyczących zalecanych zastosowań testu i umożliwiających jego użytkownikowi dokonanie prawidłowej interpretacji wyników testu, powinny zostać przedstawione wszystkie dostępne informacje wskazujące na te hipotetyczne psychologiczne właściwości, które znajdują swoje odbicie w wariancji wyników testowania (rekomendacja C.18 – por. tamże: 27). Bardzo pożądane byłoby również przywołanie teorii, którą przyjęto jako podstawę testu, i wyników badań stanowiących jej fundament (rekomendacja C.19 – por. tamże: 28).

Te dwa rekomendowane standardy informowania na temat testu poprzedzone zostały komentarzem wprowadzającym w pojęcie trafności teoretycznej. Rozpoczyna się on stwierdzeniem, iż:

Ocena trafności teoretycznej dokonywana jest w trakcie poszukiwania tych jakości psychologicznych, które mierzone są za pomocą testu, to znaczy na drodze demonstrowania związku pomiędzy określonymi pojęciami (*constructs*) eksplikatywnymi a wykonywaniem danego testu (tamże: 14).

W związku z powyższym, w logicznych i empirycznych „studiach nad trafnością teoretyczną przeprowadzamy walidację teorii leżącej u podłoża testu” (tamże). Informacje dotyczące wyników badania tego aspektu trafności – na przykład eksperymentu czy korelowania wyników pomiaru z miarami danej i innych własności – mają znaczenie dla użytkownika testu w procesie formułowania sądu określającego stopień, w jakim osoba testowana charakteryzuje się właściwościami desygnowanymi przez wspomniane pojęcia.

W komentarzu wprowadzającym w pojęcie trafności teoretycznej brak jednak precyzyjnej definicji nowego terminu. Autorzy dokumentu mówią co prawda o typach trafności, lecz jednocześnie posługują się takimi wyrażeniami, jak „trafność testu” (*validity of a test*) i „trafność inferencji” (*validity of an inference*) czy „trafność interpretacji” (*validity of an interpretation*), zaciemniając tym samym definicyjny kontekst samego pojęcia trafności.

Z niejasnością towarzyszącą introdukcji terminu „trafność teoretyczna”, stanowiącą – zgodnie z wyważoną uwagą Cronbacha i Meehla (1955: 281) – konsekwencję zbiorowego autorstwa dokumentu, usiłowali uporać się dwaj członkowie wspomnianej wcześniej komisji Amerykańskiego Towarzystwa Psychologicznego: jej przewodniczący, a także jeden z członków dwuosobowej podkomisji (w składzie: Meehl i R. C. Challman), w trakcie prac której nietradycyjny sposób myślenia o trafności został po raz pierwszy wyrażony (por. tamże). W klasycznym już artykule *Construct validity in psychological tests*, po partiach wstępnych, rekapitułujących własności

przedstawionych w *Technical recommendations* kategorii procedur walidacyjnych i akcentujących trudności wiążące się z opartym o dostępne kryterium badaniem trafności (*criterion-oriented validation procedure*), Cronbach i Meehl podejmują próbę objaśnienia podstawowych pojęć konstytuujących – ich zdaniem – naukowe podejście do problemu trafności.

„Konstrukt” (*construct*), rozumiany jako pewna postulowana własność hipotetycznie odbita w wynikach testowania (por. Cronbach, Meehl, 1955: 281), i „sieć nomologiczna” (*nomological network*), określona jako system praw wiążących dane empiryczne i pojęcia teoretyczne, którego integralnym elementem staje się w trakcie badania trafności teoretycznej ów wyodrębniony konstrukt (por. tamże: 290), składają się na takie imperatywy metodologiczne, jak chociażby:

Aby poddać walidacji roszczenie, iż dany test mierzy pewien konstrukt, musi istnieć jakakolwiek sieć nomologiczna otaczająca ten koncept [...]. Kiedy badanie postępuje, konstrukt ten wypuszcza w wielu kierunkach korzenie, które przytwierdzają go do coraz większej ilości faktów i innych konstruktów (tamże: 291)

czy

Użytkownik testu, który odrzuca teorię konstruktora testu, [...] musi sam przeprowadzić walidację testu, jeżeli zamierza pokazać, że zdefiniowany przez niego konstrukt jest w teście reprezentowany (tamże).

