

JERZY MARIAN BRZEZIŃSKI

Uniwersytet im. Adama Mickiewicza w Poznaniu
Instytut Psychologii UAM

O OSOBLIWOŚCIACH METODOLOGICZNYCH BADAŃ NAUKOWYCH I DIAGNOSTYCZNYCH PROWADZONYCH PRZEZ PSYCHOLOGÓW KLINICZNYCH

Artykuł skupia się na metodologicznych osobliwościach praktyki badawczej (badania naukowe i badania diagnostyczne) psychologii klinicznej. Autor wskazuje na typowe odstępstwa tej praktyki od standardów metodologicznych współczesnej psychologii – zwłaszcza tych, które wpisane zostały we wzorzec diagnozy opartej na dowodach empirycznych i praktyki psychologicznej opartej na dowodach empirycznych: lekceważenie prymarnej roli empirycznych teorii psychologicznych, zbyt częste odstępstwa od standardów psychometrycznych w procedurach konstruowania i stosowania testów psychologicznych (bywa, że są to pseudotesty, jak np. Test Drzewa Kocha). W artykule podkreślono ważność trafności teoretycznej testu psychologicznego. Podniesiono też problem zapewniania badaniom naukowym prowadzonym przez klinicystów trafności wewnętrznej i zewnętrznej.

Słowa kluczowe: diagnoza psychologiczna; teoria psychologiczna; proces badawczy; EBA; EBPP; trafność wewnętrzna; trafność zewnętrzna; psychologia kliniczna.

Do podjęcia tego tematu skłoniła mnie powtórna, po około trzydziestu latach, lektura klasycznego artykułu jednego z najoryginalniejszych psychologów klinicznych, Saula Rosenzweiga (1907-2004)¹. Mam na myśli jego artykuł: „The experimental situation as a psychological problem” (Rosenzweig, 1933). Rosenzweig w swojej pracy pisał o „osobliwościach” (*peculiarities*) badań eksperymentalnych w psychologii. W skrócie, zwrócił on wagę na to, że:

– badacz staje się elementem sytuacji badawczej;

Adres do korespondencji: JERZY MARIAN BRZEZIŃSKI – Instytut Psychologii, Uniwersytet im. Adama Mickiewicza w Poznaniu, ul. Szamarzewskiego 89AB, 60-568 Poznań; e-mail: brzezuam@amu.edu.pl

¹ O Saulu Rosenzweigu – por. Kaufman (2007).

– wpływ na zachowanie się osoby badanej w sytuacji badawczej mają takie zmienne związane z osobą badaną ją charakteryzujące, jak: osobowość, motywacja itp.;

– zawiązuje się interakcja: badacz–osoba badana.

Ja zaś chciałbym – wykraczając poza materię opisaną przez S. Rosenzweiga i, tak po prostu, ją akceptując (por. też: Brzeziński, 1994) – w tym szkicu metodologicznym rozważyć problem osobliwości badań (ale oglądanych z perspektywy „twardych” standardów metodologicznych) – naukowych i diagnostycznych – prowadzonych przez psychologów klinicznych (dalej: klinicystów). Zaś zależności opisane przez Rosenzweiga zaliczyłbym raczej do *psychologii społecznej badania psychologicznego* – że nawiążę do tytułu głośnej w latach 70. XX wieku książki Arthura G. Millera (1972): *The social psychology of psychological research*.

Rzecz jasna, psychologia kliniczna jest działem psychologii po prostu (rozumianej jako nauka empiryczna, i nie inaczej). Od razu też zaznaczę, że nie jest ona odrębną psychologią. W tej kwestii zajęliśmy (Brzeziński i Toeplitz-Winiewska, 2008, s. 305; też: Brzeziński, 1996, 2013) następujące stanowisko:

W „odwiecznym” sporze psychologów uniwersyteckich z psychologami praktykami (bo tu mniej więcej biegnie linia demarkacyjna) zajmujemy [...] analogiczne stanowisko do wyrażonego przez Matarazzo (1987) czy Ellisa (1992), a dające się wyrazić słowami następującej deklaracji: jest tylko jedna psychologia, a nie wiele psychologii stosowanych; jej ustalenia empiryczne są stosowane w różnych dziedzinach praktyki społecznej – jako odpowiedź sfery nauki na zgłaszane zapotrzebowania społeczne.

Zatem – i nie jest to tylko mój punkt widzenia – „osobliwość” metodologiczna praktyki badawczej i diagnostycznej psychologii klinicznej nie jest związana z rozwiązywaniem przez nią specyficznych „zamówień” płynących ze sfery praktyki społecznej (wedle schematu opisanego przeze mnie w: Brzeziński, 2013). Pod tym bowiem względem (innymi słowy: z tego poziomu ogólności rozpatrując ów problem) nie różni się ona od tak wyspecjalizowanych, nachylnych praktycznie działów psychologii, jak psychologia pracy, psychologia penitencjarna czy psychologia wychowawcza. Podkreślmy zatem raz jeszcze, że wymienione przeze mnie, tytułem przykładu, nachylone praktycznie „sub-psychologie” nie są odrębnymi, w sensie metodologicznym, psychologiami. Nie oznacza to jednak, że praktyka badawcza (naukowa i diagnostyczna) klinicystów nie jest szczególnie podatna na występowanie odstępstw od standardów metodologicznych obowiązujących w psychologii po prostu. O owym nachyleniu, charakteryzującym się lekceważeniem empirycznych teorii psychologicznych i, w konsekwencji, dryfowaniu w stronę paranauki, chciałbym podyskutować.

Rzecz jasna niezauważanie współczesnych standardów metodologicznych, lekceważenie empirycznych teorii psychologicznych czy nadmierne przywiązanie do osobistych doświadczeń (praktyki) klinicystów jako głównego źródła wiedzy diagnostycznej nie jest jedyną osobliwością psychologii klinicznej. Ale nie o wszystkim w tak krótkim opracowaniu można napisać.

