

Publikacja pod redakcją *Agnieszki Haber* i *Macieja Szałaja*

Współpraca redakcyjna dotycząca tekstów  
D. Batorskiego oraz P. Strawińskiego: *Rafał Trzciniński*

Redakcja merytoryczna tekstu A. Haber, M. Szałaja: *Jacek Pokorski*

Redakcja językowa: *Anna Krasucka*

Redakcja techniczna: *Michał Moczarski*

© Copyright by **Polska Agencja Rozwoju Przedsiębiorczości**

ISBN 978-83-60009-86-4

Wydawca: **Polska Agencja Rozwoju Przedsiębiorczości**

ul. Pańska 81/83

00-834 Warszawa

tel.: (22) 432 80 80, 652 80 80

fax: (22) 432 86 20, 432 84 04, 652 86 20, 652 84 04

**[www.parp.gov.pl](http://www.parp.gov.pl)**

Wydanie I

Przygotowanie techniczne publikacji do druku:

Studio Graficzne *MIMO* Michał Moczarski, tel. (22) 642 01 18

Druk i oprawa:

Agencja Reklamowo-Wydawnicza

Arkadiusz Grzegorzcyk

[www.grzeg.com.pl](http://www.grzeg.com.pl)

## Spis treści

<i>Agnieszka Haber, Maciej Szalaj</i> Wprowadzenie .....	9
---	---

### **Część I. Ewaluacja w organizacji i społeczeństwie**

<i>Kazimierz W. Frieske</i> Nauki społeczne w służbie spraw publicznych – polskie tradycje .....	13
<i>Dariusz Zalewski</i> Możliwości i ograniczenia wykorzystania wyników ewaluacji przez instytucje publiczne .....	31
<i>Stanisław Mazur</i> Zarządzanie wiedzą w polskiej administracji publicznej .....	51
<i>Leszek Korporowicz</i> Interakcyjne aspekty procesu ewaluacyjnego: pomiędzy analizą a animowaniem zmian społecznych .....	71
<i>Agnieszka Haber, Maciej Szalaj</i> Kapitał społeczny – warunek i efekt udanych ewaluacji .....	91

### **Część II. Wybrane zagadnienia warsztatu badań ewaluacyjnych**

<i>Aleksander Surdej</i> Metoda analizy kosztów i korzyści: imperatyw efektywności w politykach publicznych .....	109
<i>Cezary Trutkowski</i> O uwarunkowaniach stosowania metod jakościowych w badaniach społecznych i ewaluacyjnych .....	127
<i>Tomasz Kaczor</i> Modele ekonomiczne w ewaluacji, czyli o czym jej odbiorca wiedzieć powinien .....	145

*Stanisław Bienias*

Wykorzystanie wyników modelowania ekonomicznego  
w procesie ewaluacji polityki spójności ..... 155

*Dominik Batorski*

Metody analizy sieci i ich zastosowanie w ewaluacji ..... 167

*Paweł Strawiński*

Quasi-eksperymentalne metody ewaluacji ..... 193

**Szanowni Państwo,**

Wraz z rozwojem systemu wdrażania funduszy strukturalnych oraz wielokrotnym zwiększeniem dostępnych środków wzrasta rola ewaluacji jako elementu zarządzania interwencją publiczną. Znaczące poszerzenie grona instytucji odpowiedzialnych za prowadzenie ewaluacji funduszy strukturalnych tworzy potrzebę pilnych działań ukierunkowanych na rozwój potencjału ewaluacyjnego. Kształtowanie i rozwój rynku usług ewaluacyjnych, zarówno po stronie wykonawców, jak i po stronie zamawiających, rodzi również konieczność systematycznej dyskusji na temat możliwości i warunków niezbędnych dla efektywnego wykorzystania i implementacji narzędzia, jakim jest ewaluacja.

Ufam, iż jedną z możliwych odpowiedzi na powyższe potrzeby będzie prezentowana Państwu propozycja wydawnicza PARP *Środowisko i warsztat ewaluacji*. Jej charakter, podobnie jak w przypadku poprzedniego tomu, zatytułowanego *Ewaluacja ex-post. Teoria i praktyka badawcza*, wynika z chęci zaprezentowania zróżnicowanych punktów widzenia oraz uchwycenia możliwie szerokiego zakresu zagadnień.

Mam również nadzieję, iż przygotowana przez PARP publikacja dostarczy Państwu inspiracji przydatnych w Państwa praktyce zawodowej. Liczę, iż lektura zawartych w niej artykułów stanie się impulsem do dyskusji nad głównymi problemami polskich badań ewaluacyjnych oraz najważniejszymi wyzwaniem stojącymi przed środowiskami zaangażowanymi w ich realizację. Jeżeli stanie się tak w istocie, będzie to najbardziej miarodajnym wyznacznikiem sukcesu naszej propozycji wydawniczej oraz dowodem na trafność tego przedsięwzięcia.

**Danuta Jabłońska**  
Prezes Polskiej Agencji  
Rozwoju Przedsiębiorczości



## Wprowadzenie

Przedstawiamy Państwu kolejny zbiór tekstów poświęconych dziedzinie, która już na dobre zaistniała nie tylko w świadomości, ale również w praktyce wielu polskich instytucji publicznych.

Ewaluacja systematycznie zyskuje na znaczeniu, poszerza kręgi swojego oddziaływania, sięga po nowe podejścia i metody badań oraz zapewnia sobie wpływ na działania podejmowane w polityce publicznej. Liczba realizowanych w Polsce ewaluacji regularnie rośnie i przybiera coraz bardziej zinstytucjonalizowaną formę. Obserwując szybką ewolucję tej stosunkowo młodej dziedziny, nie należy jednak tracić z oczu jej dwóch głównych funkcji. Bezpośrednio wynikają one ze wzrostu publicznego nacisku na efektywność, celowość reagowania, trafność alokowania środków i odpowiedzialność działań rządu, czyli nurtu „nowego zarządzania publicznego”. Pierwszą funkcją jest dostarczenie wsparcia organizacjom oraz jednostkom w osiąganiu ich celów (z respektowaniem wartości i norm społecznych), budowanie strategii rozwoju i alokacji zasobów. Drugą jest zabezpieczenie interesu publicznego poprzez dostarczenie obywatelom informacji – w oparciu o sprawdzone dane i obiektywne sądy – na temat skuteczności, efektywności i oddziaływaniu polityk, programów rządowych oraz szeroko rozumianym sektorze usług publicznych (funkcji tzw. adwokata społecznego).

Dlatego też przygotowując drugi tom poświęcony ewaluacji, obok szeregu podejść badawczych oraz profesjonalnych instrumentów oceny stosowanych w ramach stale udoskonalanego warsztatu badawczego, podjęliśmy tym razem próbę spojrzenia na ewaluację nieco szerzej niż na sam proces jej realizacji.

Dziś nikt nie kwestionuje poglądu, że sukces ewaluacji (wykorzystanie jej wyników) zależy w równej mierze od poprawności metodologicznej, prowadzenia dialogu z otoczeniem instytucjonalnym, zarządzania wiedzą, kultury organizacyjnej i politycznej oraz kapitału społecznego. W myśl tak zarysowanej problematyki dokonaliśmy też podziału niniejszej książki. Pierwsza część poświęcona została zagadnieniom funkcjonowania ewaluacji w środowisku społecznym, a także w środowisku organizacji, druga tradycyjnie przybliży wybrane zagadnienia warsztatu badań ewaluacyjnych.

Mamy nadzieję, że zaproponowany wybór tekstów przybliży Czytelnikowi ewaluację w jej kolejnych wymiarach oraz ułatwi poruszanie się w jej

złożonym otoczeniu i implementacje najlepszych rozwiązań w wymiarze zarówno projektowym, jak i instytucjonalnym. I choć książka ta z pewnością nie obejmuje wszystkich zagadnień związanych z szeroko rozumianym środowiskiem i warsztatem ewaluacyjnym, mamy nadzieję, że zainicjuje dyskusję, która przyczyni się nie tylko do refleksji nad codzienną praktyką ewaluacyjną, ale również do rozwoju jej teorii oraz zarządzania politykami publicznymi.

*Agnieszka Haber i Maciej Szalaj*

Część I

**Ewaluacja w organizacji i społeczeństwie**





*Kazimierz W. Frieske*

## **Nauki społeczne w służbie spraw publicznych – polskie tradycje**

### **Społeczeństwo oparte na wiedzy**

Prezentowany dalej wywód jest zorganizowany wokół kilku tez podstawowych. W tej czy innej postaci były one formułowane w stosownej literaturze przedmiotu, choć służyły zwykle jako argumenty prowadzące do nieco innych konkluzji. Nad rozmaitymi formami praktycznego wykorzystania wiedzy naukowej, czy – szerzej – jakiegokolwiek systematycznej wiedzy, zastanawiało się wielu metodologów i filozofów nauki, dochodząc do wniosków tyleż oczywistych, co ciągle nieco amorficznych.

Po pierwsze, nie ma jakichkolwiek wątpliwości co do tego, że jedną z najbardziej typowych form praktyczności systematycznej wiedzy o świecie jest to, że kształtuje ona coraz silniej – wypierając zdrowy rozsądek i potoczne doświadczenie – rozmaite praktyki dnia codziennego. To jedna z najbardziej oczywistych interpretacji sloganu mówiącego o „społeczeństwie opartym na wiedzy”. Obecność nauk przyrodniczych w naszej codzienności zaznacza się w tysiącach działających wokół nas urzędzeń, bez których nie wyobrażamy sobie życia, choć zwykle nie pamiętamy, że ich powstanie stało się możliwe dzięki bardzo, niekiedy, zaawansowanej refleksji teoretycznej. Ale to samo można powiedzieć o wiedzy gromadzonej w naukach społecznych. Przecież to, jak urządzamy nasz świat społeczny, jest, przynajmniej w jakimś stopniu, konsekwencją tego, co w systematyczny sposób wiemy o rządzących nim prawidłowościach. Jako formę rządu wybieramy demokrację reprezentatywną, przekonani o tym, że lepiej stabilizuje ona życie zbiorowe aniżeli jakakolwiek tyrania – nawet, gdyby miała to być tyrania cnót merytokratycznych. Wszyscy, czy prawie wszyscy, żyjemy w jakichś porządkach organizacyjnych, ale kształt i struktura tych porządków jest funkcją tego, co o nich wiemy. Poszukiwania rozsądnego kompromisu między ideą „wolnego rynku” wraz z jego proefektywnościową zasadą konkurencji i ideą regulacji to wprawdzie przedmiot tyleż ideologicznych, co merytorycznych sporów, ale przecież wiemy, że właśnie rozsądny kompromis, nie zaś nacisk na rozwiązania skrajne, dobrze służy gospodarce. Popularna literatura pełna jest wielonakładowych porad-

ników opowiadających o efektywnych sposobach osiągnięcia indywidualnego sukcesu i – czegokolwiek by się o tych poradnikach, a także o zawodowych kreatorach wizerunku, nie powiedziało – starannie odsiewając bzdurę, przekonujemy się, że tkwi w tym wszystkim systematyczna, nieuchwytna w potocznym doświadczeniu, wiedza o funkcjonowaniu człowieka w jego społecznym otoczeniu.

Po drugie, wiedza gromadzona w naukach społecznych bywa wykorzystywana praktycznie jako instrumenty sprawowania władzy politycznej. W mniejszym stopniu idzie tu o kierunek myślenia wyznaczony przez dorobek Machiavellego, w większej mierze natomiast o wszystko, co wiąże się z administrowaniem sprawami publicznymi i procesami podejmowania decyzji. Jeśli nawet za tymi decyzjami, w stopniu większym, niż byśmy tego chcieli, kryją się rozmaite grupowe interesy, to jednak – o czym poucza przypadek słynnego „peronu we Włoszczowej” – istnieją rozsądne granice lekceważenia reguł alokacyjnej racjonalności, zaś ich przekraczanie łączy się zwykle z politycznymi skandalami. Mówiąc najkrócej, coraz częściej chcemy, żeby – *ex ante* – decyzje w sprawach publicznych były uzasadniane jakimiś analizami kosztów i korzyści, chcemy też, żeby – *ex post* – były one uzasadniane studiami ewaluacyjnymi wyłaniającymi, na przykład, tzw. dobre praktyki. Nawiasem mówiąc, warto też brać pod uwagę i to, że zarządzanie politykami publicznymi często sprowadza się nie tyle do racjonalizacji decyzji alokacyjnych, służących rozwiązaniu problemu, ile do przedstawienia informacji sugerujących jego pilność czy publiczną doniosłość. To akurat jest przypadek dość charakterystyczny, bo przecież bywa tak, iż uruchomienie procesu dla wypracowania lepszego czy gorszego rozwiązania okazuje się trudniejsze, aniżeli jego znalezienie – wtedy zwłaszcza, gdy każda z wchodzących w grę opcji decyzyjnych narusza jakieś politycznie doniosłe interesy. Jednak i wtedy nauki społeczne bywają użyteczne – choć lepiej jest mówić w takich sytuacjach o ich nadużywaniu. Wiemy wszak, że jedną z typowych strategii doświadczonych administratorów, usiłujących uniknąć podjęcia jakiegokolwiek decyzji, jest powołanie panelu ekspertów i niespieszne oczekiwanie na to, że uzgodnią oni swoje rekomendacje.

Po trzecie, systematyczna wiedza płynąca z nauk społecznych wkracza w ludzką codzienność, podsuwając takie jej wizje, które nie mogłyby powstać w oparciu o to, co dyktuje zdrowy rozsądek, skumulowane indywidualne doświadczenia czy „życiowe mądrości”. Nie widać powodów, dla których te ostatnie należałoby lekceważyć, mają one jednak tę wadę, że łatwiej znoszą wewnętrzne sprzeczności, tych zaś w nauce staramy się unikać. To – i tak brzmi jedna z prezentowanych dalej tez – ważna forma wykorzystania czy stosowania nauk społecznych. Jest ona tym istotniejsza, że kształtuje ludzkie postępowanie – w tym także działania tych, którzy o biegu spraw publicznych decydują – w sposób mało dostrzegalny i często taki, który wydaje się wielce kontrowersyjny. Na przykład, zwykliśmy tłumaczyć wiele zachowań

dewiacyjnych, także przestępczość, marną efektywnością rozmaitych mechanizmów kontroli społecznej i odpowiednio do tego tworzyć stosowne technologie prewencji czy resocjalizacji. Tymczasem, okazuje się, że współczesne *opportunity theories* każą traktować dewiantów jak ludzi racjonalnych, którzy w zastanej sytuacji dążą do maksymalizacji swoich indywidualnych korzyści i starannie rozważają relację między karami i nagrodami, które są konsekwencją zachowania uznawanego za społecznie naganne. Nie warto może wdawać się w złożoną, teoretyczną dyskusję tego punktu widzenia, warto natomiast wskazać, że – przenikając do codziennego, praktycznego myślenia – radykalnie zmienia on pytanie, które rozpoczyna dywagacje nad sposobami kontrolowania przestępczości. Mianowicie, zamiast pytać – jak robiły to całe pokolenia kryminologów i polityków społecznych – o przyczyny, dla których ludzie popełniają przestępstwa, pytamy o powody, dla których mieliby oni ich nie popełniać – skoro zastana struktura możliwości mówi, że ogólny bilans dokonywania rozmaitych zbrodni może okazać się pomyślny. Nie jest wykluczone, że ten punkt widzenia zakorzenia się powoli w codziennym myśleniu Polaków, zaś jego wyrazem jest – uchwytne w badaniach opinii publicznej – represyjność ich postaw wobec przestępców.

Na kolejnych stronach nieco szerzej zaprezentowana zostanie argumentacja, że ostatnia, ze wskazanych wyżej, forma wykorzystania wiedzy gromadzonej w naukach społecznych jest nie tylko ważna, lecz także – że była ona charakterystyczna dla polskiej socjologii już w jej początkach. Co więcej, początki te wiązały się z recepcją – co prawda wielce eklektyczną – dorobku Comte'a.

Socjologia też posłuży w niniejszym tekście jako ilustracja zjawisk wspólnych dla ogółu polskich nauk społecznych, których granice przez długie lata pozostawały dość płynne, jeśli brać pod uwagę zarówno ich instytucjonalne wyodrębnienie w postaci oddzielnych katedr, instytutów, czy też programów kształcenia nowych adeptów poszczególnych dyscyplin, jak i specyfikę stosowanego warsztatu.

Dalej, sięgając do mniej odległej, bo już dwudziestowiecznej tradycji polskiej socjologii, wskazuje się też na ten jej nurt, który ma wyraźnie krytyczny charakter i wiąże się z koncepcjami tkwiącymi u podstaw Instytutu Gospodarstwa Społecznego. Prace IGS-u demonstrowają przy tym i to, że idea ewaluacji – jeśli rozumieć ją jako, wspartą systematycznymi danymi empirycznymi, ocenę rozmaitych polityk publicznych – wcale nie jest nowa! Nowy jest tylko termin, zaś to, co się za nim kryje, stanowiło jedno z zasadniczych, programowych zadań instytucji, która obecna była w polskim życiu publicznym przez cały okres międzywojenny.

Wreszcie, ostatnia z prezentowanych tez mówi, że koncepcja badań ewaluacyjnych wyraża dwie formuły wykorzystania wiedzy gromadzonej w naukach społecznych. Pierwsza, bardziej popularna, sprowadza się do posługiwania się tą wiedzą w procesach zarządzania politykami publicznymi dla

ich racjonalizacji. Nie widać powodów, dla których należałoby doniosłość tych zabiegów kwestionować, choć odpowiedź na pytanie, w jakim stopniu są one wieńczone sukcesami, jest nader złożona i niepewna. Druga, rzadziej dostrzegana i mniej zbadana, choć, wedle prezentowanego tu poglądu, mniej doniosła, polega na dostrzeżeniu tego, że studia ewaluacyjne, gdy zostają zinstytucjonalizowane jako jedna ze społecznych praktyk, są ważnym elementem procesów demokratyzacji życia zbiorowego. Pozwalają nie tylko testować obietnice merytokracyjnych elit, lecz także usuwać przynajmniej część napięć, jakie powstają wtedy, gdy usiłuje się poddawać demokratycznej kontroli profesjonalny aparat administracyjny.

Odnosząc powyższe tezy do problematyki popularnych ostatnio badań ewaluacyjnych, warto pokazać, iż nie stanowią one bynajmniej nowego fenomenu, lecz są raczej powieleniem pewnej „kliszy”, typowej dla długiej tradycji stosowanych podejść do naukowej i przednaukowej wiedzy o społeczeństwie.