Dociekania, których celem jest sformułowanie i sprawdzenie hipotezy określającej trafność teoretyczną testu, rozpoczynają się zatem od analizy teorii stanowiącej podstawę interpretacji wyników pomiaru i obejmują różnorakie rodzaje danych zebranych za pomocą różnorodnych metod, zaś ich efekt nie daje się wyrazić w formie jakiegos jednego liczbowego współczynnika trafności. Co więcej, ich przebieg regulowany jest rozstrzygnięciami przyjętymi w ramach metodologii nauki.

Postulat naukowego charakteru walidacji testu stanowi wyraz specyficznej właściwości artykułu Meehla i Cronbacha. Za sprawą Meehla (co zaznaczył Cronbach (1988: 4) podczas konferencji na temat trafności, która odbyła się w siedzibie Służby Pomiaru Edukacyjnego – *Educational Testing Service* – wiosną 1986 r., i co podkreślił podczas tego samego spotkania William H. Angoff, (1988: 25)), który zajmował się tą problematyką we współpracy z Herbertem Feiglem (1902–1988), Michaellem Scrivenem i Wilfridem S. Sellarsem (1912–1989) z Centrum Filozofii Nauki w Minnesocie (*Minnesota Center for Philosophy of Science*) (por. przyp. 1 w Cronbach, Meehl, 1955: 281), w *Construct validity in psychological tests* został przerzucony bowiem pomost pomiędzy psychometrią i neopozytywistyczną filozofią nauki. Klimat filozoficzny lat 50. sprzyjał temu jak najbardziej – o ile jeszcze w latach 30., w złotym wieku amerykańskiej filozofii (*the golden age of the American philosophy*), „wielki posturą Neurath” zabiegał o względy „kruchej budowy, niewielkiego wzrostu” Johna Dewey’a (1859–1952), to już parę lat po II wojnie światowej uczniowie

Charlesa S. Peirce'a (1839–1914), George'a Santayany (1863–1952) czy Dewey'a właśnie „z trudem będą [...] zabiegać o miejsca na uniwersyteckich katedrach filozofii, zajętych przez kolegów Neuratha oraz ich następców” (Wilkoszewska, 1999: 36).

Otto Neurath (1882–1945) – po tragicznej śmierci Moritza Schlicka (1882–1936), który w 1922 r. objął po Ernście Machu (1838–1916) Katedrę Historii i Filozofii Nauk Indukcyjnych na Uniwersytecie Wiedeńskim, przywódca (por. Ayer, 2000: 17) Koła Wiedeńskiego (1925–1938), którego członkami byli m.in. wcześniej wspomniany Feigl i Rudolf Carnap (1891–1970) – w czasie wojny schronił się co prawda nie w Ameryce, lecz w Oksfordzie, za to Feigl, Carnap i inni członkowie i sympatycy Koła „znaleźli [...] sobie posady” (tamże) przede wszystkim w Stanach Zjednoczonych. Wśród tych ostatnich nie można pominąć Carla G. Hempla (1905–1997) z tzw. grupy berlińskiej, do którego publikacji z początku lat 50. odwołują się Cronbach i Meehl, przywołując je – obok artykułów Carnapa i Feigla – jako źródła ich filozofii nauki.

Jedną z kluczowych pozycji bibliograficznych, wskazanych w *Construct validity in psychological tests*, jako ważkie dla zrozumienia kontekstu filozoficznego stanowiska autorów, jest przygotowana przez Feigla i M. Brodbecka i opublikowana w 1953 r. antologia pt. *Readings in the philosophy of science* (Feigl, Brodbeck, 1953). Feigl otwiera ją własnym esejem poświęconym pożądanym właściwościom wiedzy osiągniętej za pomocą metody naukowej, do których zalicza intersubiektywną sprawdzalność, rzetelność, precyzję, wewnętrzną koherencję i otwartość (por. Feigl, 1953: 11–13). Wśród następujących potem przedruków tekstów Carnapa, Hansa Reichenbacha (1891–1953) – pod wpływem koncepcji którego znajdował się w tamtym czasie Meehl – Hempla, Schlicka, Macha, a nawet Alberta Einsteina, przypomniany został znaczący w owym czasie artykuł Kennetha MacCorquodale'a i Meehla – *On a distinction between hypothetical constructs and intervening variables* (pierwodruk: MacCorquodale, Meehl, 1948, t. 55: 95–107) – wprowadzający do słownika psychologii termin „konstrukt”.