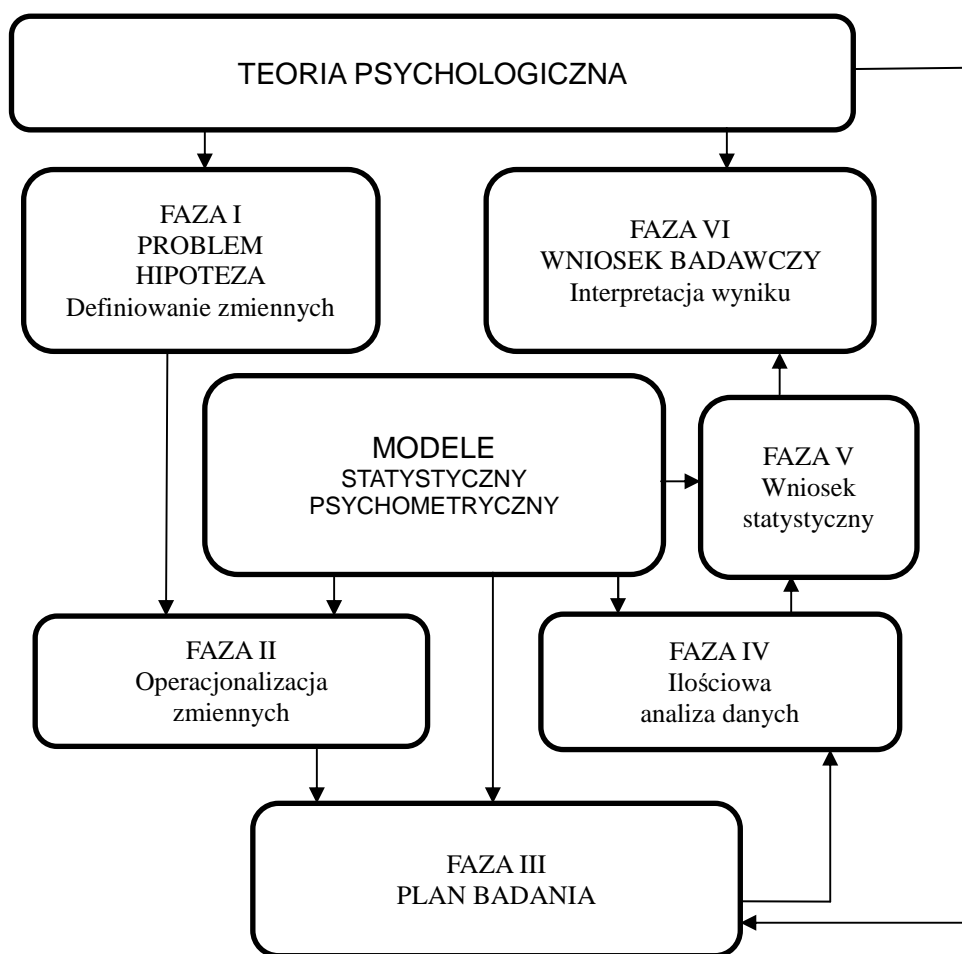
Od teorii naukowej do badania empirycznego, ale nie bez teorii

Przy różnych okazjach psychologowie (zwłaszcza ci, którzy postulują wykonywanie zawodu psychologa w perspektywie poważnej teorii psychologicznej) zwykli powoływać się na wypowiedzi słynnych uczonych eksponujących znaczenie teorii w skutecznym uprawianiu praktyki (np. klinicznej – diagnostyka czy psychoterapia). Swoistą karierę wśród psychologów zrobiła wypowiedź fizyka, Roberta Kirchoffa (dość często przypisywana, niesłusznie, psychologowi – Kurtowi Lewinowi): „Najpraktyczniejszą rzeczą jest dobra teoria”. Można też przywołać znane wypowiedzi: socjologa, Davida Silvermana (2007, s. 31): „Bez teorii takie zjawiska, jak «płeć», «osobowość», «rozmowa» lub «przestrzeń», nie mogą być zrozumiałe na gruncie nauk społecznych. W tym sensie, bez teorii nie mamy czego badać”, czy biologa, François’a Jacoba (1973, s. 32): „W wymianie między teorią a doświadczeniem teoria zawsze zaczyna dialog jako pierwsza. To ona wyznacza postać pytania, a więc również granice odpowiedzi”.

Zanim jednak zaakcentuję relację „teoria–doświadczenie (praktyka)”, odniesioną do postępowania klinicystów, przedstawię wpiery standard metodologiczny, wedle którego regulowane jest współczesne postępowanie badawcze w psychologii (szerzej: w naukach społecznych). Prezentuje go Rysunek 1, na którym wyeksponowano teoretyczne składniki postępowania badawczego (przynajmniej tak, jak ja to widzę).

Zacznijmy od tego, że postępowanie badawcze jest „zanurzone” w kontekście teoretycznym. Poza nim badanie, czy to naukowe, czy to diagnostyczne, być może – zabrzmie to mocno, ale tak uważam – nie ma żadnego sensu; zwłaszcza ów bezsens ujawnia się szczególnie wyraziście, gdy klinicyści zabierają się za „poprawianie” psychiki swoich klientów, a jedyne, co mają im do zaoferowania, to własne „doświadczenie” (po co im zatem potrzebna etykieta: „psycholog”?). Problem jednak z teorią psychologiczną, metodologicznie poprawną i znaczeniowo nośną. Tych nie mamy, niestety, w psychologii zbyt wiele. Ale, w co chcę wierzyć, wraz ze wzrostem *świadomości metodologicznej* psychologów (w czym niebagatelną rolę odgrywają studia psychologiczne prowadzone na dobrych,

badawczo zorientowanych – tu upomniałbym się o przeszczepienie na grunt polski idei amerykańskich *research universities*) prawdziwe teorie będą wypierać ich atrakcyjnie opakowane harlekinowe imitacje.



Rysunek 1. Schemat postępowania badawczego.