### **Fatamorgana „stosowania”**

Mysł o praktycznym wykorzystaniu wiedzy gromadzonej w naukach społecznych jest w tej samej mierze uniwersalna co niejasna. Podatnicy chcieliby zapewne wiedzieć, jakie to uchwytnie korzyści czerpią z finansowania instytucji naukowych i ich personelu. I trudno zadowolić ich ogólnikowym stwierdzeniem, wedle którego korzyści te polegają na lepszym czy głębszym rozumieniu społecznych procesów. Politycy, nieustannie narażeni na naciski rozmaitych akademickich grup nacisku, z niefrasobliwą hipokryzją wpisują się w polityczną poprawność, z której miałoby wynikać, że droga do sukcesu prowadzi przez zwiększanie nakładów na rozmaite, ekstrawaganckie rojenia o racjonalnie zaplanowanej zmianie społecznej. Studenci, buntujący się przeciwko wysiłkom swoich nauczycieli, bezlitośnie posługujących się symboliczną przemocą teorii społecznych, pytają dość naiwnie o to, co z tych teorii wynika dla tłumienia skrzeku rzeczywistości – słyszanego tym boleśniej, im się jest młodszym i bardziej wrażliwym. Najkrócej – „wykorzystywać”, dla mniej czy bardziej godziwych racji, chce wielu, niektórzy na stosownych deklaracjach budują swoją indywidualną czy grupową pomyślność. Wychylając się nieco z łoży szydery, w której królował niegdyś Antoni Słonimski – warto może zapytać, co właściwie robią ci, którzy mówią, że w swojej praktycznej działalności „wykorzystują wiedzę, jaką proponują akademickie nauki społeczne”? Wbrew pozorom, pytanie nie jest banalne. Wypada też stwierdzić, że nie jest oryginalne – bo było stawiane wielokrotnie. W anglosaskiej literaturze, poświęconej krytycznej refleksji nad mirażami postępu fundowanego na sukcesach nauki, pojawiło się ono, gdy stało się jasne, że transfer wiedzy między nauką i praktyką, zwłaszcza praktyką polityczną, jest procesem wielce skomplikowanym i dalece nieoczywistym.

Nie jest wykluczone, że późniejsze kłopoty z odpowiedzią na to pytanie przeczuwał przed blisko półwieczem Alvin Gouldner, gdy zwracał uwagę na różnice pomiędzy „techniczną” i „kliniczną” formułą „wykorzystania”. Gouldner pisał tak: „(...) dominujące współcześnie definicje roli badacza i roli decydenta, jakie przyjmuje większość socjologów są osadzone w klasycznym, utylitarystycznym kontekście (...). Decydent definiuje swoje kłopoty jako konsekwencje niedostatecznej wiedzy (...) zakłada on, że dysponując odpowiedzią wiedzą byłby w stanie uporać się ze swym problemem (...) mając na uwadze te założenia prosi o radę socjologa. (...) Socjolog (...) siebie samego określi najpewniej jako źródło wiedzy składającej się z faktów i cyfr (...)”<sup>1</sup>. Współcześnie już wiemy, że taka – techniczna właśnie – formuła „wykorzystania”, która miałaby polegać na poszerzaniu wiedzy o ten czy inny problem społeczny, aby przekazać ją zainteresowanemu politykowi lub administratorowi, na uwzględnianiu w procedurach podejmowania decyzji rozmaitych obiektywnych czy naukowych danych, to jedynie radosna iluzja. Nie warto może wdawać się w złożone uzasadnienia tej tezy, dość powiedzieć, że nawet najprostszy pomiar jakiegokolwiek zmiennej nieuchronnie odwołuje się do teoretycznych założeń, przyjmowanych z pewną dozą samoświadomości. Warto natomiast wskazać, że – nawet jeśli wyniki takiego pomiaru, więc właśnie zestaw owych *faktów i cyfr*, są łatwo dostępne, to z tego nie wynika wcale, że zostaną wzięte pod uwagę i wpłyną na treść decyzji.

Rozczarowania socjologów, którzy wielokrotnie przekonywali się o tym, że nie tylko mniej czy bardziej amorficzne wizje społecznej reformy, lecz także systematycznie gromadzone dane są ignorowane, przypominają doświadczenia Platona, który – w roli filozofa i politycznego intryganta – został dwukrotnie relegowany z Syrakuz. Szeroko propagowane frazesy służą bardziej legitymizacji politycznych decyzji niż ich zakorzenieniu w systematycznej wiedzy gromadzonej przez socjologów. Takie, na przykład, hasło gospodarki opartej na wiedzy, bywa wprawdzie instytucjonalizowane w postaci rozmaitych formalnych procedur, ale realia pokazują, że – starym zwyczajem – stanowi ono fasadę, za którą kryje się arbitralność zdrowego rozsądku czy partykularyzm interesów grupowych. Nieustannie przekonujemy się o tym, że wiedza gorsza – choć z tych czy innych względów wygodna – wypiera wiedzę lepszą, bo systematyczną, choć niekoniecznie zgodną ze zinstytucjonalizowanymi stereotypami. Mówiąc najkrócej, nie widać jakichkolwiek powodów, dla których reguły masowej konsumpcji nie miałyby się stosować do konsumpcji wiedzy – wszak o tym, że gorsza moneta wypiera lepszą, wiedział już Kopernik. Oczywiście nie dzieje się tak zawsze, że empiryczna ewidencja gromadzona w naukach społecznych jest kompletnie ignorowana w złożonych procesach podejmowania decyzji o ważnych sprawach publicznych. Można

<sup>1</sup> Gouldner A.W., „Explorations in Applied Social Science”, w: A.W. Gouldner, S.M. Miller (red.), *Applied Sociology, Opportunities and Problems*, The Free Press, N.Y. 1964, s. 17–18.

jednak utrzymywać z powodzeniem, że to, w jakiej mierze decyzje te są przez nią uzasadnione, jest wyznaczone przez przypadek czy przez osobliwości polityki, nie zaś przez jakość gromadzonych danych. Nie jest jednak wykluczony, że w demokratycznych społeczeństwach tak właśnie być powinno – nawet, gdyby miało się to kłócić z pryncypiami racjonalnego decydowania.

Rzecz jednak nie w tym, aby narzekać, że wiedza gromadzona w naukach społecznych bywa marnie wykorzystywana. Trzeba uświadomić sobie, że entuzjazm wobec idei „wykorzystania” – a zatem również wobec rozmaitych koncepcji socjologii stosowanej – jest funkcją tego, jak się to „wykorzystanie” rozumie. Możliwości mamy w tym względzie bardzo wiele. Spora część rozczarowań związanych z różnymi koncepcjami socjologii stosowanej wynika z wąskiej, bardzo instrumentalnej formuły „wykorzystania”, rozumianego jako wdrożenie rekomendacji wyłaniających się z prezentowanych danych empirycznych. Niechętnie stajemy w obliczu problemów strukturalizowanych wedle schematu *kochaj albo rzuć*. Mówiąc nieco inaczej: demokratyczne procesy polityczne rzadko stawiają ich uczestników wobec zero-jedynkowych wyborów, najczęściej politycznie wygodne rozwiązania sprowadzają się do propozycji, wedle których *trochę pokocham, a potem się zastanowię*, czyli do poszukiwania rozwiązań kompromisowych. Instrumentalna formuła „wykorzystania” wiedzy naukowej przez swoje konsekwencje kończy się najczęściej oddawaniem władzy grupom ekspertów – i wydaje się, że jest to zjawisko tym groźniejsze, im częściej wchodzi w ich skład anonimowi analitycy zaludniający zastępy rządowych biurokracji. Nawet gdybyśmy przyjęli, że instrumentalna formuła „wykorzystania” stanowi jedną z jego skrajnych wprawdzie i rzadko spotykanych, ale efektywnych form rozwiązywania rozmaitych problemów społecznych, czyli, że to, co dyktuje naukowy rozum, jest traktowane poważnie, nie ulega zniekształceniom w toku komunikowania jego konstatacji i że jego zalecenia są realizowane bez względu na polityczny kontekst, że nie brakuje ani pieniędzy, ani instytucjonalnej infrastruktury, ani odpowiednio wykwalifikowanego personelu, to i tak zniechęca ona tym, że kryje się za nią władza merytokratycznych elit zarządzających życie zbiorowe bez demokratycznego mandatu.

### **Nieoczywiste formy „wykorzystania”**

Fortunnie idea „wykorzystania” wiedzy gromadzonej w naukach społecznych może być też rozumiana szerzej – na przykład, jako wprowadzanie w struktury myślenia potocznego nowych pojęć czy sposobów obrazowania społecznej rzeczywistości. To zapewne druga, przez swój amorfizm również skrajna, interpretacja „wykorzystania”. Rzecz nie w tym – żeby posłużyć się określeniem Mary Douglas – iż scjentyistyczna kosmologia jest bliższa realiom, a zatem pozwala działać sprawniej, bardziej efektywnie. To kwestie długich

sporów kończących się zwykle takim czy innym wyznaniem wiary – w weberski proces racjonalizacji, wiary w wyższość naukowego poznania nad poznaniem porządkowanym przez zdrowy rozsądek i codzienne doświadczenie, wreszcie – wiary w pożytek płynący z racjonalnego projektowania zmiany społecznej. Chodzi o to, że próby nasycania zdrowego rozsądku wiedzą naukową – nawet, jeśli niekoniecznie prowadzą do lepszego odwzorowywania społecznego świata, to jednak mają tę zaletę, że umożliwiają testowanie obiegowych stereotypów i kwestionowanie tego, co uchodzi za oczywistość i jako takie bywa instytucjonalizowane.

Trzy przykłady mogą zilustrować tę myśl. Pierwszy z nich dotyczy czasów odległych, sięgających głębokiego PRL-u, to jest, końca lat 50. Funkcjonowało wtedy w prasowych enuncjacjach pojęcie najpierw „bikiniarzy”, potem zaś „chuliganów” przejmujących „kapitalistyczny styl życia” i stanowiących widomy dowód dywersji ideologicznej równie groźnej, co zachęcanie stonki ziemniaczanej do pożerania socjalistycznych kartofli. Czesław Czapów odniósł w tamtym czasie spektakularny sukces, publikując – wraz ze Stanisławem Manturzewskim – książkę pt. *Niebezpieczne ulice* wtłaczającą do myślenia o chuliganach obce klasowo pojęcie frustracji – co łączyło się ze zwróceniem uwagi na to, że nie wszystko w socjalistycznym ładzie społecznym układa się tak harmonijnie, jak chcieliby tego jego twórcy. Miało to swoje oczywiste konsekwencje praktyczne, bo wskazywało, że rozwiązywanie strukturalnych problemów młodzieży nie może polegać wyłącznie na ideologicznych szkoleniach, udziale w czynach społecznych itp. Początkowo na niewielką skalę, później na skalę zgoła masową, zaczęły powstawać rozmaite formy, dalej wprawdzie starannie kontrolowanej, ale też w jakimś stopniu autonomicznej aktywności młodych ludzi usiłujących nieco ubarwić szarość miejsca i czasu, przerebobić tzw. hotel robotniczy w miejsce nadające się do życia.

Bardzo podobne są losy popularnej współcześnie koncepcji „społecznego wykluczenia”. Termin wszedł do języka publicznych debat i zaczął organizować myślenie i praktykę pracowników socjalnych w drugiej połowie lat 90. Między innymi dlatego, że posługiwały się nim oficjalne dokumenty Komisji Europejskiej wyznaczające kierunki transferów socjalnych i zasady alokacji zasobów na rozmaite programy polityki społecznej. Do wiedzy o zjawiskach społecznej marginalności, gromadzonej przynajmniej od stu lat w ramach akademickiej socjologii, nowy termin nie wnosił nic nowego, ale – interpretowany na poziomie zdrowego rozsądku jako jeden z kluczy do rozumienia codziennych realiów – przekonywał o tym, że zbiorowy sukces i procesy społecznej modernizacji mają także swoje ofiary. Wskazywał, że są wokół nas ludzie, którzy – wbrew stereotypowi – są tacy, jacy są, nie tylko na skutek rozmaitych indywidualnych deficytów, lecz także w rezultacie rozmaitych strukturalnych determinizmów – i z tego względu warto publicznymi zasobami wspierać wysiłki zmierzające do ich społecznej reintegracji.



Przykład trzeci dotyczy pojęcia, które zdobyło sobie dość niechlubną sławę, mianowicie popularnego w publicystyce, czasem uznawanego za polityczny fantom, czasem za część otaczającej nas rzeczywistości, pojęcia „układu”. Otóż, wbrew tym, którzy istnienie i funkcjonowanie „układu” odkrywali, nie mówili oni o społecznym świecie niczego nowego: socjologowie wiedzą lub przynajmniej powinni byli wiedzieć, że jest to zjawisko ponadczasowe i ponadustrojowe. Bywało, że służyło ono do opisu realiów PRL-u lat 70. i wtedy Adam Podgórecki posługiwał się terminem „brudna wspólnota”. Bywało też, że wyrażało ono rozczarowania związane z efektywnościowymi iluzjami wolnokonkurencyjnego kapitalizmu: Mancur Olson opisywał w *Logic of Collectiva Action* powstawanie „koalicji dystrybucyjnych”, tj. właśnie „układów” składających się z ludzi monopolizujących dostęp do jakiś publicznych zasobów i rozdzielających je wśród jakichś „swoich”. Nie ma zatem w tezie o „układach” oplątujących życie zbiorowe niczego szczególnie oryginalnego, pozostaje jedynie polityczna eksploatacja zjawiska nazwanego, znanego i czasem badanego.

## Komplikowanie potoczności

Każdy ze wskazanych przykładów demonstruje to, że przynajmniej część rezultatów akademickiego namysłu wpisuje się w „światy przeżywane” praktyków, wzbogaca je, wyznacza i – co może ważniejsze – legitymizuje decyzje alokacyjne, jak i wskazuje na kierunki działania w rozległych obszarach polityki społecznej. W opisywanych sytuacjach mamy do czynienia z ważną formą „wykorzystania” wiedzy gromadzonej w ramach nauk społecznych – tym ważniejszą, że wiedza ta przekłada się nie tyle na decyzje podejmowane w politycznych czy administracyjnych establishmentach, ile na praktyki życia codziennego. Posuwając się tym tropem, możemy też powiedzieć, że amorficzną wprawdzie i pomiarowo skomplikowaną formą wykorzystania wiedzy, jaką mają pozaakademickiemu światu do zaproponowania nauki społeczne, jest posługiwanie się tą wiedzą dla uzasadniania poglądów wchodzących w szerszy społeczny obieg, czyli w strukturę publicznego dyskursu. Gdyby przyjąć taki punkt widzenia, to okazałoby się zapewne, że socjologia – także zresztą w jej przedakademickiej postaci – była „wykorzystywana” zawsze i że to „zawsze” dotyczy także początków refleksji socjologicznej w Polsce.

## Socjologia i zaangażowana publicystyka

Trudno zresztą, by było inaczej – bo początki polskiej socjologii to przede wszystkim oswajanie dorobku Comte'a<sup>2</sup>. „Prawdziwa filozofia pozytywna – pisał Comte – polega więc przede wszystkim na tym, aby wiedzieć dla przewidywania (...)”; „(...) w filozofii koniecznym przeznaczeniem wszystkich naszych zdrowych dociekań jest ciągłe ulepszanie naszych indywidualnych oraz społecznych warunków, w których żyjemy, nie zaś czcze zaspokojenie bezpłodnej ciekawości (...)”<sup>3</sup>. Ale, pamiętając o tym, warto też dostrzec to, że polska filozofia pozytywna powstawała w szczególnym kontekście doświadczeń i obserwacji, które następczo dziesiętnastowieczne społeczeństwo. Gdy Józef Supiński w 1857 r. prezentował opinii publicznej i przekazywał galicyjskiej administracji projekt stworzenia Narodowego Banku Oszczędności, to racje, które ów projekt uzasadniały, wywodził w tej samej mierze z ogólnych zasad gospodarowania i funkcjonowania bankowości, co z charakterystyki gospodarczej i diagnozy lokalnych warunków: „(...) pieniądź przechowywany bezrozumnie, szczególnie pieniądź złoty i srebrny, ukryty za strzechą, w grzędzie ogrodowej lub w lesie pod dębem, to życie naszych wieśniaków, poczciwych i wielce roztropnych (...) lecz żyjących życiem początkującym, lub życiem zwichniętym i odurzonym. Te pieniądze (...) przedstawiają część zasobu społecznego. (...) Te wartości skryte, leżące nieużytecznie (...) użyte z ręcznie i w sposób odpowiadający naszym stosunkom mogą stać się materiałem rozległej instytucji kredytowej (...)”<sup>4</sup>.

Jan Welichowski, wydawca *Pism* Supińskiego, uzupełnił druk wspomnianego tu pomysłu następującym komentarzem: „(...) projekt tedy (...) przedłożony był (...) Dyrekcji galicyjskiej kasy oszczędności, która go na jednym ze swoich posiedzeń jednomyślnie przyjęła i chciała wprowadzić w wykonanie; nie dozwoliło jednak tego ministeryum państwa (...)”<sup>5</sup>. Ci zatem, którzy narzekają na to, że władze publiczne miewają dość sceptyczny stosunek do reform instytucjonalnych proponowanych przez socjologów znajdują się w dobrym towarzystwie. Ale to, w swojej mądrości, Supiński wiedział sam, pisząc zgryźliwie, że: „(...) bank, który tu projektuję nie powstanie nigdy; bo ażeby pomyśleć o nim szczerze, trzeba pragnąć dobra powszechnego więcej, aniżeli własnych korzyści (...)”<sup>6</sup>. Instytucjonalna porażka Supińskiego była jednak zarazem zwycięstwem, bo przecież jego propozycja weszła w obieg społecznego myślenia i była uznawana za jeden z możliwych sposobów gospodarczej emancypacji. Nie bez powodów, komentując jego dorobek, J. Szacki napisał,

<sup>2</sup> Por. Szacki J. (red.), *Sto lat socjologii polskiej*, PWN, Warszawa 1995, s. 12.

<sup>3</sup> Comte A., *Rozprawa o duchu filozofii pozytywnej*, PWN, Warszawa 1973, s. 46.

<sup>4</sup> Supiński J., *Pisma. Wydanie poprawione i pomnożone*, t. IV, Warszawa 1872, s. 247.

<sup>5</sup> *Ibidem*, s. 235.

<sup>6</sup> *Ibidem*, s. 262.

