

JERZY BRZEZIŃSKI
Instytut Psychologii UAM

O WAŻNOŚCI PRETESTU ZMIENNEJ ZALEŻNEJ W EWALUACJI PROGRAMÓW SPOŁECZNYCH

WPROWADZENIE

Przed dziesięciu laty opublikowałem w ostatnim, redagowanym przez Profesora Jana Lutyńskiego numerze *Przeglądu Socjologicznego*, artykuł: *Eksperymentalne badanie zmian – metodologiczne problemy stosowania pretestu zmiennej zależnej* (Brzeziński 1989). Dziś chciałbym wrócić do tego, jak mnie mam, ważnego dla praktyki badawczej (nie tylko socjologów) zagadnienia.

Określanie poziomu wyjściowego zmiennej zależnej (inaczej: pomiaru początkowego zmiennej zależnej, albo *pretestu* zmiennej zależnej – dalej będę posługiwał się tym ostatnim określeniem) jest szczególnie ważne dla poprawnej oceny skuteczności różnorodnych programów społecznych (profilaktycznych, edukacyjnych, promocyjnych itp.). Jego pomijanie, jako prosty sposób radzenia sobie z trudnościami metodologicznymi, na które naraża badacza posługiwanie się narzędziami pomiarowymi uwrażliwiającymi osoby badane (w sensie: Cook, Campbell 1979; Brzeziński 1997; Webst 1997) i tym samym obniżającymi *trafność wewnętrzną* i *zewnętrzną* (w sensie: Cook, Campbell 1979; Brzeziński 1997; Webst 1997) samego badania nie jest godnym polecenia sposobem postępowania.

Moim zdaniem nie można rezygnować z pretestu zmiennej zależnej, gdyż może to doprowadzić do różnych nadużyć interpretacyjnych końcowego rezultatu ewaluacji jakiegoś programu społecznego. W przedmowie do jednego z podręczników ewaluacji programów profilaktycznych (Hawkins, Nederhoad 1994), napisanej przez redaktorów naukowych przekładu czytamy (Grzelak, Wolniewicz-Grzelak 1994):

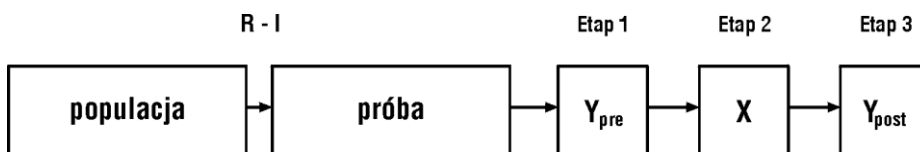
„wprowadzaniu w życie nowych programów społecznych coraz częściej towarzyszą pytania o to, jaka jest skuteczność tych programów, czy prowadzą one do osiągnięcia celów, dla których zostały stworzone, czy nakłady ponie-

sione przy ich realizacji są współmierne do ich społecznych efektów... Pytania te zadają i ci, którzy za realizację programów płacą, i ci, którzy te programy realizują”.

Aby na te pytania odpowiedzialnie odpowiedzieć, niezbędne jest odpowiednie przygotowanie metodologiczne postępowania ewaluacyjnego. Pocześnie miejsce zajmuje w nim przeprowadzenie *pretestu zmiennej zależnej*.

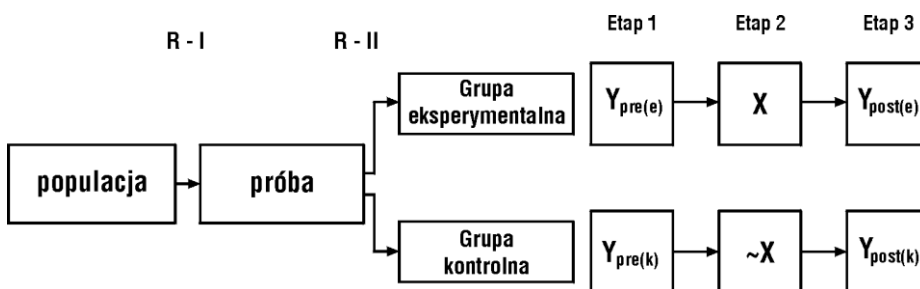
PRETEST ZMIENNEJ ZALEŻNEJ – PROBLEMY METODOLOGICZNE

Badania eksperymentalne i quasi-eksperymentalne dają się ująć w jeden z dwóch schematów. Wedle pierwszego, badacz rejestruje jakiś stan początkowy interesującego go zjawiska (pretest – np. poziom lęku grupy pacjentów depresyjnych), następnie próbuje oddziaływać na to zjawisko (np. za pomocą psychoterapii ukierunkowanej na redukcję poziomu lęku), a badanie kończy rejestracja stanu końcowego (posttest – np. pomiar poziomu lęku pacjentów po przeprowadzeniu psychoterapii) – tak jak to pokazano na Ryc. 1.



Ryc. 1

Z kolei wedle drugiego schematu badacz dzieli (losowo!) próbę osób – *reprezentatywną* dla danej populacji – na dwie grupy (tak jak to pokazano na Ryc. 2).



Ryc. 2

Po przeprowadzeniu pretestu zmiennej zależnej (np. postawy wobec jakiegoś obiektu) badacz w sposób zróżnicowany traktuje osoby z obu grup. Jedna grupa poddana jest specjalnym oddziaływaniom (np. stosuje się wobec niej techniki indoktrynacji). Druga grupa poddana jest oddziaływaniom neutralnym względem interesującej badacza postawy. Badanie kończy ponowny pomiar (posttest) natężenia zmiennej zależnej.

Na ogół badania nastawione są na znalezienie odpowiedzi na pytanie o istotność różnicy między pierwszym i drugim pomiarem natężenia zmiennej zależnej, czyli między posttestem i pretestem. Chodzi tedy o stwierdzenie, czy – aby odwołać się do powyższych przykładów – nastąpiła redukcja poziomu lęku pacjentów depresyjnych pod wpływem stosowania wobec nich psychoterapii, czy też chodzi o sprawdzenie skuteczności danej techniki indoktrynacji zastosowanej wobec danej grupy osób przejawiających określoną postawę wobec jakiegoś obiektu. Na ową skuteczność wskazuje tu wielkość zmiany postawy zarejestrowana przez porównanie pretestu i posttestu w pierwszej grupie, jak i porównanie wielkości zmiany w drugiej grupie. Badacz może też porównać dwie (i więcej) grupy traktowane w zróżnicowany sposób (np. dwa konkurencyjne sposoby leczenia, dwa konkurencyjne programy profilaktyczne). Wówczas szuka on odpowiedzi na pytanie o to, który ze sposobów leczenia, czy sposobów przeciwdziałania wystąpieniu niepożądanych zjawisk społecznych (narkomanii, alkoholizmu wśród młodzieży szkolnej, nietolerancji wobec mniejszości narodowych itp.) jest efektywniejszy.

Stosowanie pretestu zmiennej zależnej w badaniach psychologicznych, socjologicznych, pedagogicznych, psychiatrycznych, a więc w badaniach, w których „obiektem badanym” jest człowiek, którego bada inny człowiek, rodzi swoiste problemy metodologiczne (a także etyczne), których źródłem jest interakcja: „badacz – osoba badana”. To ona, poprzez takie zmienne jak: lęk przed oceną (Rosenberg 1991), wskazówki sugerujące hipotezę badawczą (Orne 1993) nastawienia interpersonalne badacza (Rosenthal 1991), aprobata społeczna (Nederhof 1991) zniekształca wyniki badania, czyniąc je nie reprezentatywnymi dla danej populacji. Każdy badacz, a już w szczególności psycholog, socjolog czy pedagog nie może nie dostrzegać owych zniekształceń. Co więcej jest zobowiązany do takiego postępowania metodologicznego, które – jeżeli nie wyeliminuje – to przynajmniej zminimalizuje wariancje błędu, której niesystematycznym źródłem jest interakcja: „badacz – osoba badana”. Niestety, także i pretest zmiennej zależnej, którego nie sposób pominąć, jest źródłem kłopotów metodologicznych. Badacz staje tedy przed trudnym pro-

blemem, jak postąpić, aby nie tracąc ważnych informacji, które niesie pretest jednocześnie pominąć jego dwojaki destrukcyjny wpływ na wyniki posttestu:

1. Osoby poddane pretestowi będą inaczej reagowały na zmienną niezależną, aniżeli osoby w populacji, wobec których zechcemy zastosować – wypróbowany w tym badaniu – program oddziaływania (zmienna niezależna). Pretest może „pomagać” zmiennej niezależnej w „dotarciu” do osób badanych. W szczególności może się okazać, że tak naprawdę, to *sam* pretest (treść pytań kwestionariusza czy skali postaw) jest źródłem wariacji zmiennej zależnej, a badacz mylnie sądzi, iż zaobserwowana – poprzez porównanie posttestu z pretestem – zmiana wyników (postaw, wiedzy) jest rezultatem testowanego przez niego programu oddziaływania na osoby badane. W takim przypadku będziemy mówili o *interakcji*: „*pretest – zmienna niezależna*”, jako o istotnym źródle wariacji zmiennej zależnej. Jej destrukcyjny (w sensie metodologicznym) wpływ na rezultat badania odnosi się do jego *trafności zewnętrznej* (w sensie: Cook, Campbell 1979; Brzeziński 1997) czyniąc go *niereprezentatywnym*.
2. Dwukrotne posłużenie się tym samym narzędziem pomiarowym (pretest + posttest) może zniekształcić wyniki pomiaru końcowego, gdyż osoby badane mogą zapamiętać odpowiedzi udzielone w pierwszym badaniu i – w zależności od rozpoznanego celu badania i nastawienia wobec badacza i samego badania – dostosować do nich odpowiedzi udzielane w postteście. Oczywiście możliwe jest metodologicznie eleganckie wybrnięcie z tej trudnej sytuacji przez posłużenie się tzw. *formami równoległymi* stosowanych narzędzi (ich równoważność można sprawdzać za pomocą znanego wzoru „połówkowego” Spearmana-Browna, albo poprzez prostą korelację ich wyników przy kontroli charakterystyk psychometrycznych poszczególnych pozycji obu równoległych form narzędzi). Także i w tym przypadku może wystąpić zjawisko *niereprezentatywności* rezultatu badawczego (jego *przeszacowanie* lub *niedoszacowanie*).

Wymienione wyżej problemy metodologiczne, aczkolwiek poważne, nie stanowią dziś przeszkody dla doświadczonego badacza i może on spokojnie posłużyć się pretestem w badaniu skuteczności programów ewaluacyjnych. W moim artykule: *Eksperymentalne badanie...* (Brzeziński 1989) podałem dwa profesjonalne rozwiązania tego problemu – odwołanie się do wielozmiennowych modeli statystycznych, takich, jak: analiza kowariancji (ANCOVA) i analiza wielokrotnej regresji liniowej (MLR) odniesionych do danych zebranych w badaniu przeprowadzonym według tzw. *czterogrupowego*

planu Solomona. Zaletą opisanego tam postępowania metodologicznego jest stosowanie pretestu, połączone z pełną kontrolą wariancji cząstkowej zmiennej zależnej, której źródłem jest ów pretest.