Teoria dostarcza badaczowi języka – z jego „gramatyką” i „słownikiem” – w którym można budować sensowne (w sensie psychologicznym) zdania. Te zaś służą nie tylko do opisanie i wyjaśniania zjawisk, ale też jako teoretyczna podpora świadczonej przez klinicystów profesjonalnej pomocy. Nauka jest wielopara-

dygmatyczna (i psychologia też). Badacz dokonuje wyboru (najlepiej, gdy jest on w pełni uświadamiany) między *paradygmatami* (w sensie: Kuhn, 2001). Opowiedzenie się za jednym z paradygmatów umożliwia „zejście” na niższy poziom analizy teoretycznej i analizy empirycznej. Tutaj jednak łatwo popaść w kłopoty, gdy chciałby on łączyć elementy teorii zaczerpniętych z różnych, wzajemnie wykluczających się paradygmatów. Otóż wszystkie zmienne teoretyczne wprowadzone do teorii muszą być definiowane w języku tego samego paradygmatu. Badacz powinien przestrzegać podstawowej zasady: *zasady niesprzeczności paradygmatycznej*. Klinicyści, na ogół prowadzący badania na poziomie subparadygmatycznym, nie zawsze zdają sobie sprawę z ważności owej zasady. W języku danej teorii lub teorii niesprzecznych ze spajającym je paradygmatem definiuje się zmienną zależną oraz zmienne niezależne tworzące *obraz przestrzeni zmiennych istotnych dla danej zmiennej zależnej $O(P_Y)$* . Z tak zdefiniowanych zmiennych badacz formułuje problemy badawcze i hipotezy badawcze (Faza I).

Każda teoria psychologiczna wymaga zinterpretowania empirycznego. Terminy (zmienne) teoretyczne muszą być powiązane z terminami obserwacyjnymi. Mówiąc inaczej, przejście do Fazy II wymaga nadania sensu empirycznego zmiennym teoretycznym. Bardzo ważna jest więc adekwatna do założeń teorii *operacjonalizacja* wymiarów skonstruowanego przez badacza $O(P_Y)$. Program operacjonalizacji zmiennych ważnych dla Y, jak i dla samej Y musi być wywieziony z teorii i – w szczególności – z nią niesprzeczny. Metodologicznie niepoprawny jest taki program operacjonalizacji zmiennych, który odwołuje się do różnych teorii (a dokładnie – do definicji zmiennych teoretycznych zbudowanych na gruncie teorii między sobą niezgodzonych (pochodzących z różnych paradygmatów). No bo jak pogodzić metodę Hermanna Rorschacha (jaka empiryczna teoria psychologiczna ją usensawnia?) z kwestionariuszem BDI Aarona Becka? Oczywiście jest też i to, że język interpretacji rezultatów pomiarów jest językiem tej samej teorii (lub z nią niesprzecznej), w której zdefiniowano zmienne i sformułowane na jej gruncie hipotezy badawcze. Badacz jest więc zobowiązany do przestrzegania *zasady zgodności programu operacjonalizacji z programem teoretycznym*.

Stosowane, dość chętnie, przez klinicystów testy psychologiczne czy niezliczone kwestionariusze osobowości wymagają kilka zdań komentarza (pomijam tu popularne w środowisku klinicystów tzw. metody projekcyjne: metodę Rorschacha, Test Apercepcji Tematycznej czy Test Zdań Niedokończonych jako metody o wątpliwej wartości naukowej). Szczególnie popularna w tym środowisku jest Skala Inteligencji Davida Wechslera, i to od momentu skonstruowania

przez Wechslera w 1939 roku jej pierwszej wersji: *Wechsler-Bellevue Intelligence Scale W-B*. Klucyści posługują się nią nie tylko jako miarą IQ, ale też jako swoistym testem osobowości i klinicznych odchyień od normy (por. Wechsler, 1941/1998; Rapaport, 1945; Zimmerman i Woo-Sam, 1973; Frank, 1984; Kowalik, 1998; Tulsky, Saklofske, Chelune i in., 2003).

Skale Inteligencji D. Wechslera są przykładem metody, która została skonstruowana na podstawie doświadczeń klinicznych i intuicji diagnostycznych jej twórcy, który „pod jednym dachem” zgromadził najbardziej przydatne w diagnostyce psychologicznej testy, tworząc z nich swoistą baterię testów (Boake, 2002). W sensie „teoretycznym” spajała ją zaproponowana przez niego definicja inteligencji jako:

[...] jest to zagregowana (ogólna) zdolność jednostki do podejmowania działań celowych, racjonalnego myślenia i do efektywnego radzenia sobie we własnym środowisku. Jest ona globalna, ponieważ charakteryzuje zachowanie jednostki jako całości, jest zagregowana, gdyż składa się z elementów (zdolności), które chociaż nie są całkowicie niezależne, ale są jakościowo odróżniane. Dokonując pomiaru tych zdolności, dokonujemy w efekcie oszacowania poziomu inteligencji (Wechsler, 1939/1998, s. 16).

To, co psychologowie próbowali później zrobić, to zrekonstruować podstawy teoretyczne tej najbardziej chyba w świecie znanej metody testowej. Jak napisano w przewodniku do *Wechsler Adult Intelligence Scale WAIS-III* z 1997 r.:

Testów Wechslera nie tworono w oparciu o teorię – być może z wyjątkiem teorii *g* Spearmana (1927), czy też ogólnej teorii inteligencji – lecz w oparciu o ujęcia praktyczne i kliniczne. [...] W zamierzeniu Wechslera, testy inteligencji miały być sposobem na uzyskanie wglądu w osobowość jednostki. Wiele lat po powstaniu skal Wechslera snuto szeroko zakrojone teoretyczne spekulacje na temat natury i znaczenia tych testów oraz ich wyników, jednak pierwotnie testy te tworzone były bez odniesienia do teorii (Kaufman i Lichtenberger, 1999, s. 3).

Przytoczony tu fragment wskazuje na brak oryginalnych teoretycznych podstaw nie tylko tej wersji Skali Inteligencji, lecz także wersji ją poprzedzających (por. Hornowska, 1988). W latach, które minęły od daty opublikowania Skali Inteligencji Wechsler-Bellevue, podejmowano próby dokonania *ex post* rekonstrukcji teoretycznych podstaw skal Wechslerowskich. Próbowano „ulokować” testy składające się na baterię Wechslerowską w ramach Charlesa Spearmana *teorii czynnika g* (badanie nasycenia czynnikiem *g* poszczególnych testów – por. Leckliter, Matarazzo i Silverstein, 1986). Dwie znaczące teorie inteligencji, z którymi usiłowano powiązać testy Wechslera, to: Raymonda B. Cattella i Johna L. Horna *teoria płynnej i skryształizowanej inteligencji* oraz Joya P. Guilforda *model struktury intelektu*. Także w przewodniku do WAIS-III i WAIS-IV wymie-

nia się te próby szukania teoretycznych podstaw Wechslera Skali Inteligencji (por. Kaufman i Lichtenberger, 1999; Lichtenberger i Kaufman, 2009).

Jeśli zaś chodzi o kwestionariusze osobowości, to z trzech, wymienianych przez Bogdana Zawadzkiego (2006, s. 77), strategii budowy kwestionariuszy: teoretycznej (zwanej też dedukcyjną), zewnętrznej (zwanej też kryterialną) i wewnętrznej (zwanej też indukcyjną) zdecydowanie zgodna z podstawowym założeniem, iż to teoria poprzedza operacjonalizację zmiennych, jest strategia pierwsza. Dziś, po opublikowaniu przełomowej w psychometrii pracy, która doprowadziła do uwzględnienia jeszcze jednej, najważniejszej odmiany trafności – *trafności teoretycznej (construct validity)* (Cronbach i Meehl, 1955; też: Cronbach, 1989), nie mamy wątpliwości, iż poza empiryczną teorią psychologiczną test jest tylko bezwartościową podróbką. Odnosząc się do owej klasycznej pracy Lee J. Cronbacha i Paula E. Meehla, Drew Westen i Roberta Rosenthala (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 psychologii. Stanowi istotę każdego badania, w którym badacze używają narzędzia pomiarowego jako wskaźnika zmiennej nie podlegającej bezpośredniej obserwacji (takiej jak np. inteligencja, agresja czy pamięć robocza). Jeśli test psychologiczny (lub szerzej: procedura psychologiczna, w tym również manipulacja eksperymentalna) nie wykazuje trafności teoretycznej, wyniki otrzymane za pomocą tego testu lub procedury będą trudne do zinterpretowania [...].

Podobnie autorzy rozszerzenia *klasycznej teorii testów* uważają, iż trafność teoretyczna jest, jak napisali w swoim fundamentalnym dziele (por. Lord i Novick, 1968, s. 278), „najważniejszą właściwością testu”. Nie mamy wątpliwości, że jest to najważniejszy aspekt ustalania trafności testu.

Mówiąc krótko, test, który nie jest wywiedziony z teorii empirycznej, nie jest godny tego miana. Jednym z uznanych sposobów jej ustalania jest zaproponowana przez Donalda T. Campbella i Donalda W. Fiske’go (Campbell i Fiske, 1959) procedura badania trafności zbieżnej i różnicowej poprzez analizę macierzy „wielu cech – wielu metod” (*convergent and discriminant validation by the multi trait – multi method matrix*).

Rysunek 1 ukazuje, że do empirycznej teorii psychologicznej, która przesądza o trafności teoretycznej testu, dołączają jeszcze dwie, nie-psychologiczne teorie (modele) – *statystyczna* (określająca ramy interpretacji ilościowej wyniku testowego) oraz *psychometryczna*, w ramach której konstruowany jest test i na gruncie której określone są podstawowe parametry *dobroci psychometrycznej* testu – przede wszystkim jego rzetelność i wielkość *błędu standardowego pomiaru*. Dzięki nim możliwe jest budowanie *przedziału ufności* dla wyniku praw-

dziwego (według *klasycznej teorii testów* Harolda O. Gulliksen, 1950). Jego znajomość jest niezbędna dla prawidłowej interpretacji wyniku testowego (por. AERA/APA/NCME, 1999/2007). Niestety, znajomość standardów statystycznych i psychometrycznych nie jest najmocniejszą stroną klinicystów. Znaczna ich część jest przekonana o swojej klinicznej nad-kompetencji i odrzuca podejście statystyczne – tak jakby przed laty nie ukazała się burząca te naiwne przekonania głośna książka Paula E. Meehla: *Clinical versus statistical prediction. A theoretical analysis and a review of the evidence* (Meehl, 1954).

W Fазie III przygotowany jest plan badania empirycznego. Tu niezbędna jest znajomość podstaw statystycznych badania eksperymentalnego i korelacyjnego – np. model ANOVA/MANOVA dla eksperymentalnych badań (z odwołaniem się do zasady randomizacji). I znowu, bez znajomości modeli psychometrycznych i statystycznych niemożliwe będzie poprawne zrealizowanie Fazy IV i V badania.

W Fазie VI badania zinterpretowany wynik (wniosek statystyczny) użytej metody statystycznej (miary korelacji, testy istotności różnic) powinien być przekształcony na wniosek badawczy. Piszę o tym, gdyż nazbyt często badacz zadawała się stwierdzeniem, iż potwierdziła się jego hipoteza o zachodzeniu związku między zmiennymi czy o istotności różnicy między średnimi wartościami zmiennej zależnej w porównywanych grupach: eksperymentalnej i kontrolnej. „Sukcesem” badawczym jest otrzymanie poziomu istotności $p = 0,05$. Dobrze, gdy informacja o wartości p będzie uzupełniona informacją o wartości *wielkości efektu* ES (*effect size*; np. współczynnik d Cohena zastosowany do wyników testu t Studenta), czyli o rzeczywistej sile oddziaływania zmiennej X na zmienną Y (od kilku lat jest to standardowe wymaganie stawiane autorom tekstów empirycznych; por. APA, 2010; też: Wilkinson and Task Force on Statistical Inference, 1999). Wartości ES powinny być zinterpretowane przedziałowo (JARS Group, 2008; Wilkinson, 1999).

Nie stawiam znaku równości między *wnioskiem statystycznym* (wyżej wskazana sekwencja: wartość statystyki testowej, np. „ t Studenta \rightarrow poziom istotności statystycznej, np. nie przewyższający $p = 0,05 \rightarrow$ wartość ES, np. d Cohena”) i *wnioskiem badawczym*. Podzielał stanowisko zajęte przez Bruce’a M. Kinga i Edwarda W. Miniuma (King i Minium, 2009, s. 25) w kwestii podziału wniosków na: *wnioski statystyczne* (*statistical conclusions*) i *wnioski badawcze* (*research conclusions*). Dodałbym jeszcze dwa uzasadnienia dla tego rozróżnienia. Pierwsze: ocena warunków, w których zastosowano dany test statystyczny; czy miały one wpływ na jego wynik. Porównywane próby nie są przecież, tak naprawdę, pobrane według losowania zwrotnego z populacji (a takie jest założenie modelu statystycznego testu istotności). Czy rzeczywiście badacz kontrolował

wszystkie wpływy zewnętrzne, które mogły zniekształcać zachowanie się osób badanych w sytuacji badawczej? Oczywiście, że nie. Czy w badaniach o charakterze eksperymentalnym stosował się do zasady randomizacji? Trzeba zatem wrócić do Fazy III i krytycznie przyrzeć się planowi badania. Przykład: gdy zachodzi uzasadnione podejrzenie, że w badaniu eksperymentalnym może wystąpić *efekt pretestu*, to badanie powinno być przeprowadzone według *planu Solomona*. Drugie: ocena konsekwencji praktycznych zaakceptowanego przez badacza ryzyka popełnienia błędu I⁰ (zawsze w fazie planowania badania, a nie już po jego zakończeniu!). Zbyt rygorystyczny poziom p może sprawić, że „utopimy” ciekawą hipotezę. Z kolei zbyt liberalny poziom p może doprowadzić do upowszechnienia się fałszywego wyniku, który będzie stanowił podstawę dla jakiejś, powiedzmy, procedury terapeutycznej, która może okazać się szkodliwa (np. źle sprawdzony lek, który okaże się toksyczny). Może zatem nie trzymać się sztywno poziomu $p = 0,05$? Może czasami trzeba sięgnąć po $p = 0,001$, a czasem wystarczy $p = 0,10$ (np. w badaniach typu eksploracyjnego)?

Tu też – zwłaszcza w kontekście badań prowadzonych przez klinicystów raczej w warunkach terenowych (np. kliniki) aniżeli w ściśle kontrolowanych warunkach laboratoryjnych – trzeba zaakcentować ważność *trafności zewnętrznej* (zwłaszcza!) i *trafności wewnętrznej* planu badania. Z jednej strony, w trosce o wysoką trafność wewnętrzną, należałoby – w jak największym stopniu – izolować badanie od zakłócających wpływów zewnętrznych i przeprowadzić je w warunkach laboratoryjnych, z odwołaniem się do *zasady randomizacji* (maksymalizacja trafności wewnętrznej). Z drugiej strony taka izolacja, bywa że doprowadzona do perfekcji, czyni je „odrealnionym”, „sztucznym”, o mocno obniżonej trafności zewnętrznej. Wyjście z laboratorium pogarsza trafność wewnętrzną badania, ale poprawia jego trafność zewnętrzną. Jest to jeden z podstawowych problemów badań nad psychoterapią, prowadzonych według dwóch standardów: *efficacy vs effectiveness* (por. np. Hunsley, Elliott i Therrien, 2013).

Niestety, chęć sprostania wysokim standardom metodologicznym, dla zapewnienia badaniu klinicznemu wysokiej trafności wewnętrznej, na ogół nie idzie w parze z uwzględnieniem wysokiej trafności zewnętrznej tego badania. *Trafność wewnętrzna* wymaga dużej precyzji pomiarowej, grup kontrolnych, randomizacji, izolacji od wpływów zewnętrznych itp. Kryteria trafności wewnętrznej spełniają badania typu RCT (*Randomized Controlled Trials*). Z kolei *trafność zewnętrzna* wymaga zbliżenia warunków badania do warunków życiowych i prowadzenia badania na grupach reprezentatywnych. Przykładem niech będą badania poprowadzone przez Martina Seligmana nad skutecznością psychoterapii metodą *Consumer Reports* (Seligman, 1996; krytycznie o tych badaniach:

Nathan, Stuart i Dolan, 2000; Jaworska, 2001). Co z tego, że badania poprowadzone na *grupie reprezentatywnej* (co na swoją obronę podkreślił Seligman) metodą ankiety pocztowej miały stosunkowo wysoką *trafność zewnętrzną*, jeśli ich *trafność wewnętrzną* była nie do zaakceptowania (m.in. brak grupy kontrolnej).

Podsumowując ten wątek „trafnościowy” mojego artykułu, zwróciłbym szczególną uwagę na źródła niskiej trafności badań prowadzonych przez klinicystów:

– *trafność wewnętrzną*: brak grupy porównawczej (kontrolnej), brak randomizacji;

– *trafność zewnętrzną*: niska reprezentatywność próby, niska reprezentatywność warunków badania.

Czy można, aby uzyskać bliższy oczekiwaniom, jeden rodzaj trafności przedłożyć nad drugi – *trafność zewnętrzną* nad *trafność wewnętrzną* – jak to zdają się czynić klinicyści? Odpowiem jak metodolog. Nie można, bo co z tego, że uzyskano wysoką *trafność zewnętrzną* rezultatów badawczych, które są dość wątpliwe, jeśli chodzi o ich *trafność wewnętrzną*. Jednocześnie zdaję sobie sprawę z tego – też jako metodolog – że nie da się w pełni pogodzić wymogu dopełnienia obu warunków jednocześnie. Tak się może zdarzyć jedynie w świecie wyidealizowanym, ale ten, nas otaczający, jest taki, jaki jest. Badacz może jedynie próbować uzyskanie pewnego kompromisu. Może też poświęcić „trochę” jeden rodzaj trafności na rzecz drugiego.

Sumując, empiryczna teoria psychologiczna i „towarzyszące” jej teorie (modele) – statystyczny i psychometryczny – kształtują badanie naukowe jako:

- źródło definicji zmiennych,
- podstawa operacjonalizacji zmiennych (nadawania im sensu empirycznego),
- rama interpretacji ilościowej (statystycznej) zebranych danych,
- rama interpretacji psychologicznej wyniku badawczego (traktowanego jako zagregowane dane empiryczne).

Od empirycznej teorii psychologicznej do EBA i EBPP

Od lat 90. ubiegłego wieku, wzorując się na medycynie², kładzie się szczególny nacisk na to, aby formułowane przez psychologów diagnozy oparte były na dowodach empirycznych (*Evidence-Based Assessment*, EBA; por. Stemplew-

² W kwietniu 2015 r. brałem udział w *debacie* (poprowadzonej według formuły *Oxford Union Debate*: „This house believes that...”) zorganizowanej przez redakcję czasopisma *Nowotwory. Journal of Oncology*. Ich uczestnicy zostali poinformowani, aby prezentowali swoje stanowiska: „[...] w duchu evidence-based medicine, tj. na podstawie wiarygodnych i aktualnych badań naukowych [...]”.

ska-Żakowicz i Paluchowski, 2008; Stemplewska-Żakowicz, 2009; Paluchowski, 2010). Ten model postępowania diagnostycznego wpisał się w ogólniejszy model praktyki psychologicznej opartej na dowodach empirycznych (*Evidence-Based Practice in Psychology*, EBPP), który *American Psychological Association* w 2006 r. (por. APA, 2006) zdefiniowała następująco: „*Evidence-based practice in psychology (EBPP) is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences*” (s. 273).

Tak naprawdę to w polskiej psychologii klinicznej, w dużym stopniu formowanej przez prace Andrzeja Lewickiego (lata 60. ubiegłego wieku – por. Lewicki, 1969), od dawna wzorowano diagnozę na procesie badawczym, a więc z dominującą rolą teorii i rzetelnej metodologii (por. np. Brzeziński i Kowalik, 1991; Kowalik i Brzeziński, 1991a, 1991b). Pisał przed laty założyciel pierwszej polskiej Katedry Psychologii Klinicznej na Uniwersytecie im. Adama Mickiewicza w Poznaniu, A. Lewicki (1969, s. 84):

Z metodologicznego punktu widzenia diagnostyczne **postępowanie psychologa klinicznego należy określić jako formę badania naukowego**, zastosowanego do rozwiązania praktycznych problemów klinicznych. Istotą każdego badania naukowego jest: a) postawienie problemu, b) wysunięcie hipotezy, tj. prawdopodobnego rozwiązania problemu, c) weryfikacja hipotezy przez zastosowanie odpowiednich metod badawczych. Badanie kliniczne spełnia te postulaty [wyróżn. – J. M. B.].

Czy nie mamy tu do czynienia z wizją EBA? Owszem, zgadzam się z modelem EBA, jeśli chodzi o standard postępowania diagnostycznego, a szczególnie jeśli chodzi o eliminowanie z instrumentarium psychologów (zwłaszcza tych klinicystów, którzy jak zaczarowani przywiązali się do szamańskich praktyk diagnostycznych odwołujących się do takich pseudotestów psychologicznych, jak Test Szondiego, Test Kolorów Lüschera czy Test Drzewa Kocha). Podpisuję się pod rzetelną krytyką tych „testów” przeprowadzoną przez Katarzynę Stemplewską-Żakowicz (2009); trzeba zauważyć, że te trzy „testy” nie wyczerpują owej czarnej listy naukowo nieuprawnionych narzędzi diagnostycznych, z upodobaniem stosowanych przez sporą część społeczności klinicystów. Ważnym uzupełnieniem tej listy mogłaby być druga czarna lista procedur, za pomocą których niedouczeni czy goniący za tanim zyskiem (np. przeróżne „gabinety psychologiczne”) paraklinicyści oddziałują na swoich naiwnych klientów; tu znalazłoby się miejsce dla *ustawień rodzinnych* Berta Hellingera, NLP czy przeróżnych magicznych praktyk psychoanalitycznych. Nie uważam, aby spełniały one kryteria EBPP.

Władysław J. Paluchowski (2010, s. 11), charakteryzując jedną z istotnych osobliwości modelu EBA, pisze, że odrzuca się w nim założenie, iż:

[...] obserwacje pochodzące z praktyki klinicznej (własnej lub innych osób) stanowią wiarygodne i wystarczające źródło praktycznej wiedzy medycznej. Osobiste doświadczenie i praktyka są przeceniane co do ich wartości, zawsze bowiem mają one charakter lokalny, ograniczony do przypadków niereprezentatywnych dla wszystkich pacjentów, ale – kontekstowo – dobranych (próba obciążona, stronicza), a ich ocena zależy od specyficznych cech podmiotowych lekarza (rodzaj wykształcenia czy preferencja poznawcza). Ponadto bazowanie na doświadczeniu autorytetów częściej ma za podstawę wiarę niż mądrość, która zastępowana bywa przez utrwalone stereotypy i modę (Sułkowski, 2007).

Moim zdaniem odnosi się to też do tradycyjnej praktyki klinicystów.

Chciałbym jednak, aby znacząco wyraziściej zaakcentowano rolę empirycznej teorii psychologicznej, gdyż to ona, przede wszystkim, przesądza o wartości diagnozy psychologicznej (także klinicznej). To należyte wyeksponowanie teorii znalazło się w przygotowanych przez Sąd Najwyższy USA tzw. *kryteriach Dauberta*, które są stosowane wobec ekspertyz sądowych w sądownictwie amerykańskim³. Owe siedem kryteriów Dauberta (*seven Daubert guidelines*; za: Ritzler, Erard i Pettigrew, 2002, s. 202-203) to:

- 1) Czy proponowana teoria (lub technika), na której ma być oparte zeznanie, jest sprawdzalna?
- 2) Czy proponowana teoria (lub technika) została przetestowana przy użyciu trafnych i rzetelnych procedur, z pozytywnym rezultatem?
- 3) Czy proponowana teoria (lub technika) została poddana recenzji?
- 4) Co wiadomo na temat oszacowania potencjalnej wielkości błędu w przypadku tej teorii lub techniki?
- 5) Jakie standardy, służące do kontroli działania proponowanej techniki, maksymalizują jej trafność?
- 6) Czy proponowana teoria (lub technika) została ogólnie zaakceptowana jako trafna przez właściwe środowisko naukowe? (Grove i Barden, 1999, s. 226)
- 7) [Dodane później] Czy wnioski eksperta wynikają w uzasadniony sposób z zastosowania w danym przypadku tej teorii (lub techniki)? (Grove i Barden, 1999, s. 226) [wyróżn. – J. M. B.].

Zauważmy, że terminem najczęściej się powtarzającym jest „teoria”. Jeżeli klinicyści sporządzający ekspertyzy (np. dla sądów: orzekanie o poczytalności, seksuologia, rozwody itp.) nie będą nadać za rozwojem teoretycznym i metodologicznym psychologii (teorie i narzędzia), to skazują się na ośmieszenie (np.

³ Jasson Daubert i Eric Schuler wystąpili z pozwem sądowym przeciwko koncernowi farmaceutycznemu Merrell Dow Pharmaceuticals Inc., gdyż uznali, że na skutek brania przez ich matki w okresie ciąży leku *Bendectin* urodzili się z fizycznymi uszkodzeniami ciała. Sąd postanowił w tym postępowaniu, aby sporządzane przez specjalistów ekspertyzy były zgodne z kryteriami, nazwanymi kryteriami Dauberta właśnie. Te kryteria stały się zaleceniami dla sędziów w sądownictwie amerykańskim.

trzymając się kurczowo testu Kocha) i wykluczenie z poważnych debat i profesjonalnych zleceń (choć niekoniecznie z łatwych zarobków).

Konkluzja

Praktyka badawcza psychologów klinicznych (naukowa i diagnostyczna) jeszcze nie w pełni oparta jest na standardach metodologicznych empirycznej psychologii, w której nie ma miejsca na niewiele mające wspólne z nauką spekulacje samozwańczo mianujące się „teoriami” – często o proveniencji psychoanalitycznej (a ta, jak pokazał K. Popper, odstaje od racjonalnego modelu nauki – nie jest *falsyfikowalna*).

O naukowości klinicznego badania naukowego (też diagnostycznego) (jak starałem się to ukazać, omawiając Rysunek 1 – eksponującą rolę: empirycznej teorii psychologicznej, modeli psychometrycznych i modeli statystycznych) przesądza w dużej mierze oparcie się na empirycznie sprawdzonej teorii psychologicznej.

Praktyka diagnostyczna odwołuje się dziś do standardów EBA; odwołuje się też (zob. kryteria Dauberta) do sprawdzonych teorii i poprawnie skonstruowanych, dostarczających *replikowanych* wyników (to też dość skuteczna zaporą przed oszustwami naukowymi) narzędzi. W ich zbiorze nie ma miejsca na przekonanie o swoich szczególnych kompetencjach wynikających z lat praktyki czy osobiste odczucia i wrażenia z kontaktu z osobami badanymi jako źródła rzetelnych danych empirycznych (por. listę *Multiple Types of Research Evidence*, zamieszczoną w raporcie APA Presidential Task Force on Evidence-Based Practice, 2006, s. 274).

Doskonalenie praktyki badawczej klinicyistów – to rozwijanie rzetelnego warsztatu i pozyskiwanie replikowalnych danych.

LITERATURA CYTOWANA

- 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.
- APA, American Psychological Association (2010). *Publication manual of the American Psychological Association* (wyd. 6). Washington, DC: Author.
- APA, APA Presidential Task Force on Evidence-Based Practice (2006). Evidence-Based Practice in Psychology. *American Psychologist*, 61(4), 271-285.
- Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, 24(3), 383-405.

- Brzeziński, J. (1994). Dimensions of diagnostic space. W: J. Brzeziński (red.), *Probability in theory-building. Experimental and non-experimental approaches to scientific research in psychology* (s. 197-223). Amsterdam–Atlanta, GA: Rodopi. „Poznań Studies in the Philosophy of the Sciences and the Humanities”, t. 39.
- Brzeziński, J. (1996). Theory and social practice. One or two psychologies? W: A. Zeidler-Janiszewska (red.), *Epistemology and history. Humanities as a philosophical problem and Jerzy Kmity's approach to it* (s. 351-364). Amsterdam–Atlanta, GA: Rodopi. „Poznań Studies in the Philosophy of the Sciences and the Humanities”, t. 43.
- Brzeziński, J. (2013). Methodological awareness and ethical awareness in the context of university education (on the example of psychology). W: B. Bokus (red.), *Responsibility. A cross-disciplinary perspective* (s. 261-277). Warszawa: Lexem.
- Brzeziński, J. i Kowalik, S. (1991). Diagnoza kliniczna w kontekście praktyki społecznej. W: H. Sęk (red.), *Spółeczna psychologia kliniczna* (s. 187-212). Warszawa: Państwowe Wydawnictwo Naukowe.
- Brzeziński, J. i Toeplitz-Winiewska, M. (2008). Model zawodowy psychologa klinicznego. W: H. Sęk (red.), *Psychologia kliniczna* (wyd. 2, t. 1, s. 299-324). Warszawa: Wydawnictwo Naukowe PWN.
- Campbell, D. i Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105; polski przekład w: J. Brzeziński (red.) (2005), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 431-459). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Cronbach, L. J. (1989). Construct validation after 30 years. W: R. L. Linn (red.), *Intelligence. Measurement theory and public policy* (s. 147-171). Champaign: University of Illinois Press.
- Cronbach, L. i Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302; polski przekład w: J. Brzeziński (red.) (2005), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 403-430). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Ellis, H. C. (1992). Graduate education in psychology. Past, present, and future. *American Psychologist*, 47(4), 570-576.
- Frank, G. (1984). *Wechsler enterprise*. Oxford: Pergamon Press.
- Grove, W. M. i Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under Daubert/Kumhoanalyses. *Psychology, Public Policy, and Law*, 5(1), 224-242.
- Gulliksen, H. O. (1950). *Theory of mental tests*. New York: J. Wiley.
- Hornowska, E. (1988). Skala inteligencji Davida Wechslera w świetle teorii inteligencji. W: J. Brzeziński i E. Hornowska (red.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (wyd. 2, s. 21-48). Warszawa: Państwowe Wydawnictwo Naukowe.
- Hunsley, J., Elliott, K. i Therrien, Z. (2013). The efficacy and effectiveness of psychological treatments, http://www.cpa.ca/docs/File/Practice/TheEfficacyAndEffectivenessOfPsychologicalTreatments_web.pdf (Pozyskano: 3 maja 2015).
- Jacob, F. (1973). *Historia i dziedziczność*. Warszawa: Państwowy Instytut Wydawniczy.
- JARS Group, APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology. Why do we need them? What might they be? *American Psychologist*, 63(9), 839-851.
- Jaworska, A. (2001). Badanie efektywności psychoterapii metodą Consumer Reports: Analiza metodologii. *Czasopismo Psychologiczne*, 7(2), 227-240.

- Kaufman, A. S. i Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.
- Kaufman, M. (2007). In memoriam: The idioverse of Saul Rosenzweig (1907-2004). *Journal of Psychotherapy Integration*, 17(4), 363-368.
- King, B. M. i Minium, E. W. (2009). *Statystyka w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe PWN.
- Kowalik, S. (1988). Diagnostyczne i kliniczne właściwości Skali Inteligencji: W-B I/II, WAIS, WAIS-R. W: J. Brzeziński i E. Hornowska (red.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (wyd. 2, s. 386-462). Warszawa: Państwowe Wydawnictwo Naukowe.
- Kowalik, S. i Brzeziński, J. (1991a). Diagnoza kliniczna. W: H. Sęk (red.), *Spoleczna psychologia kliniczna* (s. 213-239). Warszawa: Wydawnictwo Naukowe PWN.
- Kowalik, S. i Brzeziński, J. (1991b). Protodiagnoza kliniczna. W: H. Sęk (red.), *Spoleczna psychologia kliniczna* (s. 240-254). Warszawa: Wydawnictwo Naukowe PWN.
- Kuhn, T. S. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.
- Leckliter, I. N., Matarazzo, J. D. i Silverstein, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. *Journal of Clinical Psychology*, 42, 322-342.
- Lewicki, A. (1969). Psychologia kliniczna w zarysie. W: A. Lewicki (red.), *Psychologia kliniczna* (s. 9-155). Warszawa: Państwowe Wydawnictwo Naukowe.
- Lichtenberger, E. O. i Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. New York: Wiley.
- Lord, F. M. i Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Matarazzo, J. D. (1987). There is only one psychology, no specialties, but many applications. *American Psychologist*, 42, 893-903.
- Meehl, P. E. (1954). *Clinical versus statistical prediction. A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.
- Miller, A. G. (red.) (1972). *The social psychology of psychological research*. New York: The Free Press.
- Nathan, P. E., Stuart, S. P. i Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin*, 126(6), 964-981.
- Paluchowski, W. J. (2010). Diagnoza oparta na dowodach empirycznych – czy potrzebny jest „polski Buros”? *Roczniki Psychologiczne*, 13(2), 7-27.
- Rapaport, D. A. (1945). *Diagnostic psychological testing* (t. 1). Chicago, IL: Yearbook Publ.
- Ritzler, B., Erard, R. i Pettigrew, G. (2002). Protecting the integrity of Rorschach expert witnesses. A Reply to Grove and Barden (1999) Re: The Admissibility of Testimony Under *Daubert/Kumho* Analyses. *Psychology, Public Policy, and Law*, 8(2), 201-215.
- Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review*, 40(4), 337-354.
- Seligman, M. E. P. (1996). Skuteczność psychoterapii. Badania ankietowe. *Czasopismo Psychologiczne*, 2(2), 101-112.
- Silverman, D. (2007). *Interpretacja danych jakościowych. Metody analizy rozmowy, tekstu i interakcji*. Warszawa: Wydawnictwo Naukowe PWN.
- Spearman, C. E. (1927). *The abilities of man*. New York: Macmillan.
- Stemplewska-Żakowicz, K. (2009). *Diagnoza psychologiczna. Diagnostowanie jako kompetencja profesjonalna*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Stemplewska-Żakowicz, K. i Paluchowski, W. J. (2008). Podstawy diagnozy psychologicznej. W: J. Strelau i D. Doliński (red.), *Psychologia. Podręcznik akademicki* (t. 2, s. 23-63). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

- Sułkowski, H. (2007). *Co lekarz praktyk o medycynie opartej na dowodach wiedzieć powinien*. Zaczepnięte 3 maja 2015. Strona internetowa: <http://www.umed.lodz.pl/kmz/doc/ebm.pdf>
- Tulsky, D. S., Saklofske, D. H., Chelune, G. J., Heaton, R. K., Ivnik, R. J., Bornstein, R., Prifitera, A. i Ledbetter, M. F. (red.) (2003). *Clinical interpretation of the WAIS-III and WMS-III*. San Diego, CA: Academic Press.
- Wechsler, D. (1939/1998). Definicja i natura inteligencji. W: J. Brzeziński i E. Hornowska (red.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (wyd. 2, s. 15-20). Warszawa: Wydawnictwo Naukowe PWN.
- Wechsler, D. (1941/1998). Diagnostyczne i kliniczne właściwości Skali Wechsler-Bellevue. W: J. Brzeziński i E. Hornowska (red.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (wyd. 2, s. 363-385). Warszawa: Wydawnictwo Naukowe PWN.
- Westen, D. i Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, 84(3), 608-618.
- Wilkinson, L. and Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.
- Zawadzki, B. (2006). *Kwestionariusze osobowości. Strategia i procedura konstruowania*. Warszawa: Wydawnictwo Naukowe Scholar.
- Zimmerman, I. L. i Woo-Sam, J. M. (1973). *Clinical interpretation of the Wechsler Adult Intelligence Scale*. New York: Grune & Stratton.