

Kiedy odwołując się do testów psychologicznych postępujemy nieetycznie? Analiza kontekstu psychometrycznego

Jerzy M. Brzeziński*

Instytut Psychologii, Uniwersytet im. Adama Mickiewicza, Poznań

WHEN DO WE ACT UNETHICALLY RESORTING TO THE RESULTS OF PSYCHOLOGICAL TESTS? THE ANALYSIS OF PSYCHOMETRIC CONTEXT

One of the characteristic features of research and diagnostic practice in psychology is resorting to specific instruments – psychological tests. Unfortunately, apart from fully qualified researchers and practitioners (M.A. or Ph.D. degree in psychology) such tests are also conducted by persons who can use them only at a very superficial level. The author of the article is not so much interested in the ethical context of applying psychological tests which derives from respecting the rights of individuals to privacy and confidentiality or respecting the rule of informed consent to research as in ethical consequences of dissociating the test from the psychological theory (the consequence – the absence of construct validity) and omitting – at the construction stage and scores interpretation stage – the principles of test theories (e.g. true score theory or generalizability theory). In the author's opinion the most serious ethical consequences stem from incorrectly established validity and reliability of psychological testing and omitting the standard error of measurement (SEM) value in the quantitative interpretation of test scores as well as not using the confidence intervals method while analysing test results.

Już ponad stuletnia historia testów inteligencji¹ pozwala na dokonanie oceny ich użyteczności, a także wskazania możliwych do popełnienia nadużyć, których źródłem jest sam test. Obok testów inteligencji i uzdolnień, w praktyce badawczej, diagnostycznej i eksperckiej znajdują zastosowanie także **kwestionariusze osobowości** oraz tzw. **techniki projekcyjne** (tu na plan pierwszy wysuwają się 3 najbardziej rozpowszechnione w świecie narzędzia: *Test Rorschacha*, *TAT* i *Test Zdań Niedokończonych*²).

Źródłami nadużyć są:

1. Wadliwie skonstruowany test oraz niepoprawnie obliczone i nietrafnie zinterpretowane jego wyniki;
2. Wadliwe (nietrafne) decyzje, podjęte odnośnie do losów osób poddanych badaniom testowym, podjęte przez specjalistę (tak naprawdę niedo-

statecznie przygotowanego do profesjonalnego posługiwania się testem psychologicznym), który na podstawie wyników badania testo-

¹ Przypomnijmy, że zapoczątkowała ją słynna *Skala Inteligencji* stworzona w 1905 r. przez Alfreda Bineta i Theodore'a Simona (por. Binet i Simon, 1905a, 1905b, 1905c; jej najnowsza wersja to: *Stanford-Binet Intelligence Scales. Fifth Edition, S-B 5* – por. Roid, 2003a, 2003b, 2003c) oraz, wychodząca z innych założeń konstrukcyjnych Skala Inteligencji opracowana przez Davida Wechslera (jej najnowsza wersja to: *Wechsler Adult Intelligence Scale. Fourth Edition, WAIS-IV* – por. Wechsler, 2008) – por. Boake, 2002; także: Brzeziński, 2003).

² W przeprowadzonej, wśród amerykańskich psychologów klinicznych ($n=412$), w pierwszej połowie lat 90. ubiegłego wieku ankiecie (por. Watkins, Jr., Campbell, Nieberding i Hallmark, 1995) te techniki znalazły się wśród 7 najbardziej popularnych narzędzi diagnozy psychologicznej (wśród nich znalazła się także *Skala Inteligencji D. Wechslera* – dla dzieci i dla dorosłych oraz kwestionariusz MMPI-2). Mimo znacznej popularności techniki projekcyjne budzą różnorakie kontrowersje (o czym w: Lilienfeld, Wood i Garb, 2002; Garb, Wood, Lilienfeld i Nezworski, 2002; ten problem został też podniesiony przez Paluchowskiego i Stemplewską-Żakowicz, 2008, a także przez Stemplewską-Żakowicz, 2004, 2009).

* Korespondencję dotyczącą artykułu można kierować na adres: Jerzy Brzeziński, Instytut Psychologii, Uniwersytet im. Adama Mickiewicza, ul. Szamarzewskiego 89, 60-568 Poznań.
e-mail: brzezuam@amu.edu.pl

wego podejmuje wobec tych osób określone działania modyfikujące ich zachowania (np. działania terapeutyczne czy selekcyjne).

3. Wadliwe (nierzetelne i nietrafne) rezultaty badań naukowych w których operacjonalizacja zmiennych została przeprowadzona za pomocą wadliwych testów psychologicznych oraz „ni-by-testów”.

W niniejszym opracowaniu analizie będzie poddana pierwsza kategoria nadużyć, którą można podzielić na dwie subkategorie: a) związane z niepoprawnością konstrukcji testu oraz b) związane z błędną interpretacją wyników badania testowego. Dwie pozostałe kategorie (druga i trzecia) są pochodnymi względem tej pierwszej.

Zanim jednak poddamy analizie etyczność praktyki testowania (ograniczając ją, zgodnie z zapowiedzią, do kontekstu konstruowania testów i interpretowania ich wyników), to wpieryw odpowiadamy na, jak się wydaje, podstawowe pytanie: Kiedy osoba uprawniona do posługiwania się testami psychologicznymi (rozstrzygnięcie: kto jest uprawniony do posługiwania się testami wcale nie jest oczywiste) odwołuje się w swojej pracy do wyników testów psychologicznych? Moim zdaniem, można wskazać na trzy typowe sytuacje w których sięga się (zakładam, w trybie idealizującym, że w sposób przemyślany) po test psychologiczny:

1. **Badania naukowe (podstawowe i stosowane)** – testy dostarczają wyników, które wkomponowane w większą całość pozwalają na empiryczne sprawdzanie hipotez oraz teorii. Z kolei empirycznie sprawdzone teorie pozwalają osobom działającym w sferze praktyki społecznej nie tylko lepiej zrozumieć określony wycinek rzeczywistości, ale – co może najistotniejsze – pozwalają na optymalizację podejmowanych przez specjalistów działań praktycznych (np. psychoterapii). Można tu mówić o sekwencji trzech kroków: (I) teoria psychologiczna (empirycznie sprawdzona, spełniająca standardy metodologiczne) → (II) test psychologiczny (spełniający standardy psychometryczne) → (III) praktyka (diagnostyczna i badawcza – efektywna i etyczna).
2. **Badania kwalifikacyjne i selekcyjne** – testy psychologiczne stosowane są przez różne instytucje w celu trafnego doboru osób, które mają wykonywać niekiedy bardzo złożone zadania wymagające unikatowych uzdolnień (np. pilot), doświadczenia (np. menadżer), umiejętności (np. nauczyciel), a także braku patologicznych cech osobowości (np. nadmier-

na agresywność). Z pomocą specjalistycznych testów psychologicznych (na ogół wspartych innymi narzędziami) możliwe jest zoptymalizowanie pracy jakiejś instytucji.