ZNACZENIE PRETESTU ZMIENNEJ ZALEŻNEJ

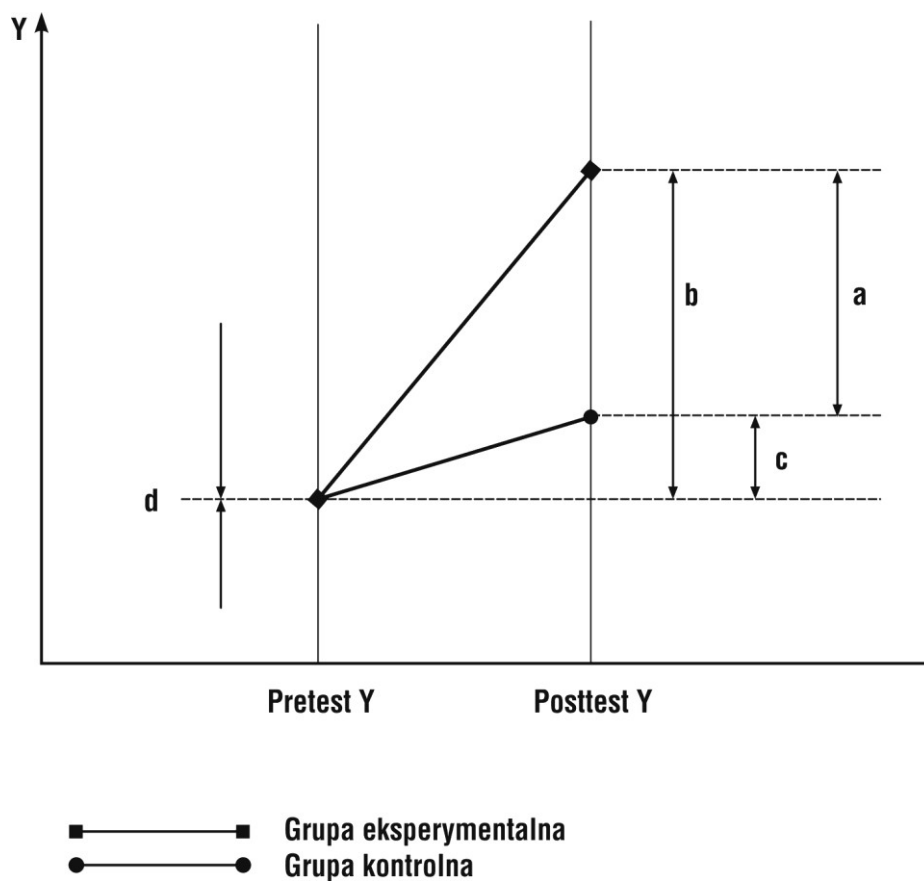
Stosowanie pretestu Y w badaniach nad skutecznością programów ewaluacyjnych spełnia trzy, jak sądzę, ważne role:

1. Badacz uzyskuje informacje o poziomie zmiennej zależnej przed wprowadzeniem działań wynikających z testowanego programu ewaluacyjnego. Przez przyrównanie tego poziomu, do poziomu zmiennej zależnej, określonego po ustaniu działań, badacz może określić *wielkość zmiany* jaka zaszła u konkretnej osoby badanej poddanej działaniom wynikającym z konkretnego programu ewaluacyjnego.
2. Określenie poziomu wyjściowego zmiennej Y – w obu grupach jednocześnie – pozwoli badaczowi rozwiązać ewentualne wątpliwości co do „równego startu” grup – eksperymentalnej i kontrolnej.
3. Badacz może określić relatywną wielkość przeciętnej zmiany zarejestrowanej w grupie eksperymentalnej wobec wielkości przeciętnej zmiany zaobserwowanej w grupie kontrolnej. Hipoteza robocza powinna przyjmować postać, w której „wpisany” będzie relatywizm porównań. Taka prawidłowo sformułowana hipoteza musi się opierać na porównaniu dwóch wyników różnicowych: (a) różnicy posttestu i pretestu w grupie eksperymentalnej z (b) różnicą posttestu i pretestu w grupie kontrolnej. Mówiąc inaczej, zarówno w grupie eksperymentalnej jak i w kontrolnej oczekujemy wystąpienia zmiany, tyle tylko, że będzie ona *relatywnie* większa w grupie eksperymentalnej, aniżeli w grupie kontrolnej.

Co by się jednak mogło stać, gdyby badacz nie przeprowadził pretestu Y w obu grupach? Omówmy je teraz w dwóch do siebie podobnych grupach. Dość typowe sytuacje przedstawione zostały na Ryc. 3.1 – 3.2 (sytuacja I) oraz na Ryc. 3.3 – 3.8 (sytuacja I).

SYTUACJA I (Ryc. 3.1 – 3.2)

Zacznijmy od analizy Ryc. 3.1 i 3.2. Przedstawia one bowiem najbardziej typowy układ wyników, którego wystąpienia spodziewa się na ogół każdy badacz.