W polskim piśmiennictwie naukowym artykuł ten został przynajmniej dwukrotnie scharakteryzowany – najpierw przez Stefana Nowaka (1925–1989) w studium pt. *Obserwacja i rozumienie ludzkich zachowań a problemy budowy teorii* (Nowak, 1965: 205–207), natomiast ostatnio przez Ryszarda Stachowskiego w jego *Historii współczesnej myśli psychologicznej* (Stachowski, 2000: 282). W tym miejscu zostanie zaakcentowane jedynie to, na co zwrócił uwagę Meehl, a mianowicie: w czasie, kiedy ten artykuł dotarł do czytelników „Psychological review”, w amerykańskiej psychologii panował rygorystyczny behawiorystyczny operacjonizm, postrzegany przez Meehla jako upraszczający i zwodniczy¹. Wprowadzenie terminu *construct* – definiowanego w słowniku Webstera jako wywodzący się od łacińskiego *construere*.

¹ Informacja uzyskana przez autora pracy od Paula Meehla w korespondencji ze stycznia 2003 r.

czasownik o znaczeniu zbudować (np. autostradę, zdanie) i narysować (figurę geometryczną) oraz – wtórnie – jako rzeczownik oznaczający intelektualną konstrukcję, ukonstytuowany w procesie mentalnej syntezy przedmiot myśli (por. Gove, 1993: 489) – nie było w tym kontekście jedynie propozycją nowej konwencji nazewnictwa. Określenie, iż „konstrukt” miały odsyłać do nieredukowalnego, do terminów empirycznych konceptu teoretycznego, a zatem odnosić się do procesów i bytów bezpośrednio nieobserwowalnych (por. MacCorquodale, Meehl, 1948: 104), komplikowało – między innymi – problematykę walidacji.

Inspiratorem i znawcą tej komplikacji w dziedzinie pomiaru edukacyjnego stanie się w następnych latach Cronbach. Już w zmienionej III edycji *Encyclopedia of educational research*, pomnikowego projektu Amerykańskiego Towarzystwa Badań Edukacyjnych, nad którą prace prowadzone były w drugiej połowie lat 50., to on właśnie w artykule poświęconym trafności, wyróżniając trzy typy strategii walidacyjnych, jedną z nich uczyni walidację konstruktów (albo raczej – za Choynowskim – walidację teoretyczną – czyli *construct validation*), która – zdaniem Cronbacha – rozpoczyna się od precyzyjnego określenia znaczenia konstruktów, czyli zdefiniowania pojęcia teoretycznego (na przykład inteligencji), stanowiącego cel pomiaru, a zmierza do rozpoznania relacji pomiędzy testem a teorią psychologiczną czy pedagogiczną (por. Cronbach, 1960: 1554).

To także Cronbach – w II edycji *Educational measurement* – przedstawi walidację interpretacji wyników zastosowania każdej procedury zbierania danych, w tym również pomiaru edukacyjnego, jako proces badania trafności całej teorii wiążącej wykorzystane w interpretacji konstrukty, przywołując przy tym pouczający przypadek walidacji pojęcia inteligencji ogólnej jako integracji wyników wszelkich studiów, których przedmiotem była trafność Stanfordzkiej Skali Bineta jako testu w ten sposób rozumianej inteligencji (por. Cronbach, 1971: 464 i 476).

I ta właśnie, zaiste filozoficzna (czy wręcz ideologiczna) wytrwałość Cronbacha w podejmowaniu tematu trafności teoretycznej, okazała się brzemienne w skutki, począwszy od oficjalnego stanowiska w sprawie walidacji trzech liczących się towarzystw naukowych – Narodowej Rady ds. Pomiarów Edukacyjnych (NCME), Amerykańskiego Towarzystwa Badań Edukacyjnych (AERA) i Amerykańskiego Towarzystwa Psychologicznego (APA) – wyrażanego w kolejnych edycjach *Standardów dla testów stosowanych w psychologii i pedagogice*, a nie kończąc na zasadniczym przewartościowaniu roli problematyki trafności teoretycznej w układzie treści podręczników pomiaru edukacyjnego, czego przykładem jest chociażby *Standardized achievement testing* Thomasa M. Haladyny, który oscyluje w kierunku wykorzystania trafności teoretycznej jako dominanty kompozycyjnej (por. Haladyna, 2002: 210). Ale to już nie za sprawą samego Cronbacha, lecz jego sukcesora w dziedzinie teorii walidacji, wprowadzającego kontrowersyjne zmiany w pojęciu trafności teoretycznej – Samuela Messicka (1931–1998).

Bibliografia

- Angoff W. H. (1988), *Validity: an Evolving Concept*, [w:] H. Wainer, H. I. Braun (red.), *Test Validity*, Hillsdale, s. 19–32.
- American Psychological Association, American Educational Research Association i National Council on Measurements Used in Education (1954), *Technical Recommendations for Psychological Tests and Diagnostic Techniques*, „Psychological Bulletin”, t. 51, nr 2 (suplement), s. 1–38 (201–238).
- Ayer A. J. (2000), *Filozofia w XX wieku*, Warszawa.
- Brzeziński J. (1978), *Elementy metodologii badań psychologicznych*, Warszawa.
- Choynowski M. (1966), *Opracowanie zestawu testów do prognozy powodzenia w podoficerskich szkołach Milicji Obywatelskiej*, Wrocław.
- Cronbach L. J. (1960), *Validity*, [w:] C. W. Harris (red.), *Encyclopedia of Educational Research*, New York, s. 1551–1555.
- Cronbach L. J. (1971), *Test Validation*, [w:] R. L. Thorndike (red.), *Educational Measurement*, II wyd., Washington, s. 443–507.
- Cronbach L. J. (1988), *Five Perspectives on the Validity Argument*, [w:] H. Wainer, H. I. Braun (red.), *Test Validity*, Hillsdale, s. 3–17.
- Cronbach L. J., Meehl P. M. (1955), *Construct Validity in Psychological Tests*, „Psychological Bulletin”, t. 52, nr 4, s. 281–302.
- Doroszewski W. (red.) (1967), *Słownik języka polskiego*, t. IX, Warszawa.
- Feigl H. (1953), *The Scientific Outlook: Naturalism and Humanism*, [w:] H. Feigl, M. Brodbeck (red.), *Readings in the Philosophy of Science*, New York, s. 8–18.
- Feigl H., Brodbeck M. (red.) (1953), *Readings in the Philosophy of Science*, New York.
- Fernberger S. W. (1932), *The American Psychological Association: A Historical Summary, 1892–1930*, „Psychological Bulletin”, t. 29, s. 1–89.
- Gove P. B. (red.) (1993), *Webster's Third New International Dictionary of the English Language Unabridged*, Springfield.
- Grzywak-Kaczyńska M. (1960), *Testy w szkole*, Warszawa.
- Harris C. W. (red.) (1960), *Encyclopedia of Educational Research*, New York.
- Komendziński T., Szahaj A. (red.) (1999), *Filozofia amerykańska dzisiaj*, Toruń.
- Haladyna T. M. (2002), *Standardized Achievement Testing. Validity and Accountability*, Boston.
- MacCorquodale K., Meehl P. E. (1948), *On a Distinction Between Hypothetical Constructs and Intervening Variables*, „Psychological Review”, t. 55, s. 95–107.
- Niemierko B. (1975), *Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe*, Warszawa.
- Nowak S. (1965), *Obserwacja i rozumienie ludzkich zachowań a problemy budowy teorii*, [w:] S. Nowak, *Studia z metodologii nauk społecznych*, Warszawa, s. 183–244.
- Nowak S. (1965), *Studia z metodologii nauk społecznych*, Warszawa.

- Nowakowska M. (1975), *Psychologia ilościowa z elementami naukometrii*, Warszawa.
- O'Donnel J. M. (1985), *The Origins of Behaviorism. American Psychology, 1870–1920*, New York.
- Pieter J. (1934), *Nowe sposoby egzaminowania. Test wiadomości, jego założenia, konstrukcja i stosowanie*, wyd. I, Warszawa.
- Sondel J. (1997), *Słownik łacińsko-polski dla prawników i historyków*, Kraków.
- Stachowski R. (2000), *Historia współczesnej myśli psychologicznej. Od Wundta do czasów najnowszych*, Warszawa.
- Standardy dla testów stosowanych w psychologii i pedagogice* (1985), Warszawa.
- Thorndike E. L. (1904), *An Introduction to the Theory of Mental and Social Measurements*, New York.
- Thorndike E. L. (1913), *An Introduction to the Theory of Mental and Social Measurements*, wyd. II, New York.
- Thorndike R. L. (red.) (1971), *Educational Measurement*, wyd. II, Washington.
- Wainer H., Braun H. I. (red.) (1988), *Test Validity*, Hillsdale.
- Wilkożewska K. (1999), *Filozofia amerykańska w Ameryce. Oczami outsidera*, [w:] T. Komendziński, A. Szahaj (red.), *Filozofia amerykańska dzisiaj*, Toruń, s. 35–51.