3. **Badania diagnostyczne (indywidualne)** – testy psychologiczne stosowane są, aby wobec osoby oczekującej specjalistycznej pomocy zastosowano optymalne postępowanie pomocowe (np. psychoterapię wobec osób cierpiących na zaburzenia czy choroby psychiczne, albo znajdujące się, aktualnie, w trudnej sytuacji życiowej z którą same nie są w stanie sobie poradzić), albo w celu podjęcia ważnej życiowo decyzji (np. wybór szkoły dla dziecka, wybór kariery zawodowej itp.). Stosowane wobec indywidualnych osób (wyłącznie za ich zgodą lub zgodą ich prawnych opiekunów) testy psychologiczne mogą mieć pośredni wpływ na poprawę ich jakości życia. Specjalną kategorią takich badań są te, wykonywane przez powołanych przez sąd biegłych (psychologów i psychiatrów) w celu orzekania o odpowiedzialności za popełniony, naruszający obowiązujące prawo, czyn. Warto zauważyć, że w pewnych sytuacjach – np. definiowanych przez *Kodeks postępowania karnego* – osoby mające prawny status osoby podejrzanej czy oskarżonej muszą się poddać odpowiednim badaniom psychologicznym czy lekarskim. W tych pierwszych mieszczą się też badania psychologiczne przeprowadzane za pomocą testów inteligencji czy osobowości³. Do tej kategorii należy też zaliczyć badania wykonywane za pomocą testów psychologicznych, które zlecane są ekspertom przez firmy ubezpieczeniowe i instytucje przyznające prawa do renty inwalidzkiej (np. w Polsce Zakład Ubezpieczeń Społecznych) z tytułu określonej niepełnosprawności (np. upośledzenia umysłowego – osoba ubiegająca się o takie świadczenie musi, między innymi, uzyskać w badaniu przeprowadzonym za pomocą odpowiedniego testu inteligencji wartość IQ mieszczącą się poniżej pewnej wartości progowej).

³ Mówi o tym art. 74 § 2 ust. 2 *Kodeksu postępowania karnego*: „[...] § 2. Oskarżony jest jednak obowiązany poddać się: [...] 2) badaniom psychologicznym i psychiatrycznym oraz badaniom połączonym z dokonaniem zabiegów na jego ciele, z wyjątkiem chirurgicznych, pod warunkiem że dokonywane są przez uprawnionego do tego pracownika służby zdrowia z zachowaniem wskazań wiedzy lekarskiej i nie zagrażają zdrowiu oskarżonego, jeżeli przeprowadzenie tych badań jest nieodzowne; w szczególności oskarżony jest obowiązany przy zachowaniu tych warunków poddać się pobraniu krwi, włosów lub wydzielin organizmu, z zastrzeżeniem pkt 3. [...]” (podkr. JB).

Pomijam tu – ze względów oczywistych – sytuację dość nietypową; oto jakaś osoba uważa, że może dokonać **autodiagnozy** (np. swojej inteligencji czy osobowości) i sama przebadana się za pomocą dostępnych jej (też w drodze wyłudzenia) testów psychologicznych. Dość zauważyć, że test psychologiczny to „coś” więcej, aniżeli lista kilkudziesięciu pytań (to: standardowa postać kwestionariusza osobowości, np. rozpowszechniona w świecie *Skala Temperamentu* autorstwa polskiego psychologa Jana Strelaua), zestaw zadań arytmetycznych, obrazków, ciągów liczbowych i łamigłówek (to z kolei: standardowa postać złożonych testów inteligencji, np. powszechnie stosowana w świecie, też i w Polsce, *Skala Inteligencji* skonstruowana przez amerykańskiego psychologa Davida Wechslera) czy zestaw plansz z rozmazanymi plamami atramentowymi (to: powszechnie w świecie stosowany test projekcyjny autorstwa szwajcarskiego psychiatry Hermanna Rorschacha)? Jeżeli testy psychologiczne są – jak się wydaje – tak proste i umożliwiają dokonanie wglądu w osobowość czy inteligencję danej, badanej nimi osoby, to dlaczego ograniczać dostęp do nich? Może słuszne jest drukowanie takich, niby-testów w popularnych, kolorowych magazynach? Może, w ten prosty sposób, trafią one pod strzechy i przed telewizory i przyczynią się do wzrostu psychologicznej samowiedzy. Może wystarczy opatrzyć taki „test” prostym językiem napisanym psychologicznym komentarzem. Umożliwimy w ten sposób osobie rozwiązującej, w zaciszu domowym, „test” dokonać autodiagnozy jej umiejętności, preferencji estetycznych, inteligencji, dojrzałości małżeńskiej, zdolności przywódczych, a nawet wyboru ścieżki kariery zawodowej. Czy, mimo tych ułatwień, można zawierzyć tak prostej procedurze? Nie, nie można. No bo czy poznanie struktury czegoś tak tajemniczego i bardzo, bardzo złożonego jak nasza osobowość czy inteligencja, albo poznanie predyspozycji do wykonywania, czasami w dość odległej przyszłości, jakiegoś zawodu nie wymagają szacunku i troski o profesjonalny poziom odpowiedzi na nurtujące nas pytania?⁴ Psychologowie, zwłaszcza oni (!), nie powinni też zachęcać, na poważnie, do stawiania auto-diagnoz. Wręcz przeciwnie, powinni do tego zniechęcać.

KTO JEST UPRAWNIONY DO STOSOWANIA TESTÓW PSYCHOLOGICZNYCH?

Sformułowane w tytule tej części artykułu pytanie nurtowało fachowców od testów psychologicznych i ich – co chciałbym podkreślić – odpowiedzialnych

zastosowań. Test psychologiczny jest wprawdzie stworzony (przez kogo?), następnie jest rozpowszechniany (przez kogo?), potem jest stosowany i interpretowany (przez kogo?), aż, wreszcie, w oparciu o jego wyniki podejmowane są wobec osób nim badanych określone decyzje (przez kogo?). Jak piszą autorzy *Standardów dla testów stosowanych w psychologii i pedagogice* (por. AERA/APA/NCME, 1999/2007, s. 22):

W procesie badania testem bierze udział wielu uczestników, w tym osoby, które: (a) przygotowują i opracowują testy, (b) wydają i dystrybuują testy, (c) przeprowadzają badania testami i oceniają ich wyniki, (d) wykorzystują wyniki testowania do podejmowania decyzji, (e) interpretują wyniki testowe dla swoich klientów, (f) decydują się na badanie testami z wyboru, polecenia lub z konieczności, (g), sponsorują testowanie (reprezentują instytucje czy agencje rządowe i zawierają kontrakt z osobami opracowującymi testy na przygotowanie specyficznego testu lub usługi), (h) dobierają lub recenzują testy z punktu widzenia możliwości ich wykorzystania do określonych celów.

W niniejszym opracowaniu będę się skupiał na osobach wymienionych w punktach: (a), (c) – (f). To ich, przede wszystkim, mają na uwadze autorzy programów kształcenia za zakresu psychometrii i diagnostyki psychologicznej realizowanych na studiach magisterskich z psychologii (por. *Standardy kształcenia dla poszczególnych kierunków studiów i poziomów kształcenia*: http://www.bip.nauka.gov.pl/bipmein/index.jsp?place=Lead07&news_cat_id=117&news_id=982&layout=1&page=text – Załącznik nr 87).

Stosowne definicje i zalecenia zostały zawarte w kilku opracowaniach, których znajomość – co znowu chciałbym podkreślić – jest obowiązkowa dla każdej osoby, która zechce się posłużyć testem psychologicznym w swojej praktyce badawczej, diagnostycznej czy pomocowej (ich tytuły podaję w pkt. 2).

W *Standardach dla testów stosowanych w psychologii i pedagogice* określa się tzw. „użytkownika testu” (*test user*). Jest nim: „osoba, osoby lub agencje odpowiedzialne za wybór i badanie testem, za interpretowanie wyników testowych otrzymanych w określonym kontekście oraz za każdą decyzję czy działanie oparte przynajmniej w części na wynikach testowych” AERA/APA/NCME, 1999/2007, s. 301). Zaczniemy od, jak się zdaje, podstawowego pytania: Jakie właściwości powinny być przypisane owemu „użytkownikowi testu”? Najprostsza odpowiedź na

⁴ Ostrzegałem przed takim postępowaniem czytelników popularnego magazynu *Charaktery* (por. Brzeziński, 2009).

nie brzmi: właściwość kompetencji. Zatem nie każdy (nie każda instytucja), kto (która) sięgnie po test psychologiczny może – nawet po zapoznaniu się z instrukcją „obsługi” – poprawnie i odpowiedzialnie (a więc i etycznie) się nim posłużyć. Czy każdy test wymaga takich samych (w sensie poziomu ich złożoności) kwalifikacji od jego użytkownika? I, jakie to są kwalifikacje?

Ekspersi *American Psychological Association* (APA, 2000) podzielili testy psychologiczne na trzy kategorie – wedle kryterium ich złożoności i wymaganej wiedzy psychologicznej i umiejętności posługiwania się danym testem. Te trzy kategorie określone zostały za pomocą trzech poziomów kwalifikacji ich użytkowników (*test users*) – poziomy: A, B, C:

- **Poziom A:** testy przydatności zawodowej (*vocational proficiency tests*).
- **Poziom B:** testy do grupowego badania inteligencji i kwestionariusze zainteresowań (*general intelligence tests and interest inventories*), np. *Test Matryc Ravena*.
- **Poziom C:** testy do indywidualnego badania inteligencji, testy osobowości i techniki projekcyjne (*individually administered tests of intelligence, personality tests, and projective methods*), np. *Skale Inteligencji Davida Wechslera* – dla dorosłych i dla dzieci (WAIS, WISC, WPPSI), *Stanford-Binet Intelligence Scale: Fifth Edition* (ostatnia rewizja: SB-5), *NEO-FFI*, *MMPI-2*, *Rorschach*, *TAT*, *Test Zdań Niedokończonych*.

Poziom B i C zakłada wykształcenie psychologiczne (co należy odczytać, w przełożeniu na warunki polskie, jako posiadanie tytułu zawodowego „magistra psychologii”, a ponadto – na poziomie C – niezbędne jest doświadczenie praktyczne zdobyte – w zakresie posługiwania się danym testem przez psychologa – pod merytoryczną kontrolą doświadczonego psychologa.

O wprowadzeniu podobnego rozwiązania w Polsce, myśli grupa ekspertów powołana przez największą polską korporację psychologów: *Polskie Towarzystwo Psychologiczne*. Sądzę, że krokiem w dobrym kierunku były, przed laty, prowadzone przez doświadczonych psychologów szkolenia-warsztaty – w zakresie *Skali Inteligencji WAIS-R(PL)* przez Elżbietę Hornowską (z Instytutu Psychologii UAM) oraz w zakresie podejścia Exnera do interpretacji metody Rorschacha przez Michała Stasiakiewicza (z Instytutu Psychologii UAM).

KOMU SŁUŻĄ STANDARDY DLA TWORZENIA I STOSOWANIA TESTÓW PSYCHOLOGICZNYCH?

Konstruktorzy pierwszych testów psychologicznych w zasadzie nie stawiali żadnych warunków wstępnych, odnoszących się do szczególnych kwalifikacji ich użytkowników. Nawet w latach późniejszych, gdy powstały już zręby przyszłej, całościowej teorii testów umysłowych⁵ nie rezerwowano prawa użytkowania testów (nawet tych, tak złożonych interpretacyjnie, jak *Stanford-Binet Intelligence Scale*) dla psychologów. Przykładowo, w Polsce w latach 20. ubiegłego wieku prowadzono – na Uniwersytecie Poznańskim – kursy dla nauczycieli (!) w zakresie stosowania, jak pisano w fachowym kwartalniku *Psychotechnika*, „amerykańskich testów narodowych” (por. Brzeziński, 2008). Specjalnie nie przejęto się uwagami Władysława Witwickiego (1928), który bardzo krytycznie odniósł się do mechanicznego przenoszenia – w odmienne od amerykańskich warunki kulturowe – testów inteligencji stosowanych w selekcji szkolnej. Ta bardzo wnikliwa analiza kulturowej nieprzystawalności tłumaczeń treści zadań testowych do nowego kontekstu kulturowego też miała charakter pionierski i nic dziś nie straciła na swej aktualności. Niestety, nie spotkała się z należytą uwagą. Psychologowie, a zwłaszcza ci pracujący w sferze praktyki (szkoła, poradnie itp.) byli raczej zafascynowani (na co pokazuje treść większości ogłoszonych drukiem w *Psychotechnice* opracowań) nową metodą.

Najważniejszym opracowaniem zawierającym wskazówki dla tych, którzy konstruują, wytwarzają i rozpowszechniają testy oraz tych, którzy posługują się testami są, wyżej już cytowane, *Standardy dla testów...*⁶. Z wydania na wydanie zwiększała

⁵ Ta, jak powszechnie wiadomo, w najpełniejszej postaci (niejako „kanonicznej”) została przedstawiona pod postacią teorii **wyniku prawdziwego** (*theory of true and error scores*) autorstwa Harolda Gulliksen (1950), zwana też **klasyczną teorią testów** (później rozwinięta i zmodyfikowana przez Frederica M. Lorda i Melvina R. Novicka (1968) – zwana **statystyczną teorią testów** (*statistical theory of mental tests*).

⁶ To aktualne, z 1999 roku, szóste już wydanie tego ważnego opracowania po raz pierwszy ukazało się w 1954 roku i nosiło tytuł: *Technical recommendations for psychological tests and diagnostic techniques* (APA, 1954). Po nim przyszły następne wydania (i nie były to tylko proste rozszerzenia tekstu z 1954 roku): 2. wydanie 1955 roku: *Technical recommendations for achievement tests* (AERA/NCME, 1955), 3. wydanie z 1966 roku: *Standards for educational and psychological tests and manuals* (APA/AERA/NCME, 1966), 4. wydanie z 1974 roku: *Standards for educational and psychological tests* (AERA/APA/NCME, 1974/1985) oraz 5. wydanie z 1985 roku: *Standards for educational and psychological testing* (AERA/APA/NVME, 1985).

się objętość *Standardów* i poszerzała się lista podejmowanych tematów. Z dość prostego zestawu „technicznych” zaleceń *Standardy* zaczęły się przeobrażać w swoisty kodeks, którego zakazy / nakazy przestały dotyczyć wyłącznie sfery psychometrycznej. Myślę, że ostatnie wydanie spełnia postulat sformułowany przez S. Messicka (1980/2005), który akcentował potrzebę stworzenia **standardów etycznych**, które znalazłyby zastosowanie w ocenie wykorzystania testu w praktyce diagnostycznej.

Przeciętny psycholog, który nie pracuje na polu problemowym psychometrii, a jedynie stosuje udostępnione mu przez wyspecjalizowane wydawnictwa (ufając, że oferują one profesjonalne usługi) testy może, albo polegać ma wyniesionej ze studiów magisterskich i studiów podyplomowych wiedzy i umiejętnościach w zakresie diagnostycznych zastosowań testów psychologicznych, albo na specjalistycznych wydawnictwach. Na szczęście tych ostatnich nie brakuje. Są one też dostępne polskiemu psychologowi w tłumaczeniu na język polski. Moim zdaniem, każdy psycholog powinien gruntownie zapoznać się z następującymi opracowaniami:

Podstawowe:

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (por. AERA/APA/NCME, 1999/2007). *Standardy dla testów stosowanych w psychologii i pedagogice*. Lektura tego opracowania powinna być obowiązkowa dla studentów psychologii (przedmiot: psychometria).

Pomocnicze:

- American Psychological Association (por. APA, 2000). *Report of the Task Force on Test User Qualifications. Practice and Science Directorates*.
- Joint Committee on Testing Practices (por. JCTP, 1998). *Code of fair testing practices in education*. Washington, DC: APA
- International Test Commission (por. ITC, 2000). *Wytyczne międzynarodowej komisji ds. testów (International Test Commission – ITC) z 2000 roku dotyczące stosowania testów: wersja polska*.
- International Test Commission (por. ITC, 2005). *Międzynarodowe wytyczne dotyczące komputerowych i internetowych badań testowych: wersja polska*.

Standardy stanowią swoisty pomost łączący świat psychologicznych, i pokrewnych, osiągnięć badawczych (tu: teorii psychologicznych, procedur wypracowanych przez psychometrię i statystykę oraz metodologię badań psychologicznych) ze światem praktyki społecznej (tu: terapii, profilaktyki, edukacji, biznesu). Wskazują też, co nadaje im walor kodeksu etycznego, na poprawne i wadliwe postępowania w obszarze konstrukcji i stosowania testów psychologicznych oraz interpretowania otrzymanych wyników. Oczywiście nie zastąpią one ani kursów uniwersyteckich, ani lektury odpowiednich podręczników i monografii, ani też odbytej pod kierunkiem doświadczonego psychologa kursu praktycznego. Jednakże dziś trudno mi sobie wyobrazić profesjonalnego psychologa (takiego, „domowego chowu” mogę sobie wyobrazić, a nawet powiem więcej, nie jest ich wcale tak mało), który nie zaglądałby, od czasu do czasu, do wyżej wymienionych publikacji.

ŹRÓDŁA NADUŻYĆ W BADANIACH TESTOWYCH – TEST I JEGO TWÓRCA I UŻYTKOWNIK

Sygnalizowane na początku artykułu nadużycia związane są z osobą psychologa i ogniskują się, jak się wydaje, na dwóch krytycznych punktach:

- Znajomości **teorii psychologicznych**.
- Znajomości **standardów konstrukcji testu i interpretacji wyniku testowego** – co jest pochodną znajomości odpowiednich **teorii psychometrycznych i statystycznych** oraz nadbudowanych na nich umiejętności posługiwania się stosownymi procedurami.

W dwóch kolejnych punktach rozwinę te dwa, kluczowe dla dobrego-etycznego stosowania testów psychologicznych zagadnienia.

ZNAJOMOŚĆ TEORII PSYCHOLOGICZNYCH

Pisał przed laty François Jacob (1973, s. 32): „w wymianie między teorią a doświadczeniem teoria zawsze zaczyna dialog jako pierwsza”. Każdy też termin wprowadzany do nauki ma wpieryw status **terminu teoretycznego** (czy jak to się zwykło mówić w naukach behawioralnych – **konstruktu teoretycznego**), któremu nadaje się określony sens empiryczny na gruncie określonej teorii należącej do danej dyscypliny naukowej. Filozofowie nauki określali reguły wprowadzania terminów teoretycznych do nauk empirycznych i nadawania mu sensu empirycznego.

W psychologii, w latach trzydziestych ubiegłego wieku, fascynowano się metodologicznym programem **operacjonizmu**, autora głośnej przed laty monografii *The logic of modern physics* Percy W. Bridgmana (1927), którego program metodologiczny na grunt badań psychologicznych usiłowali przenieść behawiorysty (por. np. Stevens, 1935/2000; *Psychological Review*, 1945)⁷. To w „cieniu” operacjonizmu powstawały testy psychologiczne. Mało tego, lekceważąc teorię (B.F. Skinner: „theories are fun”) usiłowano definiować terminy psychologiczne bezpośrednio w kategoriach operacji pomiarowych – np. „inteligencja to to, co mierzy test inteligencji” (!). Jednakże ów teoretyczny program odszedł do przeszłości, a na właściwe miejsce wróciła teoria psychologiczna. Kamieniem milowym stał się głośny, dziś już należący do kanonu lektur psychologicznych tekst Lee J. Cronbacha i Paula E. Meehla (1955/2005) poświęcony tzw. **trafności teoretycznej**. To w nim podkreślono znaczenie teorii psychologicznej w psychometrii. Nie można tedy zajmować się testami psychologicznymi bez znajomości określonych teorii psychologicznych.

ZNAJOMOŚĆ STANDARDÓW KONSTRUKCJI TESTU I INTERPRETACJI WYNIKU TESTOWEGO

Wedle współczesnego poglądu na test psychologiczny jest nim: „narzędzie lub procedura służące ocenie. Ich istotą zaś – otrzymanie w określonych warunkach próbki zachowania osób badanych, a następnie dokonanie ich oceny zgodnie z wystandardyzowanymi regułami” AERA/APA/NCME, 1999/2007, s. 24). Podobną definicję znajdziemy u Roberta J. Gregory’ego (2004, s. 30): „[...] standaryzowana procedura pozyskiwania próbki zachowania i jej opisanie za pomocą kategorii lub wyników”. Jakikolwiek, mniej czy bardziej zaawansowany podręcznik psychometrii, kładzie silny nacisk na: (1) respektowanie założeń (tu: najczęściej wywiedzionych z **teorii wyniku prawdziwego** Harolda Gulliksen’a, 1950) leżących u podstaw narzędzia pomiarowego, które jest identyfikowane właśnie jako „test psychologiczny”, (2) przestrzeganie procedury stosowania testu (spełnienie wymogu standaryzacji) oraz (3) przestrzeganie procedur ilościowej (procedury psychometryczno-statystyczne, np. analiza profilowa czy budowa przedziałów ufności symetrycznych względem wyniku otrzymanego) i merytorycznej (tu: psychologicznej,

z odwołaniem do teorii psychologicznej) interpretacji wyniku testowego.

Nie wchodząc w zbędne szczegóły, uważam, że tylko psycholog (w każdym bądź razie gdy mamy do czynienia z zaawansowanymi testami psychologicznymi *sensu proprio*, a nie np. testami osiągnięć szkolnych czy sportowych) – jako że jest on, poprzez swoje akademickie wykształcenie odpowiednio przygotowany – powinien konstruować testy psychologiczne i je stosować. Oczywiście studia psychologiczne dają tylko podstawy, ale też wskazują na drogę pogłębienia i teorii i umiejętności.

TRAFNOŚĆ

Pojęcie trafności stało się osiowym pojęciem w procedurze konstrukcji testu oraz w procedurze interpretacji jego wyniku. W klasycznym wykładzie na temat trafności, tzw. **trynitarnej** koncepcji trafności testów, mówi się o tzw. „Trójcy Świętej” (*Holly Trinity*) mając na myśli trzy jej aspekty (jak pisał Robert M. Guion, 1980/2005, s. 273: „[...] trzy różne drogi do zbawienia psychometrycznego”): (a) **trafność teoretyczna**, (b) **trafność treściowa** i (c) **trafność kryterialna (diagnostyczna i prognostyczna)**.

Przyjmuje się, dość zgodnie, następujące zalecenia odnośnie do trafności testu:

Pojęcie trafności odnosi się do stopnia, w jakim dane **empiryczne** oraz **teoria** uzasadniają interpretację wyników testowych w zakładanym kierunku. Trafność jest zatem najbardziej podstawową kategorią w procesie tworzenia i oceny testu. Proces walidacji obejmuje ciągle zbieranie danych w celu dostarczenia mocnych podstaw naukowych proponowanej interpretacji wyników testowych. Procesowi temu podlega zatem kierunek interpretacji wyników testowych, a nie test jako taki. Jeżeli wyniki testowe są interpretowane na więcej niż jeden sposób, to każdy kierunek zakładanej interpretacji powinien zostać poddany walidacji (podkr. moje) (AERA/APA/NCME, 1999/2007, s. 31); oraz: należy przedstawić informacje uzasadniające każdą proponowaną interpretację i wykorzystanie wyników testowych, w tym dane empiryczne dotyczące trafności interpretacji wyników testowych oraz teorię leżącą u podstaw testu AERA/APA/NCME (1999/2007, s. 44).

Najważniejsza, z naukowego punktu widzenia (wszak psychologia należy do porządku nauki, a nie sztuki czy magii!) – wedle wielu opinii – jest **trafność teoretyczna** (*construct validity*), która została przedstawiona w bardzo ważnej dla rozwoju psychometrii publikacji autorstwa Lee J. Cronbacha i Paula E. Meehla (1955/2005). Drew Westen i Ro-

⁷ Osobiście chciałbym zwrócić uwagę na cztery koncepcje: Rudolfa Carnapa (1959), Jerzego Kmity (1973), Leszka Nowaka (1980) oraz Elżbiety Hornowskiej (1989).

bert Rosenthal (2003, s. 608) napisali dość znamienne słowa o trafności teoretycznej i jej miejscu we współczesnej myśli psychologicznej z którymi nie sposób się nie zgodzić:

[...] trafność teoretyczna jest jednym z najważniejszych pojęć w całej psychologii. Jest ona sercem jakichkolwiek badań w których badacze używają takiej miary jaką jest wskaźnik zmiennej, która nie jest bezpośrednio obserwowalna (np. inteligencja, agresja, pamięć robocza). Jeśli testowi psychologicznemu (albo, szerzej, procedurze psychologicznej, włączając manipulację eksperymentalną) brak trafności teoretycznej, to rezultaty otrzymane z pomocą takiego testu czy takiej procedury będą trudne do zinterpretowania (podkr. moje).

Podobnie autorzy rozszerzenia klasycznej teorii testów Harolda Gulliksena (1950) uważają, iż trafność teoretyczna jest, jak napisali w swoim fundamentalnym dziele (por. Lord i Novick, 1968, s. 278), „najważniejszą właściwością testu”. Podobną, do wyżej przytoczonej, opinię na temat ważności trafności teoretycznej dla teorii i praktyki psychologicznego testowania wyrażali też: Jane Loevinger (1957), Jum C. Nunnally i Ira H. Bernstein (1994), Samuel Messick (1995/2005), Anne Anastasi i Susana Urbina (1999), Jerzy Brzeziński (1984). Aby dać świadectwo historycznej prawdzie, trzeba też przywołać krytyka podejścia Donalda T. Cronbacha i Paula E. Meehla – Harolda P. Bechtoldta (Bechtoldt, 1959). Jednakże jego krytyczne uwagi⁸ nie spowodowały zejścia z tej, jak się okazało – po trzydziestu latach (Cronbach, 1989) – dobrej drogi rozwijania problematyki trafności. Warto też zauważyć, iż trafność teoretyczna była włączona jako kryterium oceny dobroci testu do standardów APA/AERA/NCME (1966) na skutek gorącej rekomendacji Donalda T. Campbella (1960)⁹.

Moim zdaniem (Brzeziński, 2007, s. 518; także: Lord i Novick, 1968, s. 279; Drwal 1995, s. 27; McDonald, 1999, s. 211), każdy nowy test powinien być wprowadzany przez wykazanie jego **zbieżności** (aspekt zbieżny trafności) nie tylko z innymi testami, ale – przede wszystkim – z kryteriami faktycznie

zewnętrznymi (a więc, w mojej ocenie, nietestowymi!) wg których może być dokonana ocena tej samej zmiennej. Ponadto należy wykazać brak zbieżności testu (aspekt różnicowy trafności) z podobnymi miarami (testowymi i nietestowymi), ale przeznaczonymi do oceny innych zmiennych. Mówiąc krótko, idzie o to, aby badanie trafności testu przeprowadzić metodą Donalda T. Campbella i Donalda W. Fiskego (1959/2005)¹⁰. To metoda Campbella i Fiskego jest **najważniejszą metodą badania trafności teoretycznej zastosowanej procedury testowania**, a jeżeli dodatkowo przyjmiemy, że i trafność teoretyczna jest tą, do której da się sprowadzić pozostałe dwie odmiany trafności – trafność kryterialną i trafność treściową, to dochodzimy do wniosku, iż ustalanie trafności testowania psychologicznego polega na sprawdzaniu jego trafności teoretycznej metodą macierzy „wielu cech, wielu metod” (por. Fiske i Campbell, 1992; Meier, 1994; Campbell, 1996; Messick, 1995/2005).

Wniosek z tego, co napisałem w tym punkcie da się sprowadzić do jednego zalecenia: dobry (albo mówiąc inaczej, akceptowalny na gruncie współczesnej psychometrii) test psychologiczny musi (!) wychodzić z jakiejś, empirycznie sprawdzonej teorii psychologicznej; poza jej kontekstem jest on, po prostu, zestawem pytań, zdań, obrazków, łamigłówek itp. To, że niektórzy niby-psychologowie usiłują „coś” z takiego „testu” wyczytać może świadczyć jedynie o bogactwie ich wyobraźni i o wysokim poziomie tego czegoś, co Melchior Wańkiewicz nazwał „chciewstwem”. W każdym bądź razie taka praktyka niewiele ma wspólnego z psychologią.

RZETELNOŚĆ ORAZ BŁĄD STANDARDOWY POMIARU

Zacznijmy od zdefiniowania drugiej ważnej własności, którą musi spełniać procedura testowania za pomocą testu psychologicznego (dawniej tę własność odnosiło się do samego testu) – co umożliwi nam uniknięcie niepotrzebnych (a od nich, niestety, nie jest wolna literatura psychologiczna, relacjonująca wyniki badań, których źródłem są testy psychologiczne) nieporozumień – a mianowicie **rzetelności**. Rozumie się przez nią (AERA/APA/NCME (1999/2007, s. 296):

[...] stopień, w jakim wyniki testowe otrzymane dla grupy osób badanych są stabilne oraz nie zmieniają się w kolejnych badaniach tym samym testem; mogą być uznane za powtarzalne i rzetelne dla osoby badanej; także stopień, w jakim wyniki testowe są wolne od błędu pomiaru dla danej grupy.

⁸ Polemizował z nimi Donald T. Campbell (1960).

⁹ Co prawda już w pierwszym wydaniu *Standardów* z 1954 (APA, 1954) roku była mowa o trafności teoretycznej, ale dopiero artykuł Cronbacha i Meehla (nawiasem mówiąc ekspertów APA biorących udział w pracach zespołu przygotowującego owe standardy) precyzyjnie ją dookreślił.

¹⁰ Jak podaje Starkey Duncan (2003) artykuł poświęcony metodzie analizy macierzy „wielu cech – wielu metod” doczekał się największej liczby cytowań w historii czasopisma *Psychological Bulletin* (ponad 4000!).

Można, jak miemam, wyróżnić sześć metod ustalania rzetelności (por. Brzeziński, 2007, s. 467):

1. Metody oparte na porównaniu dwukrotnego badania tym samym testem.
2. Metody oparte na porównaniu form alternatywnych (równoległych) testu.
3. Metody oparte na porównaniu części (połówek itp.) tego samego testu.
4. Metody oparte na analizie właściwości statystycznych pozycji testowych.
5. Metody oparte na analizie związku pozycji testowych z ogólnym wynikiem testu.
6. Metody oparte na badaniu stopnia zgodności sędziów kompetentnych oceniających odpowiedzi testowe.

Z kolei wg *Standardów...*

[...] wyróżnia się trzy duże kategorie współczynników rzetelności: (a) współczynniki otrzymane na podstawie niezależnego, dwukrotnego badania formami równoległymi testu (współczynniki dla form równoległych), (b) współczynniki otrzymane na podstawie dwukrotnego badania tym samym testem (współczynniki test – retest lub stabilności), (c) współczynniki otrzymane na podstawie jednokrotnego badania testem i oparte na wielkościach korelacji między wynikami dla poszczególnych pozycji testu czy skal (współczynniki zgodności wewnętrznej). Jeżeli obliczanie wyników w teście wymaga skorzystania z bardziej złożonych procedur oceniania, to dodatkowo oblicza się zazwyczaj współczynnik zgodności sędziów (s. 60).

Dość często analiza rzetelności wyników testu sprowadzana jest jedynie do stosowania rozwiązania zaproponowanego ponad pół wieku temu przez Lee J. Cronbacha, to jest współczynnika *alfa* (Cronbach, 1951/2005). Niestety *alfa*-Cronbacha jest dość często nadużywana, a jej „entuzjaści” nie zdają sobie sprawy z ważnych ograniczeń (por. np. Green, Lissitz i Mulaik, 1977; Schmitt, 1996 czy Sijtsma, 2009). Uważam (podobnie jak autorzy *Standardów...*, że rzetelności testu nie da się „załatwić” jednym li tylko współczynnikiem (np. wspomnianym już współczynnikiem *alfa*-Cronbacha). No bo, jak oszacować stabilność wyniku testowego, gdy odwołujemy się tylko do jednego jego pomiaru? I znowu odwołajmy się do *Standardów...* (s. 59): „[...] Wzorcowe podejście do badania rzetelności polega na niezależnych powtórzeniach całego procesu pomiarowego”.

Tak czy inaczej, należy zgodzić się z autorami *Standardów...* (s. 65):

[...] nie istnieje jedno, preferowane podejście do ilościowego sposobu wyrażania rzetelności. Żaden pojedynczy wskaźnik nie jest w stanie uwzględnić wszystkich istotnych czynników. Żadna metoda badania nie jest optymalna w każdej sytuacji ani żaden autor testu nie jest ograniczony tylko do jednej metody badania rzetelności dla danego narzędzia pomiarowego. Wybór odpowiedniej techniki i minimalny, dający się zaakceptować poziom każdego wskaźnika zawsze są elementem profesjonalnej oceny (podkr. moje).

I w związku z powyższym (s. 67–68):

STANDARD 2.4.

Każda metoda ilościowego określania stopnia dokładności testu czy stałości wyników testowych powinna zostać jasno opisana i wyrażona za pomocą terminów statystycznych właściwych dla zastosowanej metody. Należy także przedstawić procedury doboru grup osób do badań dotyczących rzetelności oraz odpowiednie statystyki opisowe charakteryzujące te grupy.

Komentarz: Informacje dotyczące metod doboru osób badanych, liczebności prób, średnich, odchyłek standardowych oraz charakterystyk demograficznych tych prób pomagają osobom stosującym testy określić, w jakim zakresie przedstawione dane odnoszą się do ich własnych populacji osób badanych. Jeżeli wykorzystuje się metodę test – retest czy metodę form równoległych, to należy podać przedział czasowy, jaki minął między kolejnymi testowaniami. Ponieważ istnieje wiele sposobów szacowania rzetelności i każda ujawnia inne źródła błędów, to **niedopuszczalne jest proste stwierdzenie: „Rzetelność testu X wynosi .90”**. Powinno się podać następującą informację: „Współczynnik rzetelności wyników w teście X wynosi .90 i otrzymano go w efekcie korelacji wyników uzyskanych w wersjach A i B tego samego testu, przeprowadzonego w kolejnych dniach. Przedstawione dane otrzymano w próbie liczącej 400 uczniów dziesiątych klas, pochodzących z pięciu szkół. Szkoły te znajdowały się w stanie Nowy York, w okolicach zamieszkałych głównie przez klasy średnie. Skład demograficzny próby był następujący...” (podkr. moje).

Konsekwencją zaakceptowania **teorii wyniku prawdziwego** (*true score theory*), w wersji klasycznej, pochodzącej od Gulliksen, 1950 czy **teorii uniwersalizacji** (*generalizability theory*), w wersji rozszerzonej, pochodzącej od F.M. Lorda i M.R. Novicka (1968) jest „szukanie” **wyniku prawdziwego** czy **wyniku uniwersalnego** (*universe score*). Zaś zalecaną drogą owego „poszukiwania” jest **estymacja przedziałowa wyniku prawdziwego**. „Interpretując wyniki testowe trzeba brać pod uwa-

gę raczej ich przedziały ufności niż ich wartości absolutne [...]” (por. Standard J2.2., APA (1974/1985, s. 80). I w najnowszej wersji *Standardów...*: „[...] Przedział ufności dla wyniku prawdziwego, wyniku uniwersalnego czy wyniku centylowego osoby badanej pełni niemal takie same funkcje jak standardowy błąd pomiaru i może być wykorzystywany jako alternatywna forma przedstawiania danych na temat rzetelności. [...]” (AERA/APA/NCME, 1999/2007, s. 66).

Estymację wyniku prawdziwego można przeprowadzić tylko wówczas gdy użytkownik testu dysponuje informacją o wielkości **błędu standardowego pomiaru**, SEM (*standard error of measurement*).

W praktyce pomiarowej (a, tak naprawdę, to prawie 100% testów będących w obiegu diagnostycznym opartych jest na modelu H. Gulliksen) najczęściej wykorzystywany jest tylko ten jeden z błędów standardowych, a mianowicie **błąd standardowy pomiaru SEM**. Zauważmy, że popularny wykład psychometrii zawarty w klasycznym podręczniku psychometrii autorstwa Davida Magnussona (1991) zawiera omówienie tylko SEM. Różne odmiany błędu standardowego (oraz wskazanie sensownego ich wykorzystania w praktyce pomiarowej) zostały kompetentnie przedstawione w pracy Andrzeja Machowskiego (1993)¹¹. Jest to, jak dotąd, jedyne opracowanie, w polskiej literaturze psychometrycznej problemu błędów standardowych.

Na co powinno się zwracać uwagę przy budowie przedziału ufności? Przede wszystkim na poziom ufności. Im bardziej jest on „wyśrubowany”, tym bardziej wzrasta nasza pewność (ufność), że zbudowany przedział (np. symetrycznie wokół, wyniku otrzymanego, gdy wychodzimy od SEM) pokrywał będzie nieznaną, a przez nas poszukiwaną wartość wyniku prawdziwego. „Decyduje” o tym wyrażenie: „ $z_{\alpha/2}$ SEM” we wzorze na SEM. Jeżeli bowiem przejdziemy od niższego poziomu ufności $P=85\%$ ($z_{0,15/2}=1.44$) do wyższego poziomu ufności $P=99\%$ ($z_{0,01/2}=2.58$), to granice przedziału wydłużą się, gdyż stałą wartość błędu standardowego będziemy przemnażać – przy obliczaniu wartości jego dolnej i górnej granicy – nie przez 1.44, ale przez 2.58. Zatem cena, jaką musimy „zapłacić” za wzrost pewności jest spadek precyzji estymacji (zamiast „krótkiego” uzyskujemy „długi” przedział ufności).

¹¹ Można wyróżnić następujące rodzaje błędów standardowych: 1) błąd pomiaru wyniku otrzymanego (czyli „nasz” SEM), 2) błąd estymacji wyniku prawdziwego, 3) błąd pomiaru wyniku prawdziwego, 4) błąd prognozy, 5) błąd zastąpienia.

Z kolei, gdy chcemy uzyskać precyzyjniejszą estymację wyniku prawdziwego (krótszy przedział ufności), to musimy wstawić do wzoru na przedział ufności wartość $z_{\alpha/2}$ jak najniższą – zamiast $z=2.58$ dla $P=99\%$ wartość $z=1.0$ dla $P=68\%$. Jednak co z tego, że będziemy dysponowali krótkim przedziałem ufności, dającym precyzyjną estymację wyniku prawdziwego, gdy zaufanie do uzyskanego wyniku jest niskie? No cóż, użytkownik testu musi znaleźć złoty środek między precyzją estymacji i stopniem zaufania do niej. Nie można podać jednej prostej reguły postępowania. Można jednak sformułować w miarę ogólną dyrektywę. Może ona brzmieć następująco: **jeżeli wynik prowadzonego badania testowego nie będzie miał dotkliwych dla osoby badanej następstw, to wybierajmy krótsze przedziały ufności (oparte na niskich wartościach statystyki $z_{\alpha/2}$).**

Rekapitulując, podejmowanie jakichkolwiek działań odwołujących się w ich warstwie diagnostycznej (pomiarowej) do wyników badań testowych będzie w tym sensie nieetyczne, gdy owe działania będą poparte (uzasadnione) wynikami testów o nieznannej (albo wadliwie wyznaczonej) rzetelności i nieznannej wartości SEM – co w efekcie uniemożliwi przeprowadzenie przedziałowej estymacji **wyniku prawdziwego**. Owe wyniki będą obciążone błędem (jak znacznym?) niedoszacowania albo przeszacowania prawdziwej wartości mierzonej cechy. Mówiąc brutalnie, takie „podsuwanie” osobom decydującym np. o losach jakiejś osoby (klienta, pacjenta, ucznia) pozornie ścisłego (bo przyjmującego naturę numeryczną) uzasadnienia stawia psychologa w tym samym rzędzie, obok, astrologa, wróżki, kabalisty, psychoanalityka itp. Czyż bowiem taki test (nie-test, niby-test) istotnie się odróżnia od kryształowej kuli, talii kart czy rozrzuconych na piasku kosteczek? Psycholog sięgający po taki niby-test psychologiczny przekreśla, swój uniwersytecki dyplom.

PODSUMOWANIE

Niniejsze opracowanie nie wyczerpuje rzecz jasna całej problematyki „zanurzenia” praktyki psychologicznego testowania w kontekst etyczny. Z konieczności skupiło się ono tylko na jednym, ale za to o fundamentalnym znaczeniu aspekcie. Jeżeli testy psychologiczne traktować poważnie, to nie można ich oderwać od ich zaplecza teoretycznego i to rozumianego dwojako.

W pierwszym jego rozumieniu na plan pierwszy wysuwa się pytanie o to, na jakie pytanie chce uzyskać odpowiedź użytkownik testu (najczęściej jest

nim psycholog, którego ukończone studia psychologiczne przygotowały do korzystania z wyników testowych)? Otóż chce on, za pomocą testu dokonać **operacjonalizacji**¹² zmiennej teoretycznej, która wcześniej została zdefiniowana na gruncie empirycznie sprawdzonej teorii psychologicznej – i tu odwołujemy się do pojęcia **trafności** testowania, a zwłaszcza do **trafności teoretycznej** (w sensie Cronbacha i Meehla), którą najlepiej sprawdzać metodą Campbella i Fiske'go (*MTMM matrix – macierz wielu cech – wielu metod*). I – powiedzmy to wyraźnie (zgodnie z tym jak się współcześnie traktuje pojęcie trafności na gruncie psychometrii) – poza kontekstem empirycznie sprawdzonej teorii psychologicznej dany zestaw pozycji (pytań, zadań, obrazków itp.) nie zasługuje na miano testu psychologicznego.

Z kolei w drugim rozumieniu, zaplecze teoretyczne testu rozumiane jest jako **teoria testu** (dziś, mimo rozbudowywanej IRT, jest to teoria Gullikse-na-Lorda-Novicka: **teoria wyniku prawdziwego i teoria uniwersalizacji**). Kluczowym zaś na jej gruncie jest pojęcie rzetelności i błędu standardowego pomiaru. W artykule zwrócono szczególną uwagę, w ślad za dwoma ostatnimi wydaniem Standardów..., na konieczność pełnego, całościowego badania rzetelności za pomocą różnych metod, które akcentują różne jej aspekty. Zwłaszcza chodzi o badanie stabilności pomiaru testowego. Tak się jakoś ułożyło w ostatnich latach, że psychologowie ograniczają badanie tylko do jednego aspektu i tylko do jednej metody – **wzoru alfa-Cronbacha**. To jednak jest istotnym zubożeniem analizy rzetelności testowania.

Pełna analiza etycznego kontekstu testowania psychologicznego musi też uwzględniać i inne ważne jego wymiary – osobę badaną (jej prawa i obowiązki czy gotowość mentalną do podania się badaniu testowemu), instytucję zamawiającą i korzystającą z wyników badania testowego. Jeśli zaś chodzi o samego użytkownika testu, to ważne są też – co nie powinno budzić zdziwienia – ograniczenia etyczne wynikające z respektowania czterech zasad etycznych: świadomej zgody, poufności, anonimowości (oczywiście nie można jej respektować w badaniach diagnostycznych) i prywatności (por. APA, 2002; AERA/APA/NCME, 1999/2007).

Bardzo dynamiczny postęp technologiczny w obszarze badawczej penetracji mózgu (np. fMRI, CT, wielokanałowy EEG) skłania niektórych badaczy o orientacji neurokognitywistycznej do formułowa-

nia dość radykalnych sądów w kwestii przyszłości testów. Zgoda, pod względem precyzji diagnostyka komputerowa związana ze skanowaniem mózgu jest bezkonkurencyjna. Jednakże istotne obniżenie jej kosztów i dostępności to jeszcze kwestia przyszłości. Póki co, jesteśmy jeszcze przez jakiś czas „skazani” na posługiwanie się testami psychologicznymi, które – być może brzmi to paradoksalnie – jednak skutecznie konkurują z ich finezyjnymi aparaturowymi odpowiednikami (czy zastępnikami?).

Ponieważ testy psychologiczne nie zostały odsłane do muzeum psychologii, więc musimy zadbać o to, aby były one odpowiedzialnie (czyli etycznie) stosowane w różnych obszarach aktywności psychologów.

LITERATURA

- Anastasi, A., Urbina, S. (1999). *Testy psychologiczne*. Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- AERA/APA/NCME; American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999/2007). *Standardy dla testów stosowanych w psychologii i pedagogice*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- AERA/APA/NCME; American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1974/1985). *Standardy dla testów stosowanych w psychologii i pedagogice*. Warszawa: Laboratorium Technik Diagnostycznych Pol. Tow. Psychol. i Wydziału Psychologii UW.
- APA/AERA/NCME; American Psychological Association, American Educational Research Association, National Council on Measurement in Education (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: APA.
- APA; American Psychological Association (1954). *Technical recommendations for psychological tests and diagnostic techniques*. Washington, DC: Author.
- APA; American Psychological Association (2000). *Report of the task force on test user qualifications. Practice and science directorates. Practice and science directorates*. <http://www.apa.org/science/tuq.pdf>. Pozyskano: 26. 01. 2008 r.
- APA; American Psychological Association (2002). *Ethical principles of psychologists and code of conduct*. www.apa.org/ethics/code2002.html. Pozyskano: 15.12.2007 r.
- Bechtoldt, H.P. (1959). Construct validity: A critique. *American Psychologist*, 14, 619–629.
- Binet, A., Simon, T. (1905a). Sur le nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence. *L'Année Psychologique*, 11, 163–190.
- Binet, A., Simon, T. (1905b). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 11, 191–244.
- Binet, A., Simon, T. (1905c). Application des méthodes nouvelles au diagnostic du niveau intellectuel chez enfants et normaux d'hospice et d'école. *L'Année Psychologique*, 11, 245–366.

¹² W sensie: Brzezińskiego (2007) i Hornowskiej (1989).

- Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, 24, 383–405.
- Bridgman, P.M. (1927). *The logic of modern physics*. New York, NY: Macmillan.
- Brzeziński, J. (2003). Jak to z pomiarem testowym IQ było? W: Z. Piskorz, T. Zaleśkiewicz (red.), *Psychologia umysłu* (s. 96–111). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Brzeziński, J. (1984). Badanie testu psychometrycznego metodą analizy wariancji. W: J. Brzeziński (red.), *Wybrane zagadnienia z psychometrii i diagnostyki psychologicznej* (s. 9–45). Poznań: Wydawnictwo Naukowe UAM.
- Brzeziński, J. (2007). *Metodologia badań psychologicznych*. 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. (2009). Mały lekarz, mały grabarz... mały psycholog? *Charaktery. Wydanie specjalne. Psychologia dziś*, 2, 7, 60–62.
- Campbell, D.T., Fiske, D.W. (1959/2005). Analiza trafności zbieżnej i różnicowej za pomocą macierzy wielu cech – wielu metod [Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105]. W: J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 431–459). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Campbell, D.T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, 15, 546–553.
- Campbell, D.T. (1996). Unresolved issues in measurement validity: An autobiographical overview. *Psychological Assessment*, 8, 363–368.
- Carnap, R. (1959). The methodological character of theoretical concepts. W: H. Feigl, M. Scriven (red.), *Minnesota studies in the philosophy of sciences* (t. 1, s. 38–76). Minneapolis, MN: University of Minnesota Press.
- Cronbach, L.J. (1989). Construct validation after 30 years. W: R.L. (red.), *Intelligence. Measurement theory and public policy* (s. 147–171). Champaign: University of Illinois Press.
- Cronbach, L.J. (1951/2005). Współczynnik alfa a struktura wewnętrzna testów. [Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334]. W: J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 177–212). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Cronbach, L.J., Meehl, P.E. (1955/2005). Trafność teoretyczna testów psychologicznych [Construct validity in psychological test. *Psychological Bulletin*, 52, 281–302]. W: J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 403–430). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Drwal, R.Ł. (1995). *Adaptacja kwestionariuszy osobowości*. Warszawa: Wydawnictwo Naukowe PWN.
- Duncan, S. (2003). Donald Winslow Fiske (1916–2003). *American Psychologist*, 58, 1092–1093.
- Fiske, D.W., Campbell, D.T. (1992). Citations do not solve problem. *Psychological Bulletin*, 112, 393–395.
- Garb, H.N., Wood, J.M., Lilienfeld, S.O., Nezworski, M.T. (2002). Effective use of projective techniques in clinical practice: Let the data help with selection and interpretation. *Professional Psychology: Research and Practice*, 33, 454–463.
- Green, S.B., Lissitz, R.W., Mulaik, S.A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827–838.
- Gregory, R.J. (2004). *Psychological testing. History, principles, and applications* (wyd. 4). Boston, MA: Allyn and Bacon.
- Guion, R.M. (1980/2005). O trynitarnej doktrynie trafności [On trinitarian doctrines of validity. *Professional Psychology*, 11, 385–398]. W: J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 272–287). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Hornowska, E. (1989). *Operacjonalizacja wielkości psychologicznych. Założenia – struktura – konsekwencje*. Wrocław: Ossolineum.
- ITC, International Test Commission (2000). *Wytyczne międzynarodowej komisji ds. testów (international test commission – ITC) z 2000*.
- ITC, International Test Commission (2005). *Międzynarodowe wytyczne dotyczące komputerowych i internetowych badań testowych: wersja polska*. Przekład autoryzowany przez Polskie Towarzystwo Psychologiczne <http://www.practest.com.pl/html/wytyczne-miedzynarodowej-komisji-ds-testow-itc>. Pozyskano 17. 01. 2010 r.
- Jacob, F. (1973). *Historia i dziedziczność*. Warszawa: Państwowy Instytut Wydawniczy.
- JCTP; Joint Committee on Testing Practices (1998). *Code of fair testing practices in education*. Washington, DC: APA <http://www.apa.org/science/programs/testing/fair-code.aspx>. Pozyskano 30. 01. 2010 r.
- Kmita, J. (1973). *Wykłady z logiki i metodologii nauk*. Warszawa: PWN.
- Lilienfeld, S.O., Wood, J.M., Garb, H.N. (2002). *Status naukowy technik projekcyjnych*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635–694.
- Lord, F.M., Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Machowski, A. (1993). *Rzetelność testów psychologicznych. Dwa ujęcia modelowe*. Warszawa-Poznań: Wydawnictwo Naukowe PWN.
- Magnusson, D. (1991). *Wprowadzenie do teorii testów* (wyd. 2.). Warszawa: Wydawnictwo Naukowe PWN.
- McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: L. Erlbaum.
- Meier, S.T. (1994). *The chronic crisis in psychological measurement and assessment: A historical survey*. San Diego: Academic Press.
- Messick, S. (1980/2005). Trafność testu a etyka oceny (diagnozy). [Test validity and the ethics of assessment. *American Psychologist*, 35, 1012–1027]. W: J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wy-*

- bór tekstów (s. 460–485). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Messick, S. (1995/2005). Trafność oceny psychologicznej. [Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749]. W: J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 384–402). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Nowak, L. (1980). *The structure of idealization*. Dordrecht: D. Reidel.
- Nunnally, J.C., Bernstein, I.H. (1994). *Psychometric theory* (wyd. 3). New York: McGraw-Hill.
- Paluchowski, W.J., Stemplewska-Żakowicz, K. (2008). Podstawy diagnozy psychologicznej. W: J. Strelau, D. Doliński (red.), *Psychologia. Podręcznik akademicki* (t. 2, s. 23–94). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Psychological Review (1945). *Symposium on operationism*, 52, 5, cały numer.
- Roid, G.H. (2003a). *Stanford-Binet Intelligence Scales. Fifth edition, Examiner's manual*. Itasca, IL: Riverside Publishing.
- Roid, G.H. (2003b). *Stanford-Binet Intelligence Scales. Fifth edition, Interpretive manual*. Itasca, IL: Riverside Publishing.
- Roid, G.H. (2003c). *Stanford-Binet Intelligence Scales. Fifth edition, Technical manual*. Itasca, IL: Riverside Publishing.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 6, 350–353.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107–120.
- Stemplewska-Żakowicz, K. (2004). *O rzeczach widzianych na obrazkach i opowiadanych o nich historiach. TAT jako metoda badawcza i diagnostyczna*. Warszawa: Wydawnictwo Academica Wydawnictwo SWPS.
- Stemplewska-Żakowicz, K. (2009). *Diagnoza psychologiczna. Diagnozowanie jako kompetencja profesjonalna*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Stevens, S.S. (1935/2000). Operacyjne definiowanie pojęć psychologicznych. W: J. Siuta, K. Krzyżewski (red.), *Behawioryzm i psychologia świadomości* (s. 107–115). Kraków: Wydawnictwo UJ.
- Watkins, Jr.C.E., Campbell, V.L., Nieberding, R., Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice*, 26, 54–60.
- Wechsler, D. (2008). *WAIS-IV. Wechsler Adult Intelligence Scale-Fourth Edition. Technical and interpretive manual*. San Antonio, TX: Pearson.
- Westen, D., Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, 84, 608–618.
- Witwicki, W. (1928). O narodowych testach amerykańskich do badania inteligencji. *Psychotechnika*, 2, 6, 23–32.

Artykuł nawiązuje do mojego referatu: *Kiedy psycholog odwołujący się do wyników testów psychologicznych postępuje nieetycznie?* wygłoszonego na konferencji: „Etyczne i prawne granice badań naukowych”, która była zorganizowana przez Polskie Towarzystwo Bioetyczne, Interdyscyplinarne Centrum Etyki UJ oraz Instytut Filozofii Uniwersytetu Jagiellońskiego w Krakowie-Przegorzałach w dniach: 29–31. 08. 2008 r.