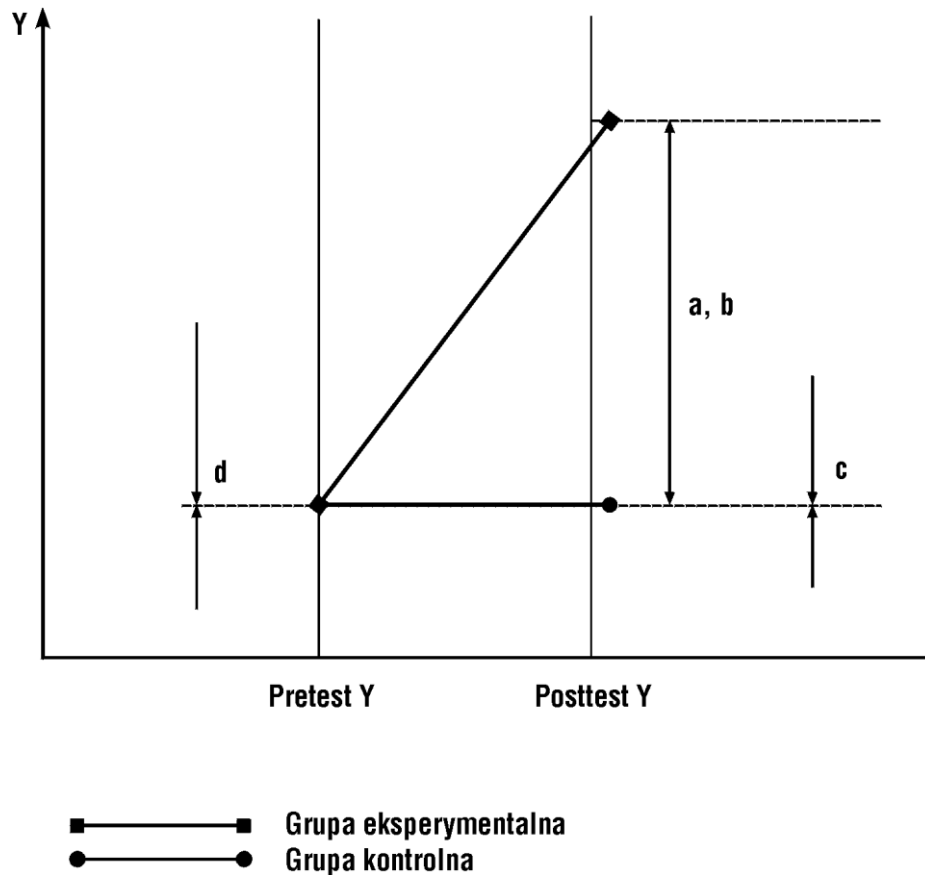


Ryc. 3.1

Jeżeli grupy były rzeczywiście równoważne („zadziałała” randomizacja i nie wystąpił *efekt czynnika selekcji*), to pretesty Y przeprowadzone w obu grupach: eksperymentalnej i kontrolnej powinny dać ten sam wynik, który daje się wyrazić pod postacią, np. równości średnich w porównywanych grupach; ilustruje to – na Ryc. 3.1 i 3.2 – zerowa wartość wskaźnika d , co graficznie zostało przedstawione jako połączenie graficznych symboli pretestów obu grup.

Po przeprowadzeniu zróżnicowanego postępowania wobec obu grup – „nowatorskiego” (o nieznanym skutecznym) w grupie eksperymentalnej i „tradycyjnego” (o znanym skutecznym) można się spodziewać, z dużą dozą prawdopodobieństwa, że wystąpi układ wyników posttestów Y, taki, jak

przedstawiony na obu omawianych rycinach. Dla jego zrozumienia niezbędne jest przeprowadzenie porównania trzech wskaźników: a , b , c .



Ryc. 3.2

Wskaźnik a informuje badacza o względnej skuteczności oddziaływań przeprowadzonych w obu grupach; im wyższą uzyskuje on wartość, tym większe oznacza on zróżnicowanie obu grup pod względem wyników posttestu Y , a o tym przecież traktują testowane hipotezy. Zauważmy, że $a = b - c$; to znaczy: o odczytanej przez badacza, na podstawie układu wyników przeprowadzonych w obu grupach posttestów Y rzeczywistej „przewadze” oddziaływania eksperymentalnego na osoby w grupie eksperymentalnej nad oddziaływaniem standardowym na osoby w grupie kontrolnej świadczy układ wartości wskaźników

b i c : $b > c$ oraz: $b > d$ i $c > d$. Jeżeli $d = 0$, jak w obu przykładach, to badacz może być pewien, że wskaźnik a informuje go rzetelnie o względnej skuteczności – ujętej w testowanej hipotezie – postępowania eksperymentalnego.

Opisane tu sytuacje różnią się pod jednym względem – ta przedstawiona na Ryc. 3.1 bliższa jest *sytuacjom życiowym*, a ta ukazana na Ryc. 3.2 bliższa jest *sytuacjom laboratoryjnym*. Jest bowiem w „życiu” tak, że jeżeli sprawdzamy skuteczność jakiegoś programu ewaluacyjnego, to (może poza jakimiś dość nietypowymi przypadkami), to wyodrębniona przez badacza grupa kontrolna też podlega zmianom (przecież pacjenci leczeni dotychczasowymi metodami, czy dzieci nauczane tradycyjnie też podlegają zmianom), a nie tylko grupa eksperymentalna. Badacz dobrze o tym wie i dlatego formułuje hipotezy kierunkowe postaci:

$$[Y_{\text{post}(e)} - Y_{\text{pre}(e)}] > [Y_{\text{post}(k)} - Y_{\text{pre}(k)}] \text{ albo inaczej: } b > c;$$

w skrajnym przypadku (takim jak na Ryc. 3.2):

$$c = 0 \text{ oraz: } a = b.$$

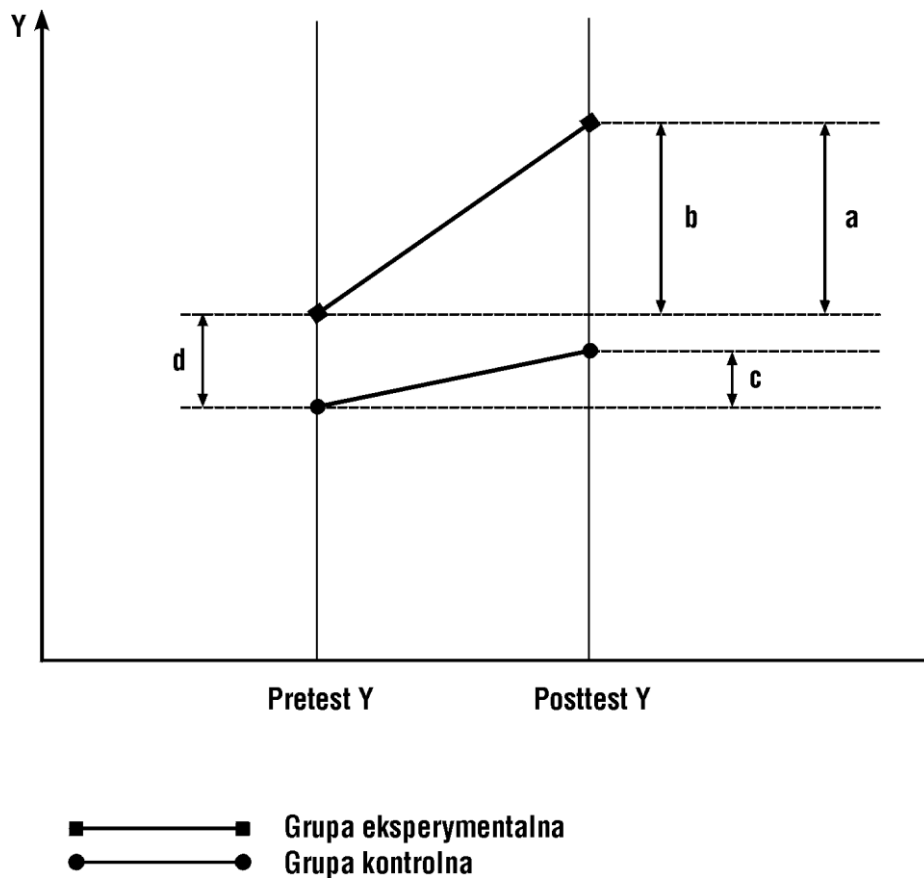
Jeżeli badacz zrezygnowałby z przeprowadzenia pretestu Y , to nie miałby podstaw do stwierdzenia, że $c = 0$. Zatem, zaobserwowana różnica posttestów (a : $Y_{\text{post}(e)} - Y_{\text{post}(k)}$) odpowiada *rzeczywistej*, względnej skuteczności testowanych programów ($b - c$). Tak jest jednak jedynie w sytuacji przedstawionej na Ryc. 3.2:

$$c = 0 \text{ oraz: } a = b - c = b - 0 = b$$

Jeżeli jednak badacz zrezygnuje z przeprowadzenia pretestu, to – co widać wyraźnie, gdy porównamy Ryc. 3.1 z Ryc. 3.2 – z faktu, że $[Y_{\text{post}(e)} - Y_{\text{post}(k)}] > 0$ wcale nie wynika fakt, że (gdyby przeprowadzone były odpowiednie pretesty, a nie były!): $d = 0$. Jest to trudna, z metodologicznego punktu widzenia, sytuacja. Badacz – z czego powinien sobie zdawać sprawę – może dokonać *niedoszacowania* rzeczywistej wartości b (o wartość c). Tedy i trafność zewnętrzna przeprowadzonej skuteczności testowanego programu społecznego będzie w jakimś stopniu (w jakim? – nie sposób to określić) zaburzona.

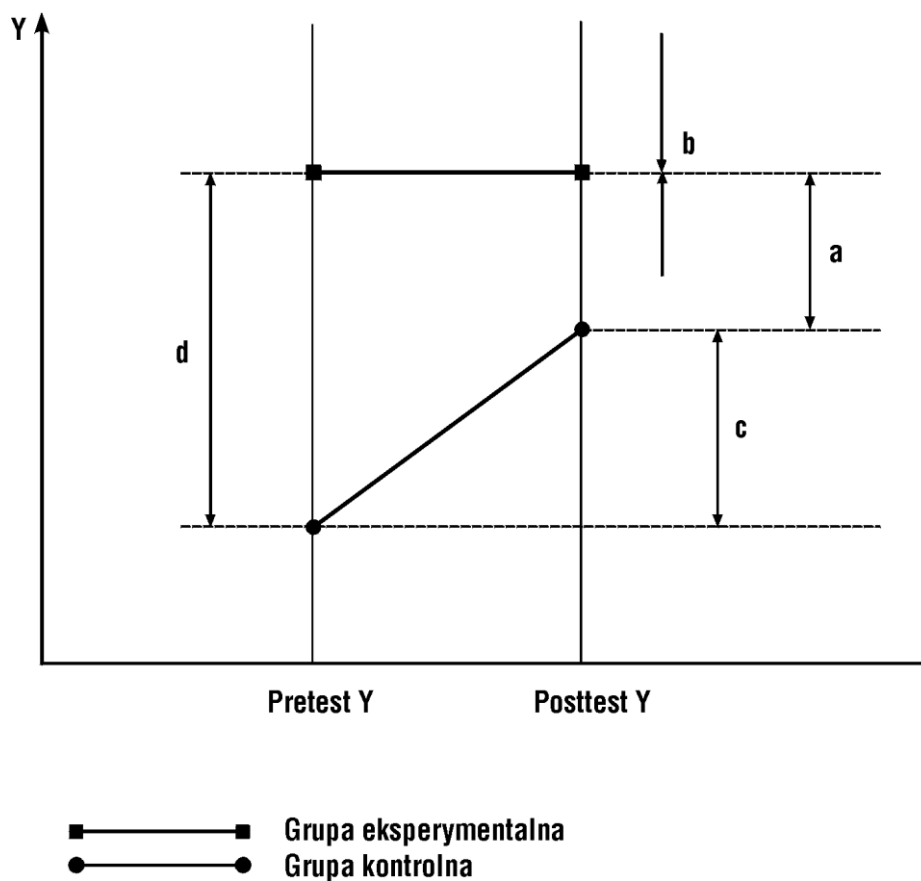
SYTUACJA II (ryc. 3.3 – 3.8)

Zauważmy, na początek, że sześć wariantów sytuacji II. jest pod jednym tylko względem do siebie podobnych – $d > 0$. Gdyby badacz nie przeprowadził w obu grupach pretestu Y, to samo tylko porównanie posttestów Y prowadziłoby do ustaleń artefaktualnych.



Ryc. 3.3

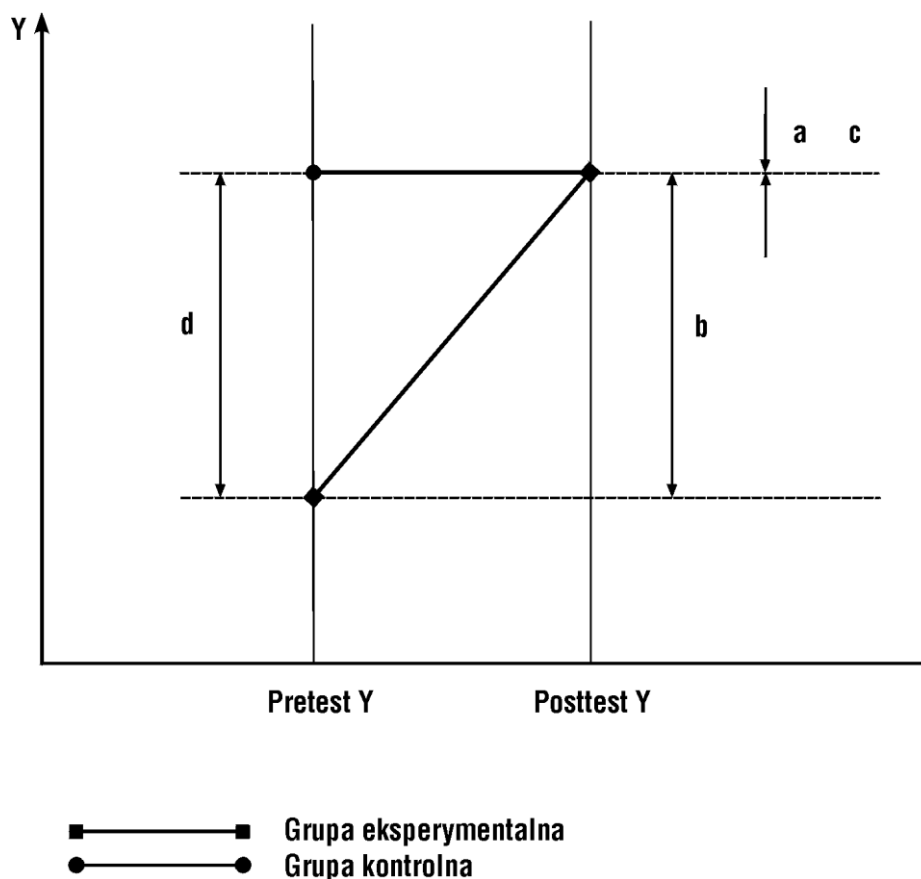
Po pierwsze, wskaźnik a nie informuje rzetelnie o względnej wielkości skuteczności programu społecznego, który wprowadzony był w grupie eksperymentalnej.



Ryc. 3.4

Po drugie, grupy nie były równoważne w momencie rozpoczęcia badania (o czym badacz, który nie przeprowadził pretestów w obu grupach, nie wiedział): $d \neq 0$. Zatem na otrzymaną przez badacza wartość wskaźnika a składają się:

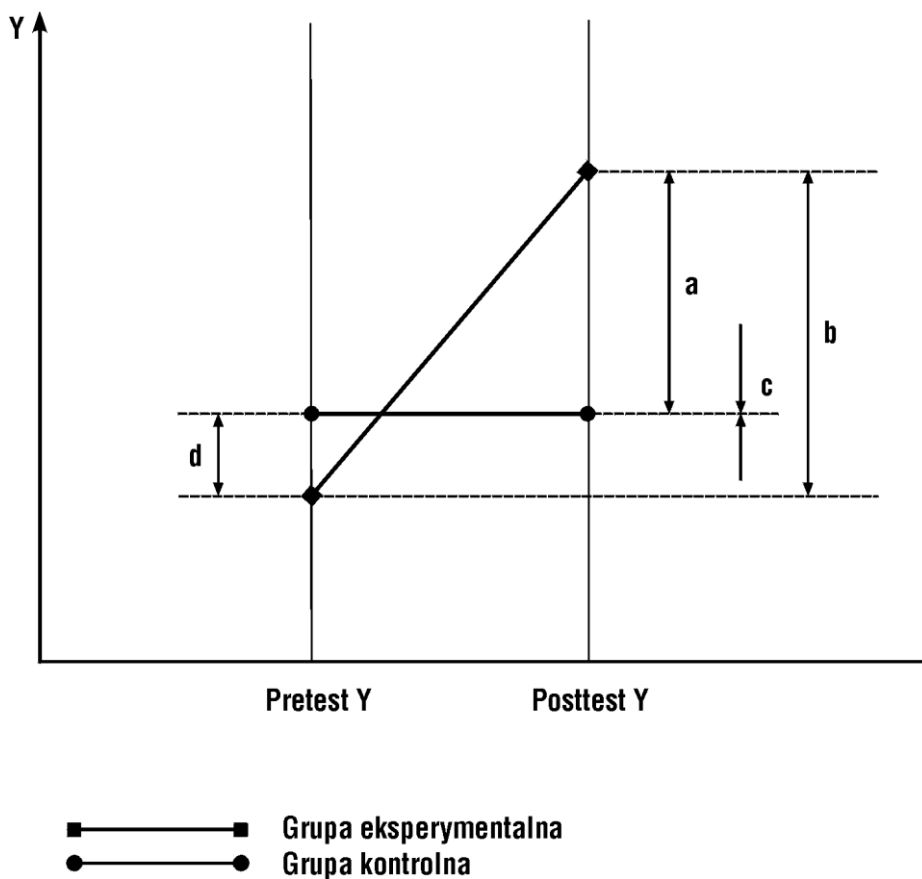
1. wartość *rzeczywistego* oddziaływania zmiennej X na grupę eksperymentalną (co wyraża wartość wskaźnika b : $b > 0$ lub, jak w przypadku Ryc. 3. 7: $b < 0$, albo też, jak w przypadku Ryc. 3.4 i 3.8: $b = 0$),
2. wartość będąca efektem korzystniejszych warunków „startowych” grupy eksperymentalnej (co z kolei wyraża wartość wskaźnika d : $d > 0$), albo gorszych warunków wyjściowych ($d < 0$ – por. Ryc. 3.5 i 3.6).



Ryc. 3.5

Badacz, który nie przeprowadziłby pretestów Y w porównywanych grupach nie byłby też w stanie dokonać, takiej, jak wyżej dokonana, analizy względnych wielkości wskaźników: b , c , d (on ich, po prostu, nie znałby). Jeszcze gorszą – z tego punktu widzenia – sytuację przedstawia Ryc. 11.8. Pokazuje ona bowiem sytuację w której badacz uznaje – kontrfaktycznie – hipotezę roboczą za potwierdzoną na podstawie dostępnych mu jedynie (wszak nie przeprowadził pretestu Y w porównywanych grupach) informacji o wartości wskaźnika a . Co z tego, że przewyższa ona zero (i jest istotna statystycznie), gdy, jednocześnie „czysta” wielkość oddziaływania eksperymentalnego równa jest ... zeru! Informują o tym – jeżeli są badaczowi dostępne!

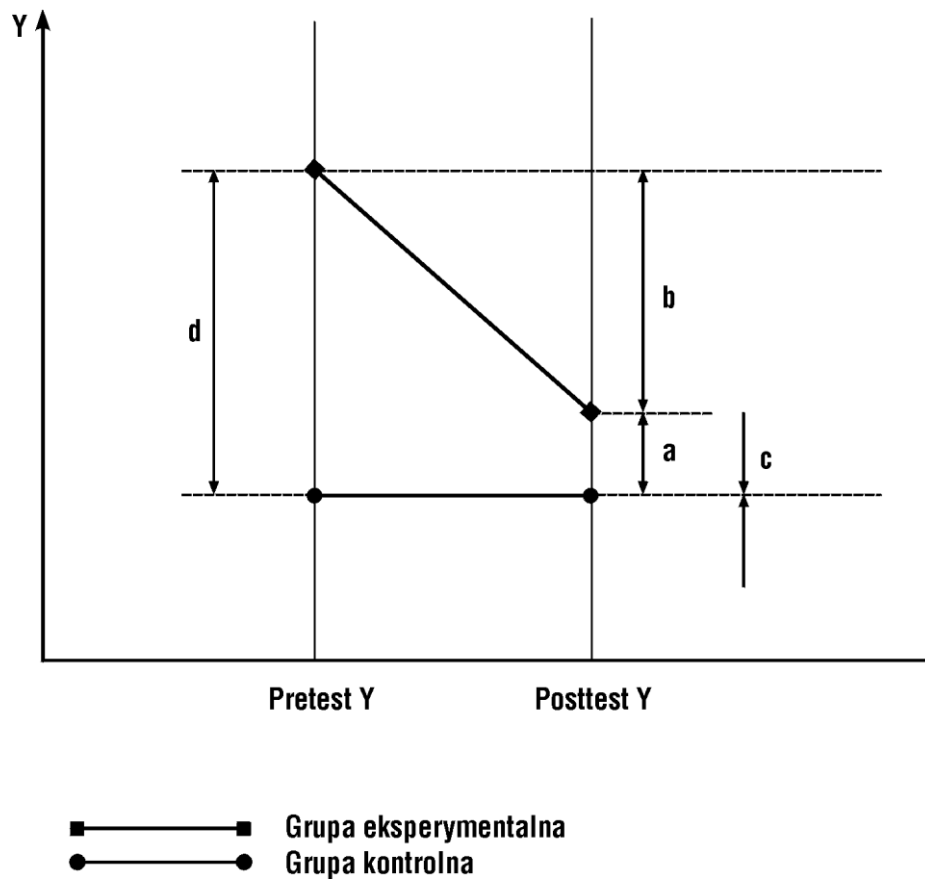
– wskaźniki b i d : $b = 0$, $d > 0$. Są one jednak dostępne tylko wówczas, gdy badacz przeprowadzi pretest zmiennej zależnej.



Ryc. 3.6

Taki nietrafny rezultat badania może być konsekwencją, albo niskiej świadomości metodologicznej badacza, a więc jego niekompetencji naukowych (efekt niedouczenia), albo, co gorsza, efektem celowej manipulacji warunkami przeprowadzonego eksperymentu (nazwijmy to po imieniu – oszustwa naukowego). Na skutek stronniczej selekcji osób do obu grup (nie respektowanie zasady randomizacji – nawiasem mówiąc wszystkie badania quasi-eksperymentalne nie przestrzegające zasady randomizacji i nie ujawnia-

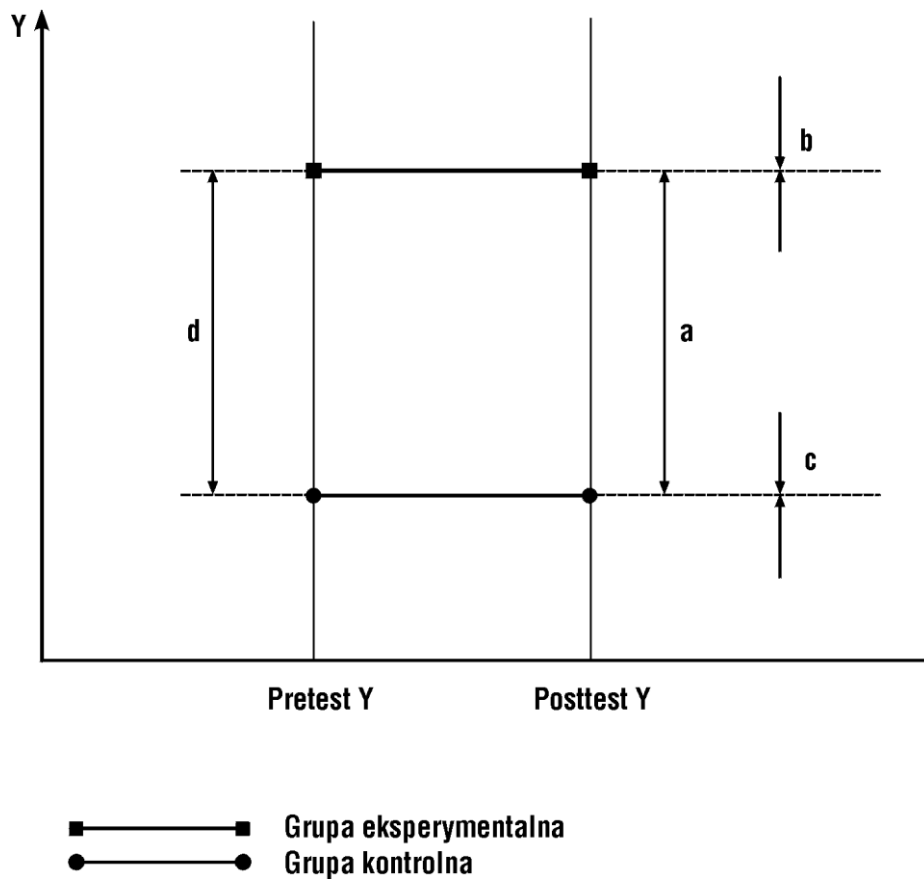
jące wyników pretestów Y uznać należy za nie zasługujące na zaufanie) pseudo-badacz mógł „dowodnie” wykazać, iż grupa eksperymentalna była „lepszą” od grupy kontrolnej.



Ryc. 3.7

Zaś faktycznie – jak to pokazuje Ryc. 11–8 – oddziaływanie zastosowane wobec grupy eksperymentalnej nie zmieniło jej. Czy jest to realnie możliwe? Uważam, że tak. Możemy przecież „wyleczyć” osoby *de facto* bardziej zdrowe, aniżeli osoby z grupy kontrolnej, jeżeli potraktujemy je jakimkolwiek preparatem (nawet czystą wodą). Rzecz jasna upraszczam nasz przykład, gdyż w rzeczywistości, pseudouczeni uciekają się do bardziej wyrafinowanych

sposobów potwierdzania tego, co chcą potwierdzić za pomocą przeprowadzanych – niekiedy dość wyrafinowanych proceduralnie i spektakularnych – pseudo-eksperymentów.



Ryc. 3.8

BIBLIOGRAFIA

Brzeziński J. (1989). Eksperymentalne badanie zmian – metodologiczne problemy stosowania pretestu zmiennej zależnej. *Przegląd Socjologiczny*, 37, 113–139.

Brzeziński J. (1997). *Metodologia badań psychologicznych* (wyd. 2). Warszawa: Wyd. Nauk. PWN.

Cook, Campbell (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin Co.

Grzelak J., Wolniewicz-Grzelak (1994). Przedmowa do wydania polskiego. [w:] J. D. Hawkins, B. Nederhood (1994). *Podręcznik ewaluacji programów profilaktycznych. Nadużywanie substancji uzależniających i inne problemy społeczne* (s. 7–8). Warszawa: PTP, Pracownia Wydawnicza.

Hawkins J. D., Nederhood B. (1994). *Podręcznik ewaluacji programów profilaktycznych. Nadużywanie substancji uzależniających i inne problemy społeczne*. Warszawa: PTP, Pracownia Wydawnicza.

Nederhof A. J. (1991). Metody korygowania zniekształcającego wpływu zmiennej aprobaty społecznej; przegląd badań. [w:] J. Brzeziński, J. Siuta (red.), *Społeczny kontekst badań psychologicznych i pedagogicznych. Wybór tekstów* (s. 33–59). Poznań: Wyd. Nauk. UAM.

Orne M. T. (1993). Eksperyment psychologiczny z punktu widzenia psychologii społecznej ze szczególnym uwzględnieniem wpływu zmiennych sugerujących hipotezę i ich implikacji. [w:] J. Brzeziński, J. Siuta (red.), *Społeczny kontekst badań psychologicznych i pedagogicznych. Wybór tekstów* (s. 15–32). Poznań: Wyd. Nauk. UAM.

Rosenberg M. J. (1991b). Warunki powstawania oraz konsekwencje lęku przed oceną. [w:] J. Brzeziński, J. Siuta (red.), *Społeczny kontekst badań psychologicznych i pedagogicznych. Wybór tekstów* (s. 61–135). Poznań: Wyd. Nauk. UAM.

Rosenthal R. (1991). Oczekiwania interpersonalne. Skutki przyjętej przez badacza hipotezy. [w:] J. Brzeziński, J. Siuta (red.), *Społeczny kontekst badań psychologicznych i pedagogicznych. Wybór tekstów* (s. 235–339). Poznań: Wyd. Nauk. UAM.

Webst S. G. (1997). Poza eksperyment laboratoryjny – plany eksperymentalne oraz quasi-eksperymentalne w otoczeniu naturalnym. [w:] I. Heszen-Niejodek, H. Sęk (red.), *Psychologia zdrowia* (s. 111–174). Warszawa: Wyd. Nauk. PWN

Jerzy Brzeziński

Institute of Psychology, Adam Mickiewicz University

ON THE IMPORTANCE OF THE DEPENDENT VARIABLE PRETEST IN EVALUATION OF SOCIAL PROGRAMMES

Summary

The author discusses methodological consequences of a resignation from performing the dependent variable pretest in the experimental evaluation of social programmes. Eight research situations have been analyzed from the point of interpretative encumbrances (underestimation or overestimation) of the result of social programme evaluation. An emphasis was laid on the indispensability of performing the dependent variable pretest for a correct evaluation of the tested social programme. The article makes reference to the author's earlier work (cf. Brzeziński 1989).