że „(...) wpłynął on bardziej na styl myślenia i postawy obywatelskie niż na szczegółowe poglądy socjologiczne (...)”<sup>7</sup>.

Polska dziewiętnastowieczna socjologia była w dużej mierze – może za wyjątkiem typowego *homo academicus*, Gumplowicza – dziełem ludzi, którzy zajmowali się nią jednocześnie z aktywnościami o charakterze obywatelskim, wyrażającymi się m.in. w publicystyce. Wszak od publicystyki i działalności o charakterze politycznym zaczynał także, już w początkach wieku dwudziestego, Stefan Czarnowski. Nina Assorodobraj pisała wprawdzie, że – jako publicysta – „(...) starał się Czarnowski zawsze wiązać swe wypowiedzi z analizą naukową (...)”<sup>8</sup>, ale byłoby zapewne przesadą utrzymywanie, że on i wcześniejsi polscy socjologowie szukali w wiedzy socjologicznej systematycznego uzasadnienia dla prezentowanych poglądów w sprawach publicznych. Byłoby to też trudne do zademonstrowania poprzez kwerendę oryginalnych tekstów. Nie jest jednak wykluczone, że to, co mieli oni do powiedzenia w sprawach publicznych, w ich wizjach życia zbiorowego było jakoś zakorzenione.

Uwaga ta charakteryzuje obecność w polskim życiu umysłowym Bolesława Limanowskiego. J. Szacki, komentując życie i dzieło Limanowskiego, pisał, że: „(...) przejawiał temperament działacza politycznego, który nauką zajmuje się nie dla niej samej, lecz ze względu na jej spodziewane pożytki praktyczne (...)”<sup>9</sup>. Istotnie, okazuje się, że Limanowski, socjolog eklektyczny, po trosze comtysta, po trosze spencerowski organicysta, spory fragment swojej *Socjologii*<sup>10</sup> poświęcił zmianom społecznym pojawiającym się w rezultacie przyrostów wiedzy naukowej: „nauki stosowane i wpływ onych bezpośredni a olbrzymi na zachowanie, przedłużenie, uprzyjemnienie i uszlachetnienie życia ludzkiego – to stanowczy dowód, że zasób umysłowy w rozwoju ludzkości ma istotnie pierwszorzędne znaczenie. (...) Jeżeli weźmiemy historię którejkolwiek z nauk stosowanych (...) to przekonamy się, że postępy tych nauk nie były przypadkowe, ale zawsze zależały od ogólnego postępu wiedzy i że wynalazki nabierały prawdziwej wartości wówczas tylko, kiedy umiano je oceniać z naukowego stanowiska (...)”<sup>11</sup>. Po stu latach wiemy, że Limanowski się mylił, że tzw. nauki stosowane generują własną wiedzę i własne teorie naukowe, które swoją strukturą i wewnętrznymi kryteriami metodologicznej poprawności rozmijają się z ogólnym modelem nauki.

Idzie tu przede wszystkim o pokazanie tego, w jaki sposób socjologiczne koncepcje Limanowskiego wpływały na sposób definiowania tego, co miał on za istotny problem praktyczny, tj. kwestię narodowej niepodległości Polaków.

<sup>7</sup> Szacki J., op.cit., s. 13.

<sup>8</sup> Assorodobraj N., „Życie i dzieło Stefana Czarnowskiego”, w: S. Czarnowski, *Dzieła*, PWN, Warszawa 1956, t. V, s. 114.

<sup>9</sup> Ibidem s. 19.

<sup>10</sup> Por. Limanowski B., *Socjologia*, Gebethner i Wolf, Warszawa 1919.

<sup>11</sup> Limanowski B., *Socjologia*, cyt. za: J. Sztumski (red.), *Socjalizm jako konieczny objaw dziejowego rozwoju*, KiW, 1989, s. 126.

Można to robić, odwołując się do wielu wątków tych koncepcji, tu wystarczy zapewne wskazać na jeden z przykładów. Otóż, znajduje się w *Socjologii* Limanowskiego wywód, poparty zresztą szeregiem danych empirycznych, z którego wynika, że pojawianie się gospodarczych monopolów jest naturalną, żeby nie rzec, ewolucyjną – konsekwencją komplikacji procesów gospodarowania. Ta teza skłania jednak naszego autora nie tyle do pytania o jej modelowe konsekwencje, lecz do pytania, co z niej wynika dla rozumienia jego współczesności: „(...) jeśli więc ma istnieć monopol, to powstaje słuszne pytanie: czy lepiej, ażeby istniał on w rękach prywatnych, czy też należał do państwa? Odpowiedź nie byłaby trudna, gdyby państwo miało na celu interesy całego narodu. Tam jednak, gdzie rząd, trzymający wszystkie sprężyny organizacji państwowej, jest nie przedstawicielem narodu, a jego wrogiem, jak np. w Polsce, a zwłaszcza w zaborach rosyjskim i pruskim; tam, gdzie państwo służy interesom jednej klasy panującej – tam monopol, powiększając siłę państwową, jest szkodliwy, a nie pożyteczny narodowi (...)”<sup>12</sup>.

Wreszcie, warto też odnotować i to, jak bardzo bliskie współczesności okazują się te poglądy Limanowskiego, które dotyczą w szczególności wiedzy gromadzonej w naukach społecznych: „(...) wpływ stosowanych nauk społecznych właściwie jest świadomym oddziaływaniem społeczeństwa na własny swój ustrój, na własne swe losy. (...) Na uniwersytetach wykładają się nauki społeczne, głównie na wydziale prawnym, ale na nim wykłada się jeszcze dotąd cała scholastyka i formalistyka praw, z lekceważeniem istotnych nauk teoretycznych, a z zupełnym pominięciem najważniejszej, socjologii. Tymczasem prawnictwo bez podstawy socjologicznej (...) jest zupełnie w podobnym stanie, w jakim byłaby nauka lekarska bez podstawy biologicznej (...)”<sup>13</sup>. Tych, którzy nie śledzą toku współczesnych reform dydaktycznych na polskich wydziałach prawa, wystarczy poinformować o tym, że w końcu pierwszej dekady XXI w. socjologia została relegowana do przedmiotów fakultatywnych, nauczanych wtedy, gdy przyszli prawnicy kończą już swoją edukację. Tyle może w kwestii losów przenikania idei do myślenia organizującego społeczną praktykę: czasem zakorzeniają się w niej trwale, czasem oportunistyczny zwany niekiedy wichrem historii wymiata je na dłużej lub krócej.

## Socjolog w służbie swojej władzy

Niezależnie od tego, w jakiej mierze i w jaki sposób socjologia wkraçała w myślenie potoczne, warto mieć na uwadze, że zakorzeniała się ona w praktyce poprzez uwikłanie samych socjologów. Pierwsi polscy socjologowie – nawet jeśli nie w ramach akademickiej socjologii, budowali sobie

<sup>12</sup> Ibidem, s. 158.

<sup>13</sup> Ibidem, s. 171–172.

swoją zawodową tożsamość – wchodzili, bywało, w skład administracji i wykonywali w niej swoje zadania, usiłując czynić to w zgodzie ze swoimi profesjonalnymi kwalifikacjami. Mówiąc najkrócej – i wykorzystując dysfunkcję R. Mertona, który w swoim czasie podzielił ekspertów na insiderów funkcjonujących w strukturach władzy i outsiderów, usiłujących forsować swoje opinie z zewnątrz, na przykład, ze świata akademii<sup>14</sup> – takim socjologiem-insiderem był przez czas jakiś Stefan Czarnowski, pracując jako oficer-specjalista, na rzecz odtwarzającej się, polskiej armii. N. Assorodobraj pisała o Czarnowskim tak: „(...) podczas służby wojskowej (...) na ogół nie zaprzestawał pracy naukowej (...)”. Do ostatniej swej służby w Sztabie Generalnym, (pamiętajmy, że Czarnowski odszedł z wojska dopiero w 1923 r. – KWF), wykonywał szereg zadań naukowych, przede wszystkim w dwóch dziedzinach: demografii i geografii wojskowej. Prace demograficzne służyły miały za podstawę do planowania akcji mobilizacyjnych. Prace z zakresu geografii wojskowej były pracami przygotowawczymi do ustalenia strategicznych podstaw przyszłych granic Polski<sup>15</sup>.

Tę informację warto zapamiętać dla dwóch przynajmniej powodów. Po pierwsze – teza o niemalże koniecznym napięciu między światem władzy i światem akademii – nawet jeśli na ogół trafna – w szczególnych sytuacjach okazuje się jawnie fałszywa. Można łatwo wyobrazić sobie, że ta czy inna władza jest przez badacza uznawana, akceptowana i traktowana jako „swoja” – więc taka, której chce się służyć nawet wtedy, gdy bez trudu można zauważyć jej polityczne ograniczenia. To, czego dowiadujemy się z biografii Czarnowskiego, to przecież wzór z powodzeniem powielany po latach przez akademickich socjologów, którzy w początkach lat 90. – rezygnując ze swoich karier uniwersyteckich – zabierali się do budowy nowej Polski i służyli politycznym establishmentom, ten cel mając na uwadze. Powód drugi sprowadza się do odnotowania tego, że socjolog podejmujący rolę insidera tych czy innych struktur politycznych, administracyjnych, usiłujący funkcjonować w nich nie tylko sprawnie, ale i nie wyzbywając się swych zawodowych umiejętności, musi zaakceptować i to, że zaczyna funkcjonować jako członek zespołu, który nie może forsować swoich profesjonalnych racji. Musi godzić się z tym, że racje te są sensowne tylko w tej mierze, w jakiej sprzyjają realizacji szerszych, grupowo zaakceptowanych celów, że zatem staje się jednym z wielu uczestników złożonych procesów podejmowania decyzji politycznych czy administracyjnych. Wprawdzie akademickie ego może z tych powodów bardzo cierpieć, ale przecież w ideę wykorzystania wiedzy gromadzonej w naukach społecznych są wbudowane wartości skoncentrowane nie na pryncypialnym dążeniu do racjonalizacji społecznego świata, lecz na jego naprawianiu. Gdyby zatem

<sup>14</sup> Por. Merton R., „Rola intelektualisty w instytucjach państwowych”, w: R. Merton, *Teoria socjologiczna i struktura społeczna*, PWN, Warszawa 1982, s. 267 i n.

<sup>15</sup> Assorodobraj N., op. cit.

okazało się, że zbiorowy pożytek płynie z decyzji, które socjologiczną wiedzę ignorują, to – bez względu na to, jak bardzo byłyby to bolesne – nie widać płynących z tego nieszczęść.

## Socjologia krytyczna

Jedną z klasycznych form wpływania na społeczną rzeczywistość, łatwych do odszukania w tradycji polskich nauk społecznych, jest wszystko to, co wiązało się z losami Instytutu Gospodarstwa Społecznego. Są one stosunkowo dobrze udokumentowane<sup>16</sup>, a co więcej, tradycja IGS-u przetrwała czasy PRL-u. Jako placówka naukowo-badawcza i dydaktyczna istnieje do dzisiaj w strukturze Szkoły Głównej Handlowej. Pamiętając o związanej z polską lewicą, zwłaszcza z Ludwikiem Krzywickim, tradycji IGS-u, stosunkowo rzadko przywołuje się okoliczności poprzedzające jego powstanie. Otóż, warto powiedzieć, że w początkach XX w. – w ramach warszawskiego Towarzystwa Popierania Pracy Społecznej – od 1910 r. funkcjonowało Biuro Pracy Społecznej pomyślane jako – powiedzielibyśmy dzisiaj – *think tank*, stanowiące intelektualną grupę wsparcia dla reprezentantów Królestwa Polskiego w rosyjskiej Dumie Państwowej. Jednym z aktywistów tego Biura był Władysław Grabski, który na krótko przed wybuchem wojny, w ramach działalności służącej upowszechnieniu systematycznej wiedzy społecznej, wydał dwa roczniki statystyczne Królestwa Polskiego. W swoich początkach Biuro Pracy Społecznej było kierowane przez Zygmunta Chrzanowskiego, prawnika o wyraźnie konserwatywnych afiliacjach, ale z biegiem czasu wśród współpracowników Biura zaczęli pojawiać się ludzie albo bliscy socjalistycznym orientacjom ideologicznym, albo ludzie związani z ruchem ludowym – T. Hołówka, M. Niedziałkowski czy W. Wakar.

W 1917 r. ten ostatni został jego kierownikiem, zaś prace Biura potoczyły się nader dynamicznie – przygotowany został projekt ordynacji wyborczej do Sejmu Ustawodawczego, wykorzystany w 1919 r., a także opracowanie *Polski Program Terytorialny* i trzy tomy raportu *Rozwój terytorialny narodowości polskiej*, zawierającego uzasadnienie dla polskiego stanowiska na konferencji wersalskiej. Ideowe i personalne konflikty doprowadziły do zawieszenia działalności Biura. W. Wakar został odsunięty od funkcji kierowniczych, zaś ideologiczna fronda złożona z jego współpracowników przeniosła się do Towarzystwa Ekonomistów i Statystyków Polskich, zakładając Instytut Gospodarstwa Społecznego, czyli tę placówkę, której dzieje znamy znacznie lepiej. Jej pierwszym szefem dalej był W. Wakar, ale w 1921 r. zastąpił go Ludwik Krzywicki.

<sup>16</sup> Por. Szturm de Sztrem T., *Instytut Gospodarstwa Społecznego 1920–1944. Przyczynek do historii instytucji naukowo-społecznych w Polsce*, PWN, Warszawa 1959.

Właściwie w tym momencie rozpoczęła się krytyczna faza działalności IGS-u, którą jeden z jego współpracowników Tadeusz Szturm de Sztrem, opisywał w sposób następujący: „W latach 1921–1923 chodziło przede wszystkim o pokazanie i zdemaskowanie istotnej roli mechanizmu działającego wówczas w gospodarstwie społecznym Polski, za pomocą którego – jak się okazało – cały ciężar powojennej odbudowy kraju oraz dokonywanych inwestycji przerzucano na barki klasy pracującej. (...) W broszurach Instytutu chodziło o wykrycie rzeczywistej wartości stawek płac, o obliczenie ich prawdziwej siły nabywczej (...) potrzebie tej uczyniło zadość parę prac, (...) które rzuciły światło na rzeczywiste położenie klasy robotniczej w okresie narastającej drożyzny (...)”<sup>17</sup>. W powstałym w 1926 r. Statucie IGS-u tak opisane zostały jego zakładane funkcje społeczne: „(...) celem Instytutu jest oparcie gospodarstwa społecznego w Polsce na naukowych podstawach (...)”. Natomiast Ludwik Krzywicki we *Wstępie* do wydanych w 1936 r. *Pamiętników Chłopów*, dziele należącym do klasyki polskiej literatury socjologicznej, pisał, że „(...) Instytut Gospodarstwa Społecznego jest instytucją o celach natury poniekąd doraźnej: usiłuje (...) trzymać rękę na pulsie naszego życia społecznego (...) celem wyjaśnienia tego czy innego niedomagania, oświecenia tej lub innej sytuacji (...)”. Tak właśnie można sobie wyobrazić szeroko rozumiane prace ewaluacyjne, usiłujące wydobyć na użytek publicznej debaty rezultaty rozmaitych politycznych obietnic, które – przy bliższym oglądzie – okazują się często jedynie propagandowym frazesem.

Rezultaty prac prowadzonych w IGS-ie wydają się zgoła imponujące: do klasyki polskich badań społecznych zaliczają się systematyczne badania nad budżetami gospodarstw domowych czy analizy obciążenia podatkami pośrednimi rozmaitych grup społeczno-zawodowych oraz – pionierskie wtedy – badania nad migracjami zarobkowymi Polaków. To, co o warsztacie tych prac mielibyśmy do powiedzenia, współcześnie nie ma większego znaczenia. Doniosłość IGS-u tkwi w tym, że wnosił on do publicznych debat systematycznie uzasadniane punkty widzenia, które bliskie były idei socjologii krytycznej, więc – socjologii kwestionującej reguły politycznej poprawności i pryncypia ideologii sukcesu. Czegokolwiek zatem nie powiedziałyby się o ideologicznych przesłankach wyboru problematyki prowadzonych w Instytucie badań, nie da się zakwestionować tego, że nadawały one społecznej refleksji nader praktyczny wymiar. Więcej nawet – program ujawniania nieoczekiwanych czy starannie pomijanych konsekwencji jakiegokolwiek praktycznej polityki społecznej to nic innego, jak program kładący nacisk na „(...) oparcie gospodarstwa społecznego na naukowych podstawach (...)”. Nie ma większego znaczenia, czy – w zgodzie ze współczesną konwencją – powiemy, że jest to program systematycznie prowadzonych ewaluacji, czy może wybierzemy inny, sensownie brzmiący termin.

<sup>17</sup> Ibidem.

## Konkluzje

O badaniach ewaluacyjnych można mówić dwutorowo: jedni chcą w nich widzieć sposób na racjonalizację procesów kształtowania decyzji alokacyjnych, podejmowanych w wyspecjalizowanych segmentach administracji publicznej, drudzy zaś widzą w nich jedną z możliwych metod demokratyzacji tych procesów. Nie ma sensu spierać się o to, która z tych funkcji ewaluacji jest ważniejsza, warto natomiast dostrzec to, że rezultaty ewaluacji są jedną z ważnych form demokratycznej kontroli merytokratycznych establishmentów. Mówiąc nieco inaczej, wyniki ewaluacji – dzięki swojej funkcji krytycznej – to jeden ze sposobów znoszenia „demokratycznego deficytu”, o czym we współczesnych demokracjach mówi się coraz częściej. W demokratycznej retoryce zwykło się przyjmować, że deficyt demokracji jest zjawiskiem groźnym w tej mierze (pomijając już kwestie prawomocności), w jakiej pomniejsza jej przewidywalność – bo utrudnia budowanie consensusu. Publiczny dyskurs, deliberacja, rozmaite formy obywatelskiej partycypacji, mają służyć właśnie poszukiwaniu rozwiązań kompromisowych, a więc – consensualnych. Tymczasem empirycznie uchwytnie argumenty nie wskazują na to, że jest tak istotnie. Takie instytucjonalne rozwiązania, które umożliwiają rozmaite formy publicznej debaty, mogą mieć wprawdzie tę zaletę, że służą jako psychologiczne wentyle pozwalające na osłabianie społecznych napięć, ale nie zbliżają uczestników do osiągnięcia zadowalającego dla wszystkich kompromisu, więc, w konsekwencji – może się okazać – nie stabilizują systemu politycznego. Co więcej, nie ma większych wątpliwości co do tego, że są one źródłem rozmaitych kosztów transakcyjnych. Gdyby tak być miało rzeczywiście, to znaczenia nabiera argumentacja wspierająca taki sposób myślenia, który kładzie nacisk na korzyści płynące z merytokracji: (...) *what matters most to citizens is the securing of certain goods – such as high employment, economic growth and environment protection (...) what people want in these fields is expertise, efficiency and equity. They look for Pareto-efficient improvements that corrects for market failure (...) There should be consultation with affected parties, but this exercise is for information gathering not to promote democratic accountability. (...) Even at the domestic level, technical regulatory issues tend to be delegated to unelected expert bodies (...)*<sup>18</sup>. Podsumowując, nie jest wykluczone, że obywatele chcą być rządzeni dobrze – co w kontekście oznacza tyle, że zgodnie z regułami technokratycznej czy ekonomicz-

---

<sup>18</sup> Por. Willets P., „The Cardoso Report on the UN and Civil Society: Functionalism, Global Corporatism, or Global Democracy?”, *Global Governance* 2006, nr 12, s. 305–324; Moravcsik A., „In Defence of the Democratic Deficit: Reassessing Legitimacy in the EU”, *Journal of Common Market Studies* 2002, vol. 40, s. 727. Podobny punkt widzenia prezentują też amerykańscy obserwatorzy problemów instytucjonalnych UE, np. Giandomenico Majone, „Europe’s Democratic Deficit”, *The Question of Standards, European Law Journal* 1998, vol. 5, nr 4.



nej efektywności – niekoniecznie zaś chcą w procesach sprawowania władzy uczestniczyć. Problem nie sprowadza się jednak tylko do refleksji wokół pytania o to, „czego oczekują od władzy obywatele – demokratycznej partycypacji czy może Pareto-racjonalnego zarządzania?”. Rzecz jest znacznie bardziej podstawowa i polega na ustaleniu, jakie okoliczności i mechanizmy społeczne stabilizują system polityczny. Tu pojawia się możliwość rozsądnego kompromisu: obywatele nie garną się – w ramach rozmaitych instytucji demokracji partycypacyjnej – do kosztownego uczestniczenia w procesach zarządzania politykami publicznymi, ale jednak chcą kontrolować ich rezultaty. Czasem jest to możliwe bez angażowania złożonego warsztatu analitycznego, bo – koniec końców – to, czy dziury w drodze są załatane, czy nie, łatwo zauważyć! Ale nieco więcej wysiłku wymaga odpowiedź na pytanie, jakim kosztem zostały one załatane, i rozstrzygnięcie, czy warto było je łątać, biorąc pod uwagę natężenie ruchu. To wymaga już bardziej złożonych analiz. Możemy je nazwać studiami ewaluacyjnymi i cieszyć się z tego, że oto pojawił się nowy sposób kontrolowania arbitralności decyzji alokacyjnych. Możemy też z pełną pokorą uświadomić sobie, że wprawdzie warsztat ewaluacji został w ciągu ostatnich dziesięcioleci znakomicie rozbudowany, ale ich idea podstawowa nie jest nowa: ta formuła wykorzystania wiedzy gromadzonej w naukach społecznych pojawiała się w nich także i wtedy, gdy wchodząca w ich skład socjologia miała jeszcze wielce prenakowy charakter.

---

**Kazimierz Frieske** – doktor habilitowany, profesor Uniwersytetu Warszawskiego i Wyższej Szkoły Handlu i Prawa w Warszawie, a także warszawskiego Instytutu Pracy i Spraw Socjalnych. W Instytucie Socjologii WFiS UW kieruje Zakładem Problemów Społecznych i Planowania Społecznego. Jego prace badawcze mieszczą się w obszarze szeroko rozumianej socjologii problemów społecznych i socjologii prawa, a także tzw. socjologii stosowanej. Jest autorem bądź współautorem lub redaktorem szeregu książek, a poza tym wielu artykułów publikowanych w profesjonalnych czasopismach. Napisał m.in.: *Pijaństwo. Interpretacje problemu społecznego* (Warszawa 1984); *Narkomania* (Warszawa 1987); *Socjologia w działaniu. Nadzieje i rozczarowania* (Warszawa 1990); *Opieka i kontrola. Instytucje wobec problemów społecznych* (Warszawa 1996) *Zjawiska marginalizacji społecznej w Polsce* (Warszawa 1997). Zajmuje się też tłumaczeniami. Przełożył na język polski takie książki, jak: *Prawnik jako negocjator* L. Hawkinsa, M. Hudsona i R. Cornalla (Warszawa 1994); *Zarządzanie firmą prawniczą* Alana Pannetta (Warszawa 1996), *Administracja publiczna w systemie politycznym* B. Guy'a Petersa (Warszawa 1999). W wolnych chwilach tłumaczy literaturę rozrywkową.

## Bibliografia

- Assorodobraj N., „Życie i dzieło Stefana Czarnowskiego”, w: S. Czarnowski, *Dzieła*, PWN, Warszawa 1956, t. V.
- Comte A., *Rozprawa o duchu filozofii pozytywnej*, PWN, Warszawa 1973.
- Gouldner A.W., „Explorations in Applied Social Science”, w: A.W. Gouldner, S.M. Miller (red.), *Applied Sociology, Opportunities and Problems*, The Free Press, N.Y. 1964.
- Limanowski B., *Socjologia*, Gebethner i Wolf, Warszawa 1919.
- Limanowski B., „Socjologia”, w: J. Sztumski (red.), *Socjalizm jako konieczny objaw dziejowego rozwoju*, KiW, 1989.
- Majone G., „Europe’s Democratic Deficit”, *The Question of Standards, European Law Journal* 1998, vol. 5.
- Merton R., „Rola intelektualisty w instytucjach państwowych”, w: R. Merton, *Teoria socjologiczna i struktura społeczna*, PWN, Warszawa 1982.
- Moravcsik A., „In Defence of the Democratic Deficit: Reassessing Legitimacy in the EU”, *Journal of Common Market Studies* 2002, vol. 40.
- Supiński J., *Pisma. Wydanie poprawione i pomnożone*, t. IV, Warszawa 1872.
- Szacki J. (red.), *Sto lat socjologii polskiej*, PWN, Warszawa 1995.
- Szturm de Sztrem T., *Instytut Gospodarstwa Społecznego 1920–1944. Przyczynek do historii instytucji naukowo-społecznych w Polsce*, PWN, Warszawa 1959.
- Willems P., „The Cardoso Report on the UN and Civil Society: Functionalism, Global Corporatism, or Global Democracy?”, *Global Governance* 2006, nr 12.



*Dariusz Zalewski*

## **Możliwości i ograniczenia wykorzystania wyników ewaluacji przez instytucje publiczne**

### **„Diaboliczne ćwiczenie”**

Czym jest ewaluacja i jakie są jej funkcje wyjaśniano na wiele sposobów. Nie ma więc powodu, aby po raz kolejny przytaczać znane podręcznikowe definicje. Kariera, jaką zrobiło to pojęcie, wydaje się być nierozzerwalnie związana z dwoma zjawiskami. Po pierwsze – z poszerzaniem się sfery publicznej i wzrostem znaczenia administracji odpowiedzialnej za realizację określonych zadań. Po drugie – z rozczarowaniem efektywnością realizacji tychże zadań przez administrację zorganizowaną wedle klasycznych zasad sformułowanych przez M. Webera. Oczywiście, musimy pamiętać, że sama ewaluacja ma znacznie dłuższą historię – sięga ona początków XX wieku, a w jej rozwoju miał swój udział m.in. Florian Znaniecki – niż krytyka weberowskiej biurokracji i stosownie do niej poszukiwanie nowych form organizacyjnych administracji publicznej dla skuteczniejszego i efektywniejszego realizowania zadań. Należałoby też dla porządku dodać, że badania ewaluacyjne, jako narzędzie oceny realizowanych programów interwencji społecznej, stosowane są nie tylko w obrębie działalności administracji publicznej, ale to głównie jej specyfice czy też niedoskonałości zawdzięczamy w znacznej mierze ich rozwój.

Wzrost zainteresowania badaniami ewaluacyjnymi w Polsce jest ściśle związany z przystąpieniem Polski do Unii Europejskiej. Przyczyna jest prozaiczna i ma charakter obligatoryjnego wymogu ewaluowania programów publicznych finansowanych przez UE<sup>1</sup>. Czarujący powab ewaluacji widoczny jest niemalże w każdym oficjalnym dokumencie UE poświęconym tejże problematyce. W pierwszym zdaniu jednego z dokumentów Komisji Europejskiej możemy przeczytać: „Ewaluacja może być traktowana przez kogoś jako diaboliczne ćwiczenie. Jednakże, jeśli ewaluacja jest dobrze przeprowadzona

---

<sup>1</sup> *Inicjatywa Komisji Europejskiej Sound and Efficient Management 2000* (znana jako SEM 2000) włącza użycie ewaluacji jako kluczowego elementu poprawy kultury zarządzania samej Komisji. Kluczową innowacją SEM będzie wymaganie, aby systematyczną ewaluacją objąć wszystkie programy UE; zob. „Evaluating UE Expenditure Programmes: A Guide”, European Commission, 01.1997, s. 7.

i jeśli rezultaty ewaluacji są wykorzystane przez decydentów, może ona przyczynić się do poprawy publicznych programów, jak również zwiększyć ich przejrzystość, odpowiedzialność i efektywność<sup>2</sup>.

Chociaż ewaluacja pełni wiele funkcji, a jej wartość powinna być oceniana przez pryzmat złożoności funkcji i celów, jakie przed nią się stawia, to jednak z uwagi na fakt, że największym odbiorcą jej wyników jest administracja publiczna, najważniejsza wydaje się być ocena racjonalności poniesionych kosztów dla osiągnięcia określonych celów. Myśl tę dość dobrze wyraża następujące zdanie: „Ewaluacja ma dostarczać kryteriów, metod i środków do oceny racjonalności działań publicznych, spełniając w sektorze publicznym rolę analogiczną do mechanizmu rynkowego w sektorze prywatnym”<sup>3</sup>. Można by oczywiście argumentować, że taka redukcja funkcji badań ewaluacyjnych jest nieuprawniona, a porównanie ich roli do mechanizmu rynkowego jest w tej mierze nietrafne, w jakiej państwo różni się od rynku, a logika działań administracji publicznej jest odmienna niż w przypadku prywatnego przedsiębiorstwa. Jednak krytyka administracji publicznej, inspirowana głównie przez przedstawicieli neoliberalnej ekonomii klasycznej oraz teorii wyboru publicznego, sprawiła, że to właśnie rynek ze swoimi zasadami konkurencji i dążeniem do ekonomicznej efektywności stał się płaszczyzną odniesienia dla proponowanych reform w zarządzaniu sektorem publicznym. Natomiast rolą ewaluacji miałyby być odpowiedź na pytanie, czy realizowane programy publiczne są efektywne w ekonomicznym tego słowa znaczeniu.

Można sądzić, że oczekiwania formułowane pod adresem ewaluacji są zbyt duże. Jeśli tak, to jakie są, z jednej strony, możliwości, a z drugiej strony ograniczenia wykorzystania wyników ewaluacji w praktyce administracyjnej? Odpowiedzi na te pytania wydają się być interesujące w kontekście poszukiwanej efektywności działania sektora publicznego, owego „Świętego Grała” wszystkich reformatorów, jak zrećcznie ujął to V. Writgh<sup>4</sup>. Analiza będzie koncentrowała się na zewnętrznych ograniczeniach możliwości wykorzystania wyników ewaluacji pomimo świadomości, że jej aplikacyjna wartość zależy w równym, jeśli nie większym, stopniu od czynników wewnętrznych procesu ewaluacyjnego (poprawności założeń metodologicznych, przestrzegania standardów badawczych, kompetencji badacza itp.). Taki wybór podyktowany jest powszechnie formułowanymi pod adresem badań ewaluacyjnych oczekiwaniami (niezwykle sprzecznymi) ze strony różnych środowisk, aby ich wyniki przyczyniały się do wzrostu efektywności realizowanych programów publicznych.

<sup>2</sup> *Evaluating UE Expenditure Programmes...*, op. cit.

<sup>3</sup> Shadish W.R., Cook T.D., Leviton L.C., *Foundations of Program Evaluation: Theories of Practice*, Sage Publications, Newbury Park, London 1991, s. 19; cyt. za K. Olejniczak, „Teoretyczne podstawy ewaluacji *ex-post*”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka*, PARP, Warszawa 2007, s. 19.

<sup>4</sup> Wright V., „Paradoxes of Administrative Reform”, w: W.J.M. Kickert, *Public Management and Administrative Reform in Western Europe*, Edward Elgar, Cheltenham 1997, s. 11.

Rozdział ten jest tak pomyślany, aby w klarowny sposób przedstawić, jakie są oczekiwania wobec ewaluacji oraz jej możliwości i ograniczenia. Ponieważ znaczenie ewaluacji może być różne w różnych koncepcjach realizacji zadań publicznych, pokażemy w skrótowy sposób propozycje zmian w administracji publicznej i ewentualne konsekwencje dla ewaluacji programów interwencji społecznych. Byłoby wskazane, aby ów tekst inspirował dyskusję o możliwościach i ograniczeniach ewaluacji nie tylko wśród przedstawicieli środowiska naukowego, gdyż ci zawsze znajdują czas, by spierać się o coś, ale także wśród przedstawicieli administracji, odpowiedzialnej za realizację zadań o charakterze publicznym.

### **Słów kilka o reformach administracji i ich konsekwencjach dla ewaluacji**

To, jaką rolę ewaluacja może odegrać w procesie efektywnego wydawania środków publicznych, w jakim stopniu przyczynić się do racjonalnego podejmowania decyzji alokacyjnych, zależy w dużej mierze od tego, w jaki sposób i kto realizuje usługi o charakterze publicznym.

Przed prawie trzydziestu laty M. Friedman wspólnie ze swoją małżonką Rose opublikowali książkę *Free to Choose*, w której przystępnie przedstawili zasady rządzące racjonalnością alokacji zasobów finansowych, ze względu na dwie pary zmiennych: czyje pieniądze wydajemy (własne bądź cudze) oraz na kogo wydajemy (na siebie bądź na kogoś innego). W wyniku połączenia tych dwóch par alternatyw uzyskujemy macierz odpowiedzi, na którą składają się cztery warianty: własne pieniądze na własne potrzeby (1), własne pieniądze na potrzeby innych (2), cudze pieniądze na własne potrzeby (3) oraz cudze pieniądze na potrzeby innych (4)<sup>5</sup>. Krytykując nieracjonalność alokacyjną administracji publicznej, Friedmanowie skoncentrowali się na wariantach 3 i 4, dowodząc, że wydawanie publicznych pieniędzy stanowi przesłankę do marnotrawstwa tych środków, gdyż nie ma wystarczających bodźców do uruchomienia racjonalnych strategii działania tak, aby optymalizować uzyskiwane przez beneficjentów korzyści do poniesionych publicznych kosztów. W podobnym duchu wypowiadają się przedstawiciele teorii wyboru publicznego, którzy swoje analizy opierają na mikroekonomicznych założeniach neoliberalnej ekonomii. *Nobody spends somebody else's money as carefully as he spends his own* – pisał przed laty W. Mitchell, analizując funkcjonowanie sektora publicznego w ramach teorii wyboru publicznego<sup>6</sup>.

<sup>5</sup> Friedman M., Friedman R., *Wolny wybór*, Wyd. Panta, Sosnowiec 1994, s. 111.

<sup>6</sup> Cyt. za: Dunleavy P., *Democracy, Bureaucracy and Public Choice*, Harvester Wheatsheaf, 1991, s. 173.

Uwagi poczynione przez Friedmanów raczej nie przyczyniły się do rozwoju badań ewaluacyjnych, a ich intencją było pokazanie, że administracja publiczna głównie dba o własne, a nie publiczne interesy. Jeden z możliwych wniosków, jaki moglibyśmy wyprowadzić z neoliberalnej krytyki funkcjonowania sektora publicznego, byłby taki, że wyniki badań ewaluacyjnych mogłyby być pomocne, by nie powiedzieć rozstrzygające, w dyskusji nad efektywnością programów realizowanych przez instytucje publiczne. Potwierdza to przekaz płynący z rozmaitych wypowiedzi przedstawicieli politycznych establishmentów, szczególnie dobrze widoczny dla polityków UE, którzy wyniki ewaluacji traktują jako czynnik racjonalizacji procesu decyzyjnego w sektorze publicznym. „Ewaluacja jest niezbędną częścią nowoczesnej praktyki zarządzania sektorem publicznym” – pisał przed kilku laty E. Liikanen, członek Komisji Europejskiej odpowiedzialny za budżet i administrację<sup>7</sup>.

Oczywiście musimy pamiętać, że ewaluacja jest tylko narzędziem służącym ocenie rozmaitych programów społecznej interwencji, podejmowanych nie tylko przez administrację publiczną, i w żadnym przypadku nie można jej traktować jako funkcjonalnej recepty na rozmaite kłopoty, których nastręcza funkcjonowanie sektora publicznego. Jeśli jednak chcemy odpowiedzieć na takie czy inne zarzuty kierowane pod adresem instytucji publicznych, często z resztą uzasadnione, lub też dostarczyć samej administracji informacji o kosztach i efektach jej funkcjonowania, to wydaje się, że powinniśmy odwołać się do wyników oceny tejże działalności, które możemy uzyskać w ramach prowadzonych badań ewaluacyjnych. Należy jednak zdawać sobie sprawę nie tylko z rozmaitych ograniczeń wykorzystania wyników tych badań, ale też i z tego, że znaczenie samej ewaluacji może być różne ze względu na to, kto i w jaki sposób będzie realizował określone zadania tradycyjnie przynależne do sektora publicznego w ramach wzajemnych powiązań państwa i rynku.

### **„Urynkowienie państwa” a ewaluacja**

Propozycje radykalnych zmian w administracji publicznej, jakie pojawiły się m.in. w latach 80. ubiegłego wieku, głównie za sprawą ekonomistów ze Szkoły Chicagowskiej oraz zwolenników teorii wyboru publicznego, nawiązywały do zasad funkcjonowania wolnego rynku z jego logiką maksymalizowania korzyści. Doświadczenie było takie, że administracja publiczna nie radzi sobie z rozwiązywaniem rozmaitych problemów, do czego została powołana, a jej efektywność alokacyjna jest zdecydowanie mniejsza niż podmiotów prywatnych działających na rynku. Uzasadnieniu

<sup>7</sup> *Evaluating UE Expenditure Programmes...*, op. cit.

utrzymywania sektora publicznego i zarządzającej nim administracji zawartemu w haśle *market failure* przeciwstawiono hasło *government failure*. Porównując zatem działanie administracji publicznej do przedsiębiorstw rynkowych (czy słusznie to zupełnie inna sprawa) neoliberalni ekonomiści zaproponowali, po pierwsze, deregulację sektora publicznego, po drugie, jego prywatyzację<sup>8</sup>.

Deregulacja sektora publicznego miała na celu przede wszystkim zredukowanie odpowiedzialności państwa za rozwiązywanie rozmaitych problemów tak, aby zwiększyć przestrzeń dla działań rynkowych. „Jeśli publiczna regulacja – argumentował George Stigler – nie jest konieczna, zatem wiele programów publicznego sektora regulujących gospodarkę powinno zostać usuniętych, włączając w to agencje, które wprowadzają te publiczne regulacje”<sup>9</sup>. Krótko mówiąc, deregulacja, a w ślad za nią prywatyzacja agencji sektora publicznego, miały zapewnić większą konkurencyjność między jednostkami odpowiedzialnymi za realizację funkcji publicznych państwa, a funkcję ewaluacyjną ich działalności miały pełnić przede wszystkim reguły gry rynkowej ze swoją zdolnością do skutecznego rozwiązywania problemów alokacji zasobów. Kłopot tylko w tym, że rynkowe czy quasi-rynkowe instytucje alokują swoje zasoby tam, gdzie istnieje możliwość uzyskania wysokiej stopy zwrotu zainwestowanych środków. Poza tym nie ma żadnych gwarancji, że konkurencja między podmiotami rywalizującymi o realizację określonych usług o charakterze publicznym będzie prowadzona wedle zasad bezstronności. Abstrahując jednak od wielu problemów i krytyk propozycji ekonomistów Szkoły Chicagowskiej czy przedstawicieli teorii wyboru publicznego, jak na przykład takich, że deregulacja wymaga nowych form regulacji<sup>10</sup>, a prywatyzacja zwiększa koszty transakcyjne, należy zauważyć, że jedną z konsekwencji tych propozycji powinno być zmniejszenie roli badań ewaluacyjnych. Tę rolę powinien w znacznym stopniu przejąć rynek ze swoją czytelną zasadą ekonomicznej efektywności. Prywatyzacja bowiem stanowi odwrócenie procesu poszerzania państwowego interwencjonizmu, który to proces był jedną z kluczowych przesłanek rozwijania badań ewaluacyjnych czy szerzej, implementacji dorobku nauk społecznych do rozwiązywania konkretnych problemów<sup>11</sup>.

<sup>8</sup> Lane J.E., „Public Sector Reform: Only Deregulation, Privatization and Marketization?”, w: J.E. Lane (red.), *Public Sector Reform. Rationale, Trends and Problems*, Sage Publications, London 1997, s. 3.

<sup>9</sup> Ibidem.

<sup>10</sup> Majone G., *The rise of the regulatory state in Europe*, West European Politics, 17, s. 77–101.

<sup>11</sup> K.W. Frieske np. podaje, że początki zaangażowania przedstawicieli nauk społecznych w „praktykę społeczną” w Stanach Zjednoczonych miały miejsce w okresie New Dealu, kiedy państwo zwiększyło obszar swojego interwencjonizmu; zob. Frieske K.W., *Socjologia w działaniu. Nadzieje i rozczarowania*, Wyd. Uniwersytetu Warszawskiego, Warszawa 1990, s. 78–79.



## „Urynkowienie władzy” a ewaluacja

Propozycje zmian w sektorze publicznym złożone przez neoliberalnych ekonomistów z pewnością stanowiły poważne zagrożenie dla interesów ewaluatorów, a w jakiejś mierze także i dla samych badań ewaluacyjnych, gdyż rychło mogłoby się okazać, że nie ma czego ewaluować albo nie byłoby dla kogo ewaluować. Takiego zagrożenia z pewnością nie ma ze strony reformatorskiej koncepcji „publicznej przedsiębiorczości” (*public entrepreneurship*), której bodaj najbardziej znanymi orędownikami stali się w latach 90. ubiegłego wieku D. Osborne i T. Gaebler, autorzy bardzo głośnej książki *Reinventing Government*. Sama propozycja, wpisująca się w szerszą strategię „nowego zarządzania publicznego” (*New Public Management*) nie była ani specjalnie nowa, ani oryginalna, a czerpała zarówno z założeń teorii wyboru publicznego (opartej na neoliberalnym modelu racjonalności jednostki), jak i neotaylorowskiej teorii organizacji i zarządzania. To, co ewidentnie różni koncepcję „publicznej przedsiębiorczości” od koncepcji urynkowienia usług publicznych, pochodzącej ze Szkoły Chicagowskiej, zawiera się w różnym podejściu do sił rynkowych, które miały zreformować sektor publiczny. O ile neoliberalni ekonomiści proponowali ograniczyć działalność państwa, aby zwiększyć przestrzeń dla sektora prywatnego, o tyle „menedżeryści” zaproponowali włączenie mechanizmów rynkowych do działalności administracji tak, aby stworzyć tzw. wewnętrzny rynek w ramach sektora publicznego. Nawiązując do neotaylorowskiej koncepcji efektywnego zarządzania, zaproponowali oni model administracji zorientowanej na osiąganie celów, a w szczególności położyli nacisk na wykorzystywanie wskaźników wydajności i skuteczności, ewaluację realizowanych programów, wynagradzanie wedle zasady uzyskiwanych rezultatów i podnoszenie jakości świadczonych usług<sup>12</sup>.

Osborne i Gaebler, zachęcając do systematycznej ewaluacji realizowanych przez administrację publiczną programów, wskazali jednocześnie na szereg problemów, jakie zwykle towarzyszą pomiarowi oczekiwanych wyników. Ich uwagi, sformułowane na bazie doświadczenia administracyjnej praktyki, miały przestrzegać, jakich błędów przy dokonywaniu pomiaru osiągniętych wyników należy unikać. Warto do nich odnieść się chociażby po to, aby pokazać, nie kwestionując ich merytorycznej zasadności, że ani nie wyczerpują one katalogu możliwych do popełnienia błędów, na jakie narażone jest wnioskowanie na podstawie wyników ewaluacji, ani zastrzeżeń, jakie pod adresem ewaluacji kierują rozmaici badacze.

---

<sup>12</sup> Keraudren Ph., van Mierlo H., „Theories of Public Management Reform and their Practical Implications”, w: T. Verheijen, D. Coombes, *Innovations in Public Management. Perspectives from East and West Europe*, Edward Elgar, Cheltenham 1998, s. 40.

- Po pierwsze, „między mierzaniem procesu działania a mierzaniem jego wyników jest wielka różnica”<sup>13</sup>. Uwaga ta brzmi dość banalnie dla każdego, kto przynajmniej w stopniu podstawowym jest wtajemniczony w arkaną ewaluacyjnego rzemiosła, choć niekoniecznie musi być równie oczywista dla przedstawicieli administracji. W gruncie rzeczy autorom chodziło o to, że administracja często koncentruje się na pomiarze nie tego, co powinno być przedmiotem oceny. „Skłonność do koncentrowania się na jakimś procesie jest naturalna – menedżerowie mierzą to, co robią powierzone im organizacje, a w organizacjach kierujących się regulaminami i przepisami ludzie traktują swoją pracę jako pewien proces, którego przebieg wyznaczają te przepisy. Jeśli zatem przestrzegają dokładnie przepisów i wytwarzają tyle, ile się od nich oczekuje, to wywiązują się ze swoich zadań. Rzadko myślą o wynikach swojej pracy, o tym, jaki wpływ ma ona na tych, którym dana agencja ma służyć. Jednakże taka, nawet doskonale wykonana praca, jest stratą czasu i pieniędzy, jeśli nie prowadzi do osiągnięcia pożądaných wyników”<sup>14</sup>.
- Po drugie, „istnieje wielka różnica między mierzaniem wydajności a mierzaniem skuteczności”<sup>15</sup>. Dla każdej jednostki organizacyjnej, powiadają autorzy, ważne powinno być skuteczne osiągnięcie zakładanego celu. Wydajność nie jest i nie powinna być celem, do jakiego zmierza organizacja, a jedynie środkiem do jego osiągnięcia. Koncentracja uwagi na wydajności (i jego pomiarze) w oderwaniu od skuteczności działania prowadzi do paradoksalnej sytuacji marnotrawstwa środków. „Nie ma nic głębszego niż wydajniejsze robienie czegoś, czego nie powinno się w ogóle robić”<sup>16</sup>. Co więcej, Osborne i Gaebler twierdzą, że: „koncentrowanie się na wydajności prowadzi również do wytworzenia u pracowników instytucji publicznych poczucia alienacji. Kiedy władze podkreślają koszt każdej jednostki pracy, to często osiągają tylko tyle, że pracownicy widzą jedynie swój odcinek, co ogranicza ich inteligencję i umiejętności. (...) Jeśli jednak ich przełożeni koncentrują się wyłącznie na ich wydajności – na tym, jak szybko wykonują każdą jednostkę pracy – to zaczynają się oni czuć, jakby byli zatrudnieni przy taśmie montażowej”<sup>17</sup>. Na marginesie warto dodać, iż wyraźnie widać, że autorzy przestrzegając przed negatywnymi konsekwencjami pomiaru tylko wydajności pracy, wychodzą poza tradycję neotaylorizmu, a nawiązują do tradycji badań z kręgu tzw. *Organizational humanism* – wyznaczonej przez takie nazwiska, jak John Dewey, Elton Mayo, Chester Bernard

<sup>13</sup> Osborne D., Gaebler T., *Rządzić inaczej. Jak duch przedsiębiorczości przenika i przekształca administrację publiczną*, Media Rodzina, Poznań 1992, s. 467.

<sup>14</sup> Ibidem.

<sup>15</sup> Ibidem, s. 468.

<sup>16</sup> Ibidem, s. 469.

<sup>17</sup> Ibidem.

czy Kurt Levin – zorientowanych na eksplorację relacji międzyludzkich wewnątrz organizacji, m.in. po to, aby testować związek między poziomem satysfakcji oraz stopniem zaspokojenia społecznych i emocjonalnych potrzeb członków organizacji a osiąganymi przez nią rezultatami<sup>18</sup>.

- Po trzecie, „istnieje też ważna różnica między «wynikami programu» a obejmującymi szerszy zakres «wynikami polityki»<sup>19</sup>. Według Gaeblera i Osborna to, co jest ważne ze społecznego punktu widzenia, odnosi się do rezultatów szerszej rozumianej „polityki”, której poszczególne programy są tylko częścią. Wymowny jest przykład, jaki podają autorzy. Otóż instytucje realizujące programy szukania miejsc pracy dla klientów systemu pomocy społecznej często obliczały liczbę osób, dla których udało się ją znaleźć, oraz wysokość ich zarobków, jednakże nie uwzględniały liczby osób przechodzących na zasiłek dla bezrobotnych, długości okresu pobierania zasiłku itd. Jakie zatem były tego konsekwencje? „Osoby bezpośrednio nadzorujące program mogą odrzec, że są odpowiedzialne za liczbę znalezionych miejsc pracy, ale nie za liczbę osób przechodzących na zasiłek i będą miały rację. Na wielkość tej drugiej liczby wpływa bowiem ogólna polityka w tej dziedzinie, to znaczy zasady ustalające, kto kwalifikuje się do uzyskania zasiłku, w jakiej relacji pozostaje zasiłek do wynagrodzenia za wymagającą małych kwalifikacji i nisko płatną pracę i ile można znaleźć takich miejsc pracy. Jeśli jednak wydział opieki nie mierzy obu zbiorów liczb, to jego pracownikom może się wydawać, że wykonują znakomitą robotę, znajdując pracę dla swoich klientów i tym samym pozbywając się ich, gdy tymczasem szeregi podopiecznych w rzeczywistości rosną. Tak właśnie dzieje się od dziesięciu lat w wielu stanach. (...) Takie statystyki podkreślają wagę mierzenia zarówno wyników programu, jak też polityki”<sup>20</sup>.

Gaebler i Osborne formułują też kilka praktycznych zaleceń dla administracji publicznej, która powinna być zainteresowana pozyskaniem wartościowych informacji pochodzących z badań ewaluacyjnych.

- Pierwsze zalecenie, które brzmi *przeprowadzaj zarówno analizę ilościową, jak i jakościową*, ma swoją bardzo obszerną literaturę i nie ma powodu aby cokolwiek w tej sprawie dodatkowo wyjaśniać<sup>21</sup>.
- Po drugie, *uważaj na amatorów łatwych kasków*. Chodzi o to, aby nie pozwolić zwieść się pozorom łatwych rozwiązań, które miałyby świadczyć o skuteczności takiego czy innego programu. Odwołując się do praktyki

<sup>18</sup> Szerzej na ten temat m.in. w McKinney J.B., Howard L.C., *Public Administration. Balancing Power and Accountability*, Praeger Publishers, Westport 1998, s. 159–165.

<sup>19</sup> Osborne D., Gaebler T., op. cit., s. 470.

<sup>20</sup> Ibidem, s. 471.

<sup>21</sup> Jasiński M., Kowalski M., „Falszywa sprzeczność: metodologia jakościowa czy ilościowa?”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka*, PARP, Warszawa 2007.

amerykańskiego sektora publicznego, autorzy piszą tak: „Usługodawcy zazwyczaj podają takie liczby, jakich się od nich oczekuje, nawet jeśli muszą w tym celu uprościć sobie pracę albo obejść przepisy. Jeśli muszą znaleźć pracę dla tysiąca osób rocznie, to wybiorą tysiąc osób, dla których najłatwiej znaleźć pracę i przeszkolą je. (...) Tak właśnie postępowały ośrodki pomocy społecznej w pierwszych latach realizacji ustawy o współudziale w kształceniu zawodowym”<sup>22</sup>.

- Po trzecie, *bądź przygotowany na silny opór*, co oznacza mniej więcej tyle, że rzetelne informacje o wydajności i skuteczności mogą stanowić duże zagrożenie dla bytu danej jednostki organizacyjnej. Autorzy podają kilka przykładów takiej sytuacji, np. Stanu Floryda, który zlikwidował program nagradzania szkół za poprawę jakości pracy czy Stanu Arizona, który zaprzestał publikowania danych dotyczących odsetka absolwentów szkół zawodowych i pomaturalnych, którzy uzyskali pracę. Powód był prozaiczny – „college i szkoły zawodowe uznały bowiem, że informacje takie stanowią dla nich zagrożenie, jako że ujawniają, jak naprawdę przygotowują młodych ludzi do podjęcia pracy”<sup>23</sup>.
- Po czwarte, *włączaj w proces tworzenia metod pomiaru usługodawców i podwładnych*. To zalecenie odwołuje się do znanych z teorii organizacji i zarządzania modeli, jak chociażby przywoływany wcześniej *Organizational humanism*, zalecających kooperację dla wypracowania wspólnych rozwiązań i uniknięcia konfliktów wewnątrz-organizacyjnych. „Najlepszym – jak piszą autorzy – sposobem przełamania oporu przed wprowadzeniem metod mierzenia wykonania i jakości pracy jest włączenie usługodawców i wszystkich pracowników w proces ich tworzenia”<sup>24</sup>.
- Po piąte, *poddawaj metody pomiaru corocznej rewizji i modyfikacji*.
- Po szóste, *nie stosuj zbyt dużo ani zbyt mało miar*. To istic arystotelesowskie zalecenie umiaru opiera się na przekonaniu, że jest on możliwy do wypracowania w ramach kooperacyjnej strategii tworzenia miar. „Jeśli (...) pracownicy i usługodawcy biorą czynny udział w tworzeniu metod pomiaru i jeśli pozwala się im na okresowe ich korygowanie, to zazwyczaj potrafią zachować umiar w tworzeniu miar”<sup>25</sup>.
- Po siódme, *strzeż się antybodźców*. Chodzi o to, aby unikać rozwiązań, które prowadzą do nieefektywnych, z punktu widzenia „całościowej polityki”, rezultatów. „Metody pomiaru wykonania pracy przyjęte początkowo w ustawie o współudziale w kształceniu zawodowym zachęcały usługodawców do wyszukiwania łatwych kąsków. Takie antybodźce mogą obrócić w niwecz wszystkie wysiłki zmierzające do stworzenia miar wy-

<sup>22</sup> Osborne D., Gaebler T., op. cit., s. 476.

<sup>23</sup> Ibidem, s. 477.

<sup>24</sup> Ibidem.

<sup>25</sup> Ibidem, s. 478.

konania pracy. Dla ich uniknięcia organizacje, przed wprowadzeniem nowych miar, powinny je przetestować, starając się znaleźć wszelkie możliwe sposoby robienia uników, które stosowałiby sprytni usługodawcy<sup>26</sup>.

- W kolejnym, ósmym zaleceniu, autorzy namawiają, aby „skupić się na maksymalnym wykorzystaniu danych o wykonaniu pracy. Samo stworzenie miar nie daje jeszcze gwarancji, że menedżerowie wykorzystają je do zmiany sposobu działania kierowanych przez siebie instytucji ani że ciała ustawodawcze skorzystają z nich, zmieniając kierunki i sposoby finansowania programów i instytucji<sup>27</sup>. Krótko mówiąc, nie należy zwlekać z wykorzystaniem dostępnym danych i stosować „politykę faktów dokonanych” w oparciu o zgromadzone informacje dotyczące efektywności i skuteczności działania.
- Po dziewiąte, „zlecaj mierzenie politycznie niezależnej, obiektywnej instytucji. Jeśli ludzie mają polegać na danych, to muszą wierzyć, że są one obiektywne. Dlatego dobrym rozwiązaniem jest korzystanie z usług niezależnej instytucji<sup>28</sup>. Realizacja tego zalecenia napotyka jednak na pewne trudności. Kłopot nie polega tylko na tym, czy „wiara” w obiektywność dostarczonych informacji, jak sugerują Gaebler i Osborne, czyni je rzeczywiście obiektywnymi. To fundamentalne dla nauki pytanie o „obiektywność” rzeczywistości i możliwość jej poznania możemy w tym miejscu uchylić, gdyż intencje autorów są bardzo klarowne. Chodziło im bowiem o to, aby pomiaru wydajności i skuteczności realizowanych programów dokonywały zewnętrzne w stosunku do wykonawców, równocześnie niezależne od nich, organizacje. Obiektywność oceny osiągniętych wyników należałoby zatem traktować jako funkcję niezależności instytucji oceniającej. I tu pojawia się problem z tzw. niezależnością instytucji oceniającej, która mogłaby dostarczyć administracji obiektywnych informacji. Można by oczywiście powiedzieć, zgodnie z sugestią autorów, że niezależność oznacza tyle, co brak personalnych i kapitałowych powiązań instytucji oceniającej z organizacją realizującą dany program. Ale taka redukcjonistyczna wykładnia „instytucjonalnej niezależności” jest zawodna w tej mierze, w jakiej nie uwzględnia złożoności rozmaitych punktów widzenia i interesów w sferze publicznej. Nie widać specjalnie dobrych uzasadnień, dla których tzw. niezależna instytucja (wszystko jedno czy prywatna czy publiczna), dokonująca ewaluacji programu, nie miałaby uwzględniać w swojej ocenie oczekiwań zleceniodawcy. Niewątpliwie uwzględnienie punktu widzenia czy interesów zleceniodawcy powinno mieć wpływ na ostateczny rezultat ewaluacji, nawet jeśli jej wyniki oparte byłyby na dobrze przygotowanych narzędziach badawczych. Ostateczny wynik

<sup>26</sup> Ibidem.

<sup>27</sup> Ibidem, s. 479.

<sup>28</sup> Ibidem, s. 478–479.

ewaluacji jest nie tylko, a może nawet nie przede wszystkim, rezultatem zastosowania określonych standardów badania, co ich ustalenia w fazie przedbadawczej.

Co więcej, można by dowodzić, że decyzje w sferze publicznej najczęściej kształtowane są nie wedle empirycznych ustaleń wolnych od osobistych wartości i poglądów decydentów, ale zgodnie z nimi. Wyniki rozmaitych badań ewaluacyjnych wykorzystywane w dyskusjach o polityce penalizacyjnej, edukacji seksualnej, aborcji itd. głównie służą jako instrument wzmacniający wyznawane przez strony poglądy czy wartości i zwykle kwestionuje się obiektywność wyników niezgodnych z wyrażanymi poglądami, niż zmienia poglądy. Dotychczas brakuje jakichkolwiek przesłanek (i empirycznych dowodów), by sądzić, że dobrze przeprowadzone przez niezależne instytucje badania ewaluacyjne, nad np. programami resocjalizacyjnymi groźnych przestępców, skłonią zwolenników kary śmierci do porzucenia swoich poglądów „penalizacyjnych” czy też odwrotnie – z zagorzałych zwolenników resocjalizacji uczynią równie gorliwych zwolenników kary śmierci. Krótko mówiąc, problem niezależności i obiektywności wyników badań ewaluacyjnych uwikłany jest w strukturę interesów i wartości wyznaczających dynamikę procesów decyzyjnych w sferze publicznej.

## Ewaluacyjne nadzieje i rozczarowania

G. Majone, znany ze swojego sceptycyzmu wobec możliwości wyznaczania racjonalności procesu decyzyjnego w sferze publicznej wedle czysto empirycznych ustaleń badań ewaluacyjnych, podkreśla, że „profesjonalna ewaluacja stanowi zaledwie niewielką część ogólnego procesu krytyki i oceny polityk publicznych, w którym wszyscy politycznie aktywni członkowie demokratycznej wspólnoty biorą udział na różne, lecz równie użyteczne sposoby”<sup>29</sup>. Dostrzegając znaczenie ewaluacji, jako rozległej i rozwijającej się dziedziny analizy polityki poświęconej „gromadzeniu, sprawdzaniu i interpretowaniu informacji na temat wdrażania i skuteczności istniejących polityk i programów publicznych”, Majone twierdzi, że dla efektywnego świadczenia usług publicznych nie wystarczy przygotowanie jakiegoś teoretycznie optymalnego programu działania. „Jeszcze ważniejsze jest poznanie, jak faktycznie przebiega wdrażanie programu, kto z niego korzysta, a kto na nim traci, czy program realizuje zamierzone cele, a jeśli tak się nie dzieje – to jak można go poprawić lub przerwać”<sup>30</sup>. Przykładowo, ocena szczególnie popularnych w krajach UE

<sup>29</sup> Majone G., *Dowody, argumenty i perswazja w procesie politycznym*, Wyd. Naukowe Scholar, Warszawa 2004, s. 249.

<sup>30</sup> *Ibidem*, s. 248.

programów aktywizacji bezrobotnych nie zawiera zwykle informacji o tym, kto na nich zyskuje, a kto traci, gdyż milcząco zakłada się, że zwiększenie zatrudnienia leży w interesie wszystkich, a zatem uzasadnione jest, aby koszt takich szkoleń wpisać w koszty opodatkowania obywateli.

Można by argumentować, że przedstawiona kontrowersja jest raczej pozorna niż rzeczywista, gdyż ustalenie „wygranych” i „przegranych” w wyniku realizacji określonych programów czy polityk publicznych, stopnia realizacji zamierzonych celów czy sposobów wdrażania rozmaitych programów, należy do kanonu zadań ewaluacji. Trudno jednak nie zgodzić się z Majone, który twierdzi, że: „Wydaje się, że wielu ewaluatorów zakłada, iż są to czysto empiryczne ustalenia, które nie wiążą się ani z wyborami wartości, ani z osobistymi poglądami. W rzeczywistości wartości i poglądy bardzo się liczą w ewaluacji, nie tylko ze względu na niejednoznaczność wyników w praktyce – trudności z przypisaniem określonych przyczyn konkretnym rezultatom, z pomiarem wyników i oceną niezamierzonych skutków, z rozróżnieniem błędnych koncepcji i niepowodzeń we wdrażaniu – lecz jeszcze w większym stopniu wskutek nieuniknionych różnic poglądów na temat rodzaju kryteriów ewaluacyjnych, które są sensowne, uzasadnione lub politycznie dopuszczalne w danej sytuacji. Takich niejasności i różnic zdań nie można w pełni rozwiązać za pomocą ulepszonych metod pomiarowych i sprawdzających, lecz można je przedstawić i wyjaśnić poprzez dyskusję i wzajemną perswazję”<sup>31</sup>.

Argumenty Majone bardzo wyraźnie wskazują na ograniczone możliwości perswazyjne wyników badań ewaluacyjnych i na niewiele, jak sugeruje autor, zda się poszukiwanie coraz bardziej wyrafinowanych metod pomiaru. Kłopot polega na tym m.in., że „wyjściowy” rezultat ewaluacji zależy od tego, jakim materiałem „wejściowym” dysponują ewaluatorzy, a jego zebranie to nie tylko kwestia zastosowania takich czy innych technik badawczych, ale także pewnych uzgodnień punktów widzenia, wartości czy interesów między zainteresowanymi realizacją programu i jego ewaluacją stronami. Stosowanie określonych standardów badawczych jest kompletnie czym innym niż ustalanie tychże i aby zdawać sobie sprawę z możliwości perswazyjnych osiągniętych rezultatów ewaluacji, konieczne jest, jak twierdzi Majone, „rozróżnienie ustalania standardów i stosowania standardów”<sup>32</sup>.

Dyskutowanie określonych standardów ewaluacyjnych, np. uzgodnienie wśród zainteresowanych stron podstawowych kryteriów oceny programu ze wszelkimi tego konsekwencjami dla ostatecznych rezultatów ewaluacji może uchronić ewaluatorów przed rozmaitymi zarzutami, np. że nie mierzą tego, co powinni, albo, że nie uwzględnili istotnych aspektów ewaluowanego programu bądź też problemu, którego rozwiązanie tenże program miał przynieść. Takie zarzuty mogą być szczególnie często podnoszone, kiedy wyniki ewa-

<sup>31</sup> Ibidem, s. 249.

<sup>32</sup> Ibidem, s. 249.

luacji zagrażają mniej lub bardziej wyartykułowanym interesom grupowym zaangażowanych w realizację programu czy też dotkniętych skutkami jego implementacji<sup>33</sup>. Powinni szczególnie dobrze pamiętać o tych zagrożeniach wszyscy ci, którzy realizują programy naruszające określony układ interesów, także wówczas, gdy dysponują empirycznymi dowodami, że utrzymywanie *status quo* jest niekorzystne dla wszystkich, a proponowana zmiana spełnia kryteria postępu wedle Pareto. Odwoływanie się jednak do argumentu, że „liczby nie kłamią”, może być równie irytujące i niezrozumiałe dla takiej czy innej grupy, jak podejmowane przed laty próby przekonywania „opornych” robotników przez żarliwych marksistów, że są oni nosicielami „fałszywej świadomości klasowej”, ale dla ich dobra trzeba to zmienić w ramach „zobiektywizowanych” procesów dziejowych.

Przyjęcie określonych kryteriów ewaluacyjnych, np. efektywności i skuteczności działania, w oczywisty sposób warunkuje ostateczny rezultat oceny programu. To, jakie kryteria oceny zostają przyjęte, to w dużej mierze konsekwencja ról społecznych, jakie odgrywają aktorzy zaangażowani w proces ewaluacji. Ten sam program inaczej będzie oceniany przez ekonomistę, dla którego najważniejsze jest poszukiwanie efektywności i skuteczności realizowanego programu, a inaczej przez prawnika, dla którego ważna jest np. ocena jego zgodności z obowiązującym prawem. Nie ma niczego złego w tym, że programy oceniane są z różnych punktów widzenia. Ważne jest, aby zachować odrębność różnych punktów widzenia, bowiem, „dopóki poglądy wyrażane z perspektywy jednej konkretnej roli nie są przedstawiane ani mylnie interpretowane jako poglądy odnoszące się do wszystkich możliwych ról lub wygłaszane w ich imieniu, mamy do czynienia ze zdrowym stanem wielokrotnej lub pluralistycznej ewaluacji”<sup>34</sup>. Kłopot tylko w tym, że w praktyce ewaluacyjnej różne punkty widzenia nakładają się na siebie, a błędem jest zakładanie pewnej uniwersalności wniosków wypływających z ewaluacji w tym sensie, że powinny być one równie istotne dla aktorów odgrywających różne społeczne role. „Być może najczęstszy problem występuje, kiedy zakłada się, że wnioski z ewaluacji dokonanej na użytek określonej roli są równie istotne z perspektywy innych ról stosujących odmienne kryteria ewaluacyjne. Ponieważ role i kryteria do siebie nie pasują, jest niemal nieuniknione, że wnioski z ewaluacji okażą się nieprzekonujące”<sup>35</sup>.

---

<sup>33</sup> „Kluczowy dylemat reformatorów można określić bardzo prosto: większość badań z zakresu public policy podkreśla fakt, że skuteczne reformy muszą uzyskać akceptację grup, których reformy najbardziej dotyczą. Jednakże wielu z tych, których reformy dotkną, są albo niechętni, albo wrogo nastawieni do reform i to nie jest zaskoczeniem, ponieważ ich miejsca pracy mogą być zagrożone; tylko najbardziej ekscentryczne indyki wyglądają świąt”; zob. Wrigth V., „The Paradoxes of Administrative Reform”, w: W. Kickert (red.), *Public Management and Administrative Reform in Western Europe*, Edward Elgar, Cheltenham 1997, s. 9.

<sup>34</sup> Majone G., op. cit., s. 251.

<sup>35</sup> Ibidem.



Szczególny kłopot wydaje się występować wówczas, kiedy nacisk kładzie się na nieco „zmitologizowaną” efektywność i skuteczność osiągania zakładanych celów, przy czym prawie zawsze w takich przypadkach chodzi o tzw. racjonalizację publicznych wydatków, a w zasadzie o ograniczenie kosztów. Szczególny dlatego, że kryteriom efektywności i skuteczności najczęściej przypisuje się najważniejszą rolę w procesie oceny publicznych programów, często zapominając, że ich podnoszenie, zgodnie z ekonomicznym standardem racjonalności, może naruszać wiele żywotnych interesów czy podzielną wartość głównych aktorów, których program dotyczy. Z drugiej zaś strony, nieuwzględnienie tychże interesów może sprawić, że dążność do racjonalizacji decyzji może prowadzić do skutków odwrotnych od zamierzonych i znacznego marnotrawstwa środków (*casus* obwodnicy Augustowa).

Zalecane w raportach ewaluacyjnych sposoby podniesienia efektywności i skuteczności mogą być też sprzeczne z panującą w organizacji, realizującą dany program, kulturą pracy czy obowiązującymi w niej standardami profesjonalizmu. Co więcej, jak wskazuje V. Wright, analizując paradoksalne konsekwencje podejmowanych reform administracji publicznej, ewaluacja może przynosić zgoła nieoczekiwane rezultaty z punktu widzenia podstawowego celu, jakim jest tzw. racjonalizacja wydatków publicznych. „Czwartym paradoksem jest to, że podczas gdy zamierzeniem wielu reform jest cięcie kosztów, ostateczny rezultat niektórych z nich może okazać się całkiem odmienny. Weźmy na przykład «ewaluację», to współczesne szaleństwo, która ignoruje koszty siły roboczej. Ewaluacja może identyfikować obszary marnotrawstwa środków, ale odkrywa także różnice nie tylko w wydajności (*performance*), lecz również osiągniętych ogólnych wyników (*outcomes*): pewni ludzie mogą być jednak relatywnie gorzej traktowani. Te różnice były wcześniej ukrywane. Rezultatem będzie wywieranie politycznej presji, aby takie różnice usuwać. Zgodnie z zamierzeniem, przekształcanie obywateli w klientów, z jasno określonymi przez cele prawami, może mieć wpływ na poprawę jakości realizowanych programów, lecz to także tworzy oczekiwania i presję zadośćuczynienia (ujawnionych nierówności – przyp. D.Z.), często na drodze sądowej”<sup>36</sup>. Jeśli uwaga Wrighta jest trafna, to nadzieje, że dzięki stałej ewaluacji programów publicznych czy działań administracji uda się zwiększyć ich ekonomiczną efektywność, wydają się być płonne w tej mierze, w jakiej nie uwzględniają one złożoności relacji międzyludzkich i powielają błędne przekonania o prostej relacji bodźca i efektu, zawarte w konwencjonalnych mechanistycznych teoriach organizacji i zarządzania.

G. Majone utrzymuje, że oprócz uwzględnienia ról, pozwalających na różnicowanie kryteriów ewaluacji, użyteczne jest też odniesienie się do trzech

<sup>36</sup> Wright V., „Paradoxes of Administrative Reform”, w: W.J.M. Kickert, *Public Management and Administrative Reform in Western Europe*, Edward Elgar, Cheltenham 1997, s. 10.

różnych trybów ewaluacji: trybu wynikowego, nakładowego i procesowego<sup>37</sup>. Szczególnie atrakcyjny wydaje się być tryb wynikowy, bowiem ocena dokonywana na podstawie osiągniętych wyników najbardziej apeluje do wyobraźni decydentów. Należy przypomnieć, że wszelkie próby racjonalizowania wydatków publicznych i optymalizowania wyników podejmowanych działań w sferze publicznej opierają się na dążeniu do pomiaru rezultatów i ich kwantyfikowaniu. Nie bez powodu płaszczyzną odniesienia dla krytyków działalności administracji publicznej jest rynek oraz prywatne przedsiębiorstwo z przypisywaną im zdolnością do pomiaru osiągniętych wyników. Założenie jest logiczne, jak też wbudowane w metodologiczny kanon ewaluacji, a mianowicie, aby móc możliwie najlepiej ocenić realizowany program, trzeba dokonać pomiaru jego efektów. Odwołanie się do czysto ekonomicznych zasad oceny programów społecznej interwencji jest jednak kłopotliwie i to nie tylko dlatego, że logiki funkcjonowania rynku i państwa oraz funkcje, jakie każde z nich pełni, są odmienne. Kłopot polega m.in. na tym, że wyniki nie zawsze poddają się kwantyfikacji, cele nie zawsze są jednoznacznie określone, a narzędzia pomiaru często zawodzą. Jeszcze większą trudność jednak stanowi to, że związek między zarejestrowanym wynikiem a jego przyczyną często jest niepewny i bardziej polegamy na intuicji niż naukowych dowodach.

E. Babbie słusznie zauważa, że: „potencjalnie jednym z najbardziej uciążliwych aspektów badań ewaluacyjnych jest ustalenie, czy badany program odniósł sukces, czy poniósł porażkę. Celem programu nauczania języka obcego może być pomoc studentom w lepszym nauczaniu się tego języka, lecz jak duży postęp jest tu wystarczający? Celem programu odwiedzin małżonków w więzieniu może być poprawa nastrojów wśród więźniów, lecz jak bardzo mają być one poprawione, by móc uzasadnić ten program? Jak można się spodziewać, nie ma ostatecznych odpowiedzi na tego typu pytania. Dylemat ten w oczywisty sposób stał się podstawą analiz kosztów i korzyści (*cost – benefit analysis*). Niestety sytuacje, z jakimi zwykle ma się do czynienia podczas badań ewaluacyjnych, rzadko kiedy poddają się prostemu rachunkowi ekonomicznemu”<sup>38</sup>.

Wydaje się jednak, że także oczekiwania, jakie ewaluatorzy mają wobec analizy kosztowo-korzyściowej, są nieco przesadzone w tym sensie, że jej wyniki bywają podważane jako nieobiektywne, gdyż sama analiza opiera się raczej na subiektywnych założeniach składających się na jej ekonomiczny model. Pamiętając o tym, że analiza kosztów i korzyści jest tylko jednym z narzędzi wykorzystywanych w badaniach ewaluacyjnych, musimy mieć

<sup>37</sup> W trybie wynikowym ewaluacja koncentruje się na rezultatach konkretnego działania. W trybie nakładowym ewaluacja kładzie nacisk na zasoby, umiejętności i ludzi, a w trybie procesowym przenosi się na metody stosowane do przekształcania nakładów w rezultaty; zob. Majone G., *Dowody, argumenty...*, op. cit., s. 253.

<sup>38</sup> Babbie E., *Badania społeczne w praktyce*, PWN, Warszawa 2004, s. 376.

świadomość ograniczonej wartości wyników. R. Formaini, w swojej książce o bardzo sugestywnym tytule *The Myth of Scientific Public Policy*, przekonuje, że nie ma czegoś takiego, jak „obiektywna” analiza kosztów i korzyści, a zatem wyniki, jakie dzięki niej uzyskujemy, nie są „obiektywne” w tym sensie, że nie dostarczają nam pewnej, możliwej do przyjęcia przez każdego, wiedzy potrzebnej do oceny realizowanego programu<sup>39</sup>. Zarzuty, jakie stawia Formaini, mają charakter ontologiczno-epistemologiczny, a dotyczą błędnych, jego zdaniem, założeń neoklasycznej ekonomii co do natury porządku społecznego, na których oparta jest analiza kosztów i korzyści. „Analiza kosztów i korzyści i neoklasyczna ekonomia są symbiotycznie powiązane ze sobą przez fakt, że analiza kosztów i korzyści nie mogłaby być dokonana bez założeń dostarczonych przez neoklasyczne metody. Kiedy analitycy mierzą, definiują i agregują «koszty» i «korzyści», robią to w ramach neoklasycznej struktury poczynionych założeń. Jest zatem uprawnione, aby argumentować, że analiza kosztów i korzyści jest «dzieckiem», a neoklasyczna teoria ekonomiczna «ojcem». (...) Koniecznym zatem jest, aby zrozumieć analizę kosztów i korzyści, trzeba najpierw poznać podstawowe założenia neoklasycznej ekonomii”<sup>40</sup>.

Dyskusja o ontologicznych podstawach modeli analizy, przeciwstawianiu modelowi „człowieka ekonomicznego” modelu „człowieka socjologicznego”, i uzyskiwanych w ich ramach wynikach stanowi kanon naukowych dociekań. Nie ma powodu, aby w tym miejscu tę dyskusję rekonstruować. Wystarczy przypomnieć tylko, że to, co dla jednych uchodzi za zobiektywizowany w ramach przyjętych teoretycznych założeń rezultat, w oparciu o który można podejmować racjonalne decyzje, dla innych jest niczym więcej niż tylko projekcją pewnych zakładanych motywów czy postaw, które nie mogą być podstawą do czynienia uogólnionych wyjaśnień w sferze teorii czy podejmowanych działań w sferze praktyki. Zarzut, jaki czyni Formaini tzw. unaukowanej *Public Policy*, jest niczym innym jak próbą obnażenia niedostatków ekonomicznego modelu „człowieka ekonomicznego” czy „kalkulującego” (*homo economicus*). Można by go uznać za zasadny także dla badań ewaluacyjnych, których orędownicy często przekonują, że o ile przeprowadzone są one wedle starannie przygotowanej metodologii, to ich wyniki stanowią podstawę do podejmowania racjonalnych decyzji. Konkluzja Formainiego brzmi, że „oparta na naukowych podstawach polityka publiczna (*Public Policy*), to marzenie, które wzrastało coraz bardziej od czasów Oświecenia i być może osiągnęło swoje apogeum pod koniec tego wieku, jest mitem, teoretyczną iluzją. Ona istnieje w naszych umysłach, analizach i metodach tylko dlatego, że szukamy by ją odkryć, a zmierzamy do odkrycia tego, czego szukamy”<sup>41</sup>.

<sup>39</sup> Formaini R., *The Myth of Scientific Public Policy*, Transaction Publishers, 1990, s. 39–65.

<sup>40</sup> Ibidem, s. 39.

<sup>41</sup> Ibidem, s. 1.

## Zakończenie

Wyciągnięcie wniosku, że wszelkie ograniczenia, na jakie narażona jest ewaluacja, czyni jej wyniki mało przydatnymi w procesach podejmowania decyzji, w szczególności dotyczących racjonalizowania wydatków publicznych, wydaje się być tak samo nieuprawnione, jak twierdzenie, że wyniki starannie przeprowadzonej ewaluacji z zachowaniem wszelkich prawideł sztuki są pewnym oparciem dla racjonalizowania decyzji w sferze publicznej. Powiedzenie jednak, że „prawda leży gdzieś pośrodku” jest obarczone takim samym prawdopodobieństwem błędu, jak wskazanie na którąś ze skrajnych opinii. Ponadto spór nie powinien przebiegać wedle logiki „kto ma rację” i wedle jakich argumentów tę rację jesteśmy skłonni przyznać jednej ze stron sporu. Raczej powinniśmy poszukiwać konsensusu tak, aby wyniki badań ewaluacji, nawet jeśli zgodzimy się co do ich rozmaitych niedoskonałości, stanowiły określoną wartość dla administracji publicznej odpowiedzialnej za podejmowanie takich czy innych decyzji w sektorze publicznym.

Dla „decyzjonistów”, by posłużyć się terminem używanym przez G. Majone, problem w zasadzie nie istnieje, gdyż racjonalne decyzje powinny być podejmowane w oparciu o informacje, jakie przynosi starannie przeprowadzona analiza kosztów i korzyści. Ponieważ jednak oponenti wskazują na wiele niedoskonałości takiego podejścia, głównie o charakterze ontologiczno-epistemologicznym, to pojawia się pytanie, co zrobić z ogromną ilością „wyprodukowanych” informacji w ramach prowadzonych ewaluacji programów interwencji społecznych. Gdyby bez jakichkolwiek zastrzeżeń uznać rację „antydecyzjonistów” kwestionujących przydatność analiz ewaluacyjnych dla procesu racjonalizowania decyzji publicznych, to musielibyśmy w jakiejś mierze przyznać, że dążność do alokacyjnej racjonalności w oparciu o ewaluację wyników przeraża się w swoje przeciwieństwo w tym sensie, że chcąc racjonalnie wydawać publiczne pieniądze, nie uzyskujemy żadnej pewnej wiedzy, za którą musimy jeszcze zapłacić.

Propozycja krytyków badań ewaluacyjnych, prowadzonych wedle założeń modelu ekonomicznego, jest taka, aby włączyć wysiłki ewaluatorów w szerszą demokratyczną debatę poświęconą ustalaniu kryteriów i standardów realizacji programów publicznych i ich oceny. Jedną ze stron takiej debaty powinna być administracja publiczna, która nie bezkrytycznie wykorzystywałaby dostarczane jej informacje pochodzące z przedsięwzięć ewaluacyjnych. Innymi słowy, wartość ewaluacji nie polegałaby na tym, że można by na jej podstawie podjąć racjonalną decyzję maksymalizującą korzyści przy określonych kosztach czy też minimalizującą koszty przy określonych korzyściach, ale na tym, że stanowiłaby ona wartościowy wkład do dyskusji między stronami reprezentującymi różne punkty widzenia czy interesy. „Obecnie – pisze Majone – konieczne jest nie tyle opracowanie «obiektywnych» wskaźników wyników

– tradycyjny cel analiz ewaluacyjnych – ile ułatwienie szerokiego dialogu wśród zwolenników różnych kryteriów<sup>42</sup>.

Gdyby udało się wyniki ewaluacji wykorzystać w szerszej debacie o zasadności prowadzenia interwencji społecznych, ich kosztach i korzyściach oraz kryteriach, wedle których oceniane byłyby rezultaty interwencji, to z pewnością moglibyśmy powiedzieć, że ewaluacja może wносить swój wkład w demokratyczny proces uzgadniania rozmaitych racji czy punktów widzenia. We współczesnych demokracjach, cierpiących na deficyt legitymizacji, taka funkcja ewaluacji wydaje się być bardzo atrakcyjna przynajmniej z dwóch powodów. Pierwszy jest taki, że prowokując debatę, w jakiejś mierze zwiększa się legitymizację procesu decyzyjnego w administracji publicznej, drugi zaś taki, że debata ta, oparta jednak na rezultatach uzyskanych zgodnie z naukowymi standardami badań empirycznych, wnosi określony potencjał wiedzy, który może przyczynić się do bardziej wysublimowanego, jak mówi Majone, postrzegania procesu kształtowania polityki publicznej. Z pewnością nie oznacza to, że wyniki badań ewaluacyjnych, z których może i powinna korzystać administracja publiczna, uwolnią nas od rozmaitych manipulacji zasobem informacyjnym, gdyż sama nauka nie jest wolna od tego typu pokus, ale to jest temat na zupełnie inną dyskusję.

Ważne jednak jest, aby możliwości i ograniczenia wykorzystania wyników badań ewaluacyjnych były szeroko dyskutowane wśród przedstawicieli administracji publicznej, gdyż to od niej zależy jakość realizowanych programów społecznej interwencji. Gdyby miało się okazać, że decydenci na poziomie systemu administracyjnego przeceniają możliwości ewaluacji, to wyniki, które dzięki niej uzyskujemy, niekoniecznie musiałyby odzwierciedlać rzeczywistą wartość realizowanych programów. Ta ostatnia, jak mówią „antydecyzyjniści”, nie zawsze jest możliwa do policzenia. Sceptycyzm wobec ewaluacji nie powinien jednak przesłaniać jej możliwości, gdyż wiele programów społecznej interwencji z powodzeniem można poddać procesom ewaluacji, dostarczając tym samym wielu cennych informacji o rozmaitych aspektach realizowanych programów. Dyskusja wokół ewaluacji powinna toczyć się zatem zarówno wśród przedstawicieli środowiska akademickiego w ramach międzydyscyplinarnej czy paradygmatycznej wymiany opinii, jak i między nimi a przedstawicielami administracji. Oni z racji swojego doświadczenia mogą znacznie więcej wnieść do procesów doskonalenia badań ewaluacyjnych, niż to wydaje się z perspektywy akademickiej katedry. Cóż to oznacza dla procesu podejmowania decyzji przez administrację? Z pewnością ewaluacja nie jest w stanie zastąpić zdrowego rozsądku i doświadczenia, które generowane jest w toku codziennej praktyki administracyjnej, a do którego często odwołują się inkrementaliści. Musimy jednak pamiętać, że postępowanie wedle określonych schematów grozi petryfikacją struktur decyzyjnych, a wyniki ewaluacji mogą

<sup>42</sup> Majone G., op. cit., s. 264.

nas przed takim niebezpieczeństwem uchronić. Arystotelesowska pochwała umiaru wydaje się być rozsądną propozycją, z której powinna korzystać także administracja publiczna.

---

**Dariusz Zalewski** – doktor, adiunkt w Instytucie Socjologii WFiS Uniwersytetu Warszawskiego (Zakład Socjologii Problemów Społecznych i Planowania Społecznego) oraz w Instytucie Pracy i Spraw Socjalnych (Zakład Zbiorowych Stosunków Pracy). Od lat zajmuje się problematyką instytucji państwa opiekuńczego, administracji publicznej i dialogu społecznego. Jest autorem i współautorem m.in. takich publikacji, jak: *Opieka i pomoc społeczna. Dynamika instytucji, Utopie inkluzji. Sukcesy i porażki programów reintegracji społecznej, Dialog społeczny na poziomie regionalnym. Ocena szans rozwoju*. Wielokrotnie uczestniczył w projektach badawczych poświęconych m.in.: formom reprezentacji interesów grupowych, sposobom realizacji zadań przez administrację publiczną, dialogowi społecznemu na poziomie centralnym i regionalnym. W ramach swoich zawodowych zainteresowań brał udział także w badaniach ewaluacyjnych, np. w ewaluacji programu wymiany młodzieży „Leonardo da Vinci” współfinansowanego przez Unię Europejską.

## Bibliografia

- Evaluating UE Expenditure Programmes: A Guide*, European Commission, 01.1997.
- Babbie E., *Badania społeczne w praktyce*, PWN, Warszawa 2004.
- Dunleavy P., *Democracy, Bureaucracy and Public Choice*, Harvester Wheatsheaf, 1991.
- Formaini R., *The Myth of Scientific Public Policy*, Transaction Publishers, 1990.
- Friedman M., Friedman R., *Wolny wybór*, Wyd. Panta, Sosnowiec 1994.
- Frieske K.W., *Socjologia w działaniu. Nadzieje i rozczarowania*, Wyd. Uniwersytetu Warszawskiego, Warszawa 1990.
- Haber A. (red.), *Ewaluacja ex-post. Teoria i praktyka*, PARP, Warszawa 2007.
- Kickert W.J.M., *Public Management and Administrative Reform in Western Europe*, Edward Elgar, Cheltenham 1997.
- Lane J.E. (red.), *Public Sector Reform. Rationale, Trends and Problems*, Sage Publications, London 1997.
- Majone G., *Dowody, argumenty i perswazja w procesie politycznym*, Wyd. Naukowe Scholar, Warszawa 2004.
- Majone G., *The rise of the regulatory state in Europe*, West European Politics, 17.
- McKinney J.B., Howard L.C., *Public Administration. Balancing Power and Accountability*, Praeger Publishers, Westport 1998.
- Osborne D., Gaebler T., *Rządzić inaczej. Jak duch przedsiębiorczości przenika i przekształca administrację publiczną*, Media Rodzina, Poznań 1992.
- Verheijen T., Coombes D., *Innovations in Public Management. Perspectives from East and West Europe*, Edward Elgar, Cheltenham 1998.



Stanisław Mazur

## Zarządzanie wiedzą w polskiej administracji publicznej

Zarządzanie wiedzą (ZW)<sup>1</sup> to podejście konceptualno-aplikacyjne, w którym generowanie i kapitalizowanie wiedzy stanowi podstawową kategorię doskonalenia organizacji i systemów organizacyjnych. Jego istotą jest gromadzenie danych i informacji, ich selekcja i konwersja na wiedzę zakorzenioną w kulturze organizacji. Tak wytworzona wiedza staje się podstawowym zasobem organizacji warunkującym jej zdolność do osiągnięcia ustanowionych dla niej celów.

Efektywne i skuteczne zarządzanie wiedzą w instytucjach sektora publicznego wymaga odpowiedniej strukturyzacji i integracji czyli systemu ZW. System ZW w polskiej administracji publicznej znajduje się *in statu nascendi*. Jego zasadnicza konstrukcja instytucjonalna i elementy konstytutywne nie zostały jeszcze określone. W systemie tym występuje nadmiar danych, niepełny zasób informacji i ograniczony potencjał wiedzy, która stanowiłaby podstawę racjonalnych ekonomicznie i społecznie decyzji podejmowanych przez państwo i jego struktury. Podstawowym mankamentem jest jego niski poziom zdolności do wykorzystywania wiedzy dla potrzeb procesów decyzyjnych.

Dla budowania systemowego rozwiązania służącego wzrostowi wiedzyochłonności administracji publicznej kluczowe jest stworzenie systemu infrastruktury ZW. Jej celem byłoby wzmocnienie w administracji publicznej kapitałów, które stanowią podstawę systemów otwartych na wiedzę, systemów refleksyjnych i uczących się, tj. kapitału: ludzkiego<sup>2</sup>, kulturowego<sup>3</sup> oraz tech-

---

<sup>1</sup> Posługiwanie się pojęciem „zarządzania wiedzą” może stanowić przykład swoistego imperializmu racjonalnego rozumu. Bardziej adekwatne jest raczej używanie pojęcia zarządzania warunkami dla rozwoju i wykorzystywania wiedzy. Zważywszy jednak na dominującą i ugruntowaną konwencję semantyczną, autor skłonny jest posługiwać się terminem „zarządzania wiedzą” w znaczeniu – kreowanie warunków do jej rozwoju, a nie zarządzanie nią w myśl klasycznych zasad nauki o zarządzaniu.

<sup>2</sup> **Kapitał ludzki** – informacja, wiedza, kreatywność, umiejętność i zdolność ich efektywnego wykorzystania, przywództwo polityczne itp.

<sup>3</sup> **Kapitał kulturowy** (organizacyjny) – organizacje otwarte, uczące się, refleksyjne elastyczne, zadaniowo zorientowane, premiujące współpracę i dzielenie się wiedzą.



nologicznego<sup>4</sup>. Realizacja tego postulatu służyłaby tworzeniu podstawowego zasobu, który państwo może oferować swoim obywatelom. Zasobem tym jest dobre rządzenie. Tworzenie tego zasobu jest procesem złożonym i długotrwałym, związanym z koniecznością zmiany kultury rządzenia, w tym kultury administracyjnej. Procesy zmian kulturowych, w szczególności w instytucjach publicznych, cechują się niską dynamiką wynikającą z ich konserwatyzmu, skłonności oraz zdolności do obrony *status quo*.

Celem niniejszego wywodu jest zarysowanie odpowiedzi na pytania: a) o istotę ZW w administracji publicznej? b) o jakość ZW w polskiej administracji publicznej? c) o to, co należałoby uczynić, aby polska administracja publiczna kapitalizowała w większym zakresie, niż ma to miejsce obecnie, korzyści wynikające z praktykowania formuły ZW.

Dlaczego w publikacji poświęconej ewaluacji zajmujemy się kwestiami zarządzania wiedzą w tak szerokim kontekście? ZW to paradygmat, w którym refleksja nad decyzjami, sposobami ich realizacji oraz ich konsekwencjami stanowi zagadnienie centralne. Takie jego ujęcie sytuuje go blisko ewaluacji i jej konstytutywnych funkcji. Co zatem stanowi jego wyróżnik względem ewaluacji? Odpowiadając na tak postawione pytanie, należy podkreślić, że zarządzanie wiedzą jest konceptem szerszym względem ewaluacji. Ewaluacja przede wszystkim bada zależności przyczynowo-skutkowe oraz ocenia konsekwencje ich wystąpienia (zarówno zamierzone, jak i niezamierzone). Tworzy korpus specjalistycznej wiedzy niezbędnej do programowania, implementacji i oceny polityk i programów publicznych. Rolą systemu ZW jest aktywizacja tej wiedzy poprzez jej standaryzację i upowszechnienie w systemie instytucji publicznych. Relacje między systemem ZW a ewaluacją mają charakter integralny – ten pierwszy tworzy architekturę dla przepływu strumieni wiedzy, a ewaluacja jest jednym ze sposobów generowania treści, które strumienie te tworzą. To też uzasadnia ujmowanie zagadnienia ZW w rozważaniach nad ewaluacją.

## Zarządzanie wiedzą – zagadnienia ogólne

Pojęcie *zarządzania wiedzą* (ZW) jest terminem równie nośnym co nie-dookreślonym. Nadaje mu się wiele znaczeń, interpretuje się go na wiele sposobów. Wspólnym rdzeniem semantycznym jego różnorodnych perspektyw definicyjnych jest wiedza i jej zastosowanie. Ekspozycja znaczenia wiedzy nie oznacza, że wcześniejsze paradygmaty zarządzania nie odwoływały się do niej jako do istotnego zasobu rozwojowego. Jednakże to

---

<sup>4</sup> **Kapitał technologiczny** – nowoczesne instrumenty realizacji ustrojowych powinności państwa, np. nowoczesne systemy analizy i programowania, systemy zarządzania, świadczenia usług, e-administracja.

w nurcie refleksji nad ZW jej generowanie i kapitalizowanie uznane zostały za kategorie o znaczeniu centralnym. W wiedzy upatruje się podstawowej determinanty doskonalenia organizacji i systemów organizacyjnych, wzrostu ich innowacyjności, wzmacniania ich pozycji konkurencyjnej oraz potencjału adaptacyjnego.

Rozważania o ZW stanowią centralny przedmiot debaty nad sposobami doskonalenia organizacji i ich systemów. Koncept ten to swoistego rodzaju przykład wielkiej narracji aspirującej do roli holistycznej perspektywy wyjaśniania istoty zmian – zarówno w wymiarze organizacyjnym, jak i społecznym. Jego siła oddziaływania przysparza mu wielu zwolenników, czyniąc go jednym z dominujących podejść konceptualnych wyjaśniania logiki jakościowej zmiany działania organizacji i ich systemów.

### Podejścia definicyjne

Badacze związani z Organizacją Współpracy Gospodarczej i Rozwoju (*Organization for Economic Co-operation and Development* – OECD) pojęcie ZW definiują, z perspektywy teleologicznej, jako: „(...) szeroki zbiór praktyk organizacyjnych luźno powiązanych z generowaniem, pozyskiwaniem, upowszechnianiem *know-how* i promowaniem dzielenia się wiedzą w organizacji i w jej zewnętrznym środowisku, obejmujących:

- rozwiązania organizacyjne (decentralizacja władzy, uelastycznienie biurokratycznych struktur, wykorzystanie technologii informacyjnych i komunikacyjnych),
- rozwój zasobów ludzkich (doradztwo, szkolenia, mobilność itp.) i umiejętności zarządczych,
- transfer kompetencji (bazy danych dotyczących personelu, opisy dobrych praktyk itp.),
- promowanie zarządzania i bodźców zachęcających do dzielenia się wiedzą (ocena jakości pracy i awans warunkowane skłonnością do dzielenia się wiedzą, ocena roli menedżerów)<sup>5</sup>.

Równie powszechnym sposobem definiowania ZW jest jego ujmowanie procesowe: „(...) zarządzanie wiedzą składa się z sześciu kluczowych procesów: lokalizowania, pozyskiwania, rozwijania, dzielenia się wiedzą i jej rozpowszechniania, wykorzystywania oraz zachowywania wiedzy”<sup>6</sup>.

Istotą ZW w instytucjach publicznych jest zatem gromadzenie danych i informacji, ich selekcja i konwersja na wiedzę organizacyjną oraz oparcie

<sup>5</sup> *The Learning Government: Introduction and Draft Results of the Survey of Knowledge Management Practices in Ministries/Departments/Agencies of Central Government* – OECD, 2003, s. 7.

<sup>6</sup> Probst G., Raub S., Rombardt K., *Zarządzanie wiedzą w organizacji*, Oficyna Ekonomiczna, Kraków 2002, s. 48.

procesu podejmowania decyzji na zasobach wiedzy instytucjonalnej (skodyfikowanej i nieskodyfikowanej) zakorzenionej w kulturze organizacji.

### **Infrastruktura zarządzania wiedzą**

Efektywne i skuteczne zarządzanie wiedzą w instytucjach sektora publicznego, podobnie zresztą jak w sektorze podmiotów gospodarczych czy też społecznych, wymaga odpowiedniej strukturyzacji i integracji wielu elementów składowych konstytuujących infrastrukturę systemu ZW. Składają się na nią:

1. Ugruntowana wśród politycznych zwierzchników administracji publicznej świadomość znaczenia wiedzy dla dobrego rządzenia, skutkująca określonymi decyzjami sprzyjającymi tworzeniu kultury ZW (np. przyjmowanie stosowanych strategii, alokowanie środków na ich realizację, wprowadzanie odpowiednich regulacji, w tym dotyczących polityki kadrowej, tworzenie konkurencyjnych warunków pracy i wynagradzania).
2. Ewolucyjna zmiana kultury organizacyjnej (tworzenie klimatu zaufania i współdziałania, kreowanie liderów promujących postawy innowacyjne, przechodzenie z proceduralnego, formalistycznego sposobu działania administracji w kierunku działania zorientowanego na osiągnięcie celów, zmiana wzorców współpracy administracji z partnerami społecznymi i gospodarczymi – odejście od imperatywności na rzecz interaktywności, odejście od myślenia wertykalnego (resortowego) na rzecz myślenia horyzontalnego (holistycznego) w procesach tworzenia i implementacji polityk oraz programów publicznych).
3. Doskonalenie kompetencji zawodowych urzędników (wzmacnianie systemów szkoleń i doskonalenia zawodowego, wprowadzanie bodźców skłaniających do podnoszenia kompetencji zawodowych, konkurencyjne systemy rekrutacji i selekcji pracowników, ich oceny oraz planowania ich rozwoju zawodowego itp.).
4. Doskonalenie metod zarządzania opartych o intensywne wykorzystanie wiedzy (np. zarządzanie jakością, zarządzanie przez cele, doskonalenie metodologii analizy, programowania oraz oceny polityk i programów publicznych, w tym wykorzystanie metod ekonometrycznych, tworzenie sieci wiedzy opartych o współdziałanie z ośrodkami akademickimi i doradczymi).
5. Nasycenie administracji publicznej nowoczesnymi technologiami informatycznymi i komunikacyjnymi (np. e-administracja, portale, Intranet, informatyczne systemy przepływu informacji i dokumentów, bazy danych, informatyczne narzędzia wspierania procesów decyzyjnych).
6. Stosowanie elastycznych i zadaniowych form organizacyjnych wyprowadzonych z koncepcji organizacji uczącej się (np. stosowanie płaskich struk-

tur organizacyjnych, wprowadzanie formuły zespołów zadaniowych oraz efektywnych systemów informacyjno-komunikacyjnych zorientowanych na szybki przepływ informacji i wiedzy).

Na rysunku 1 przedstawiono, w postaci diagramu, elementy składowe infrastruktury ZW.



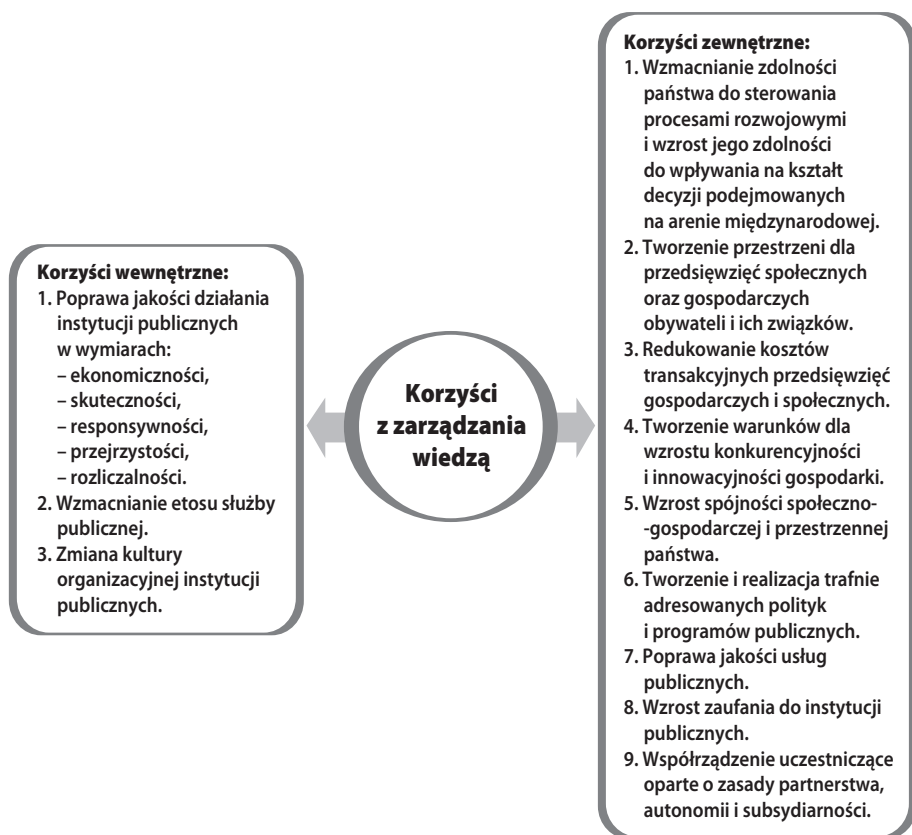
Rysunek 1. Infrastruktura systemu zarządzania wiedzą

Źródło: Opracowanie własne.

Reasumując, infrastruktura systemu zarządzania wiedzą to wielowymiarowa konstrukcja, na którą składają się: a) instytucjonalna architektura o charakterze sieciowym, b) aktorzy poruszający się w przestrzeni wytyczonej przez ową architekturę, c) zestaw bodźców służących otwarciu systemu instytucjonalnego oraz jego aktorów na internalizację i pełniejsze wykorzystanie wiedzy, d) otwartość systemu oznaczająca zarówno zdolność do wchłaniania bodźców zmiany (izomorfizm refleksyjny), jak i umiejętność zachowania dynamicznej równowagi (relacja między zewnętrznymi bodźcami, adaptacją środowiskową a wewnętrzną integralnością systemu), e) mechanizm zwrotnej refleksyjności rozumiany jako zdolność do krytycznej refleksji nad jakością podejmowanych działań oraz zdolność do rekonfiguracji zasobów i wzorców wiedzy w celu lepszego ich wykorzystania.

## Korzyści wynikające z ZW

Zapożyczając z teorii modernizacji pojęcia centrum, peryferii oraz półperyferii, przez analogię można stwierdzić, że kraje, których administracje nie adaptują się do reżimu wiedzy – nie będą w stanie oprzeć na niej procesów programowania polityk publicznych i ich implementacji – tracić będą pozycję w centrum układów i spychane będą w kierunku półperyferii i peryferii. Jeśli układy instytucjonalne państwa (w tym administracja publiczna) oraz reguły i mechanizmy ich działania pozostawac będą w opozycji do logiki reżimu wiedzy, wzrastać będzie ich bezsilność w zakresie wpływania na bieg wydarzeń międzynarodowych, a ich zdolność do sterowania wewnętrznymi procesami społecznym w coraz większym stopniu będzie iluzoryczna. Rysunek 2 przedstawia przykładowe korzyści wynikające z zarządzania wiedzą w instytucjach publicznych. Ich prezentacji dokonano w podziale na dwie kategorie, tj. korzyści wewnętrzne i zewnętrzne.



Rysunek 2. Korzyści z ZW w instytucjach publicznych

Źródło: Opracowanie własne.

## Percepcja zagadnienia ZW w instytucjach publicznych

Interesujące spostrzeżenia na temat zjawiska ZW i percepcji tego zagadnienia zawiera raport z badań przeprowadzonych przez OECD<sup>7</sup>. Z informacji tam zawartych wynika, że szybko wzrasta liczba państw, które uznają ZW za obszar o znaczeniu priorytetowym dla poprawy jakości rządzenia. Zdecydowana większość respondentów (75%) postrzega ZW jako podstawowy sposób na wzrost efektywności i skuteczności działania kierowanych przez nich instytucji. Wymiernym tego przykładem jest wzrost środków na szkolenia pracowników, większe nasycenie instytucji publicznych nowoczesnymi technologiami informatyczno-komunikacyjnymi, jak również delegowanie zadań związanych z koordynacją działań z zakresu ZW na osoby zajmujące kluczowe miejsce w hierarchii organizacyjnej.

Jako zalety wprowadzania systemu ZW wymienia się między innymi: poprawę jakości funkcjonowania organizacji, obniżenie kosztów funkcjonowania, wzrost przejrzystości działań, poprawę jakości świadczonych usług oraz wzrost zdolności adaptacyjnych. Ustalenia w raporcie wskazywać mogą, że ZW to nie tylko przejściowa moda na kopiowanie nośnych medialnie koncepcji wytworzonych w sektorze rynkowym, a raczej presja otoczenia na szeroko pojęty sektor publiczny, będąca wynikiem swoistej racjonalności systemowej gospodarki opartej na wiedzy.

Wyniki przywoływanego badania nie oznaczają bynajmniej, iż jakość ZW osiągnęła w instytucjach publicznych państw OECD stan zadowalający. Wskazuje się, bowiem na potrzebę:

- dalszej decentralizacji zarządzania,
- usprawniania poziomego przepływu informacji wewnątrz organizacji publicznych i między nimi,
- podniesienia jakości szkoleń,
- wzmacniania bodźców skutkujących dzieleniem się wiedzą,
- doskonalenia technologii informatyczno-komunikacyjnych.

## Zarządzanie wiedzą w kontekście zmian zachodzących w otoczeniu państwa i jego administracji

Administracja publiczna staje się ważnym elementem działań na rzecz szeroko rozumianego rozwoju społeczno-gospodarczego. Ze stwierdzeniem tym zgadza się coraz większa liczba ekonomistów, socjologów i politologów. W tym również, to pewien paradoks, ci, którzy jeszcze nie tak dawno wyrażali opinie odmienne.

<sup>7</sup> *The Learning Government...*, op. cit.

Coraz częściej przyjmuje się, że bez dobrej administracji nie ma sprawnego państwa, a bez dobrze działającego państwa rozwój społeczno-gospodarczy cechuje się słabą dynamiką, niepewnością oraz brakiem spójności w jej szerokim rozumieniu. Dobrze działająca administracja publiczna staje się jednym z *condicio sine qua non* dla rozwoju państwa i zapewnienia odpowiedniej jakości życia jego obywateli. Można zatem uznać, że państwom, które mają sprawną administrację, będzie się powodziło lepiej niż tym, które jej nie posiadają. Kontekstem systemowym dla wzmacniania ZW w instytucjach publicznych jest pojawienie się gospodarki opartej na wiedzy. W tym nowym paradygmacie podstawowym zasobem, który instytucje publiczne mogą wnieść, jest dobre rządzenie pojmowane jako dobro publiczne. Nie jest ono możliwe bez nasycenia tych instytucji praktykami zarządzania wyprowadzonymi z konceptu ZW.

Wzrost znaczenia administracji i charakter oczekiwań względem niej formułowanych zakorzenione są jednak w rzeczywistości różnej od tej, w której ojcowie nowożytnej administracji wytyczali zasady jej działania. Istota państwa i jego administracji ulega redefinicji. Zmieniają się jego relacje ze społeczeństwem. Bez zrozumienia roli władzy publicznej oraz jej miejsca w różnicującym się i nabierającym cech autopojetyczności<sup>8</sup> systemie społecznym, projektowanie programów doskonalenia administracji i ich realizacja może charakteryzować się dysfunkcjonalnością względem założonych celów.

Wzrost policentryczności porządku społecznego i politycznego<sup>9</sup>, narastające zróżnicowanie funkcjonalne poszczególnych segmentów społecznych oraz erozja władzy hierarchicznej, prowadzą do deficytu zaufania względem władzy publicznej i problemu jej legitymizacji. Zjawiska te pogłębiają przeciążenie państwa, kryzys polityki regulacyjnej oraz problem efektywnej, skutecznej i społecznie aprobowanej alokacji dóbr publicznych. Narastająca konkurencja i globalizacja to kolejne czynniki, które w odmienny sposób – niż to miało miejsce dotychczas – wytyczają relacje państwa, administracji i obywateli. Rządzenie za pomocą klasycznych zasad i instrumentów staje się coraz bardziej problematyczne w społeczeństwie zmiany, niepewności i ryzyka. Tworzą się nowe kody i wzorce zachowań społecznych. Wzrasta zapotrzebowanie na dobre rządzenie, którego konstytutywnym elementem jest dobra administracja, a jej warunkiem jest zdolność do kapitalizowania wiedzy dla realizacji interesu publicznego. Państwo i administracja w coraz większej

<sup>8</sup> Pojęcie autopojetyczności rozumiane jest w tym opracowaniu jako proces wewnętrzny różnicowania się elementów systemu społecznego w ujęciu opisywanym przez N. Luhmanna, *The Differentiation of Society*, Columbia University Press, New York 1982.

<sup>9</sup> Poprzez pojęcie *policentryczności porządku społecznego* rozumie się erozję instytucji tradycyjnie uważanych za dominujące ośrodki władzy i wpływu (np. państwo, Kościół) i pojawianie się wielu nowych ośrodków władzy, jak również spadek znaczenia hierarchicznych metod koordynacji działań zbiorowych.

mierze funkcjonują w reżimie wiedzy, w którym zasobem najcenniejszym jest zdolność do refleksyjnego jej wykorzystania i przełożenia na procedury operacyjne organizujące ich aktywności. Nie oznacza to, że w poprzednich reżimach wiedza nie była wykorzystywana, a jedynie, że nigdy wcześniej jej zastosowanie nie miało tak fundamentalnego znaczenia dla jakości działania struktur państwa.



Rysunek 3. Ekosystem działania administracji publicznej

Źródło: Opracowanie własne.

W kontekście nowej strategicznej roli państwa warto zwrócić uwagę na środowisko, w którym działa administracja publiczna. Funkcjonuje ona w złożonym ekosystemie instytucji, aktorów, celów, zasad, interesów, układów władzy, wartości i przekonań. Od administracji oczekuje się:

- lojalności wobec rządzących polityków, a zarazem politycznej neutralności i przestrzegania reguł demokracji reprezentatywnej,
- efektywności i sprawności w rozwiązywaniu problemów społecznych, a zarazem odpowiedzialności względem politycznych przełożonych i respektowania oczekiwań społecznych,
- kierowania się proceduralną logiką wynikającą z porządku prawnego, a zarazem niepopadania w nadmierny formalizm,
- wykorzystywania profesjonalnej, specjalistycznej wiedzy bez przyjmowania kultury technokratycznej,
- kierowania się zasadami skuteczności i ekonomicznej efektywności, a zarazem dostosowania się do oczekiwań społecznych, stosowania za-



sad rynkowych oraz działania na rzecz ograniczania wykluczenia społecznego,

- uwzględniania społecznych preferencji i oczekiwań, a zarazem niepoddawania się ich presji i oddziaływaniu grup interesu<sup>10</sup>.

Przywołane tu oczekiwania, w pewnej mierze wykluczające się, to dodatkowy element, który z jednej strony wymusza stosowanie coraz bardziej złożonych mechanizmów zarządzania, opartych o zalecania wypływające z modelu ZW, a jednocześnie zarządzanie to utrudnia.

## Zarządzanie wiedzą w polskiej administracji publicznej – diagnoza

Sprawnie działający system ZW powinien umożliwiać gromadzenie odpowiednich danych i informacji oraz tworzenie i wykorzystanie wiedzy w praktyce. Dla potrzeb dalszego opisu dokonuję arbitralnego wyróżnienia trzech typów wiedzy: a) **strategicznej**<sup>11</sup>, b) **taktycznej**<sup>12</sup>, c) **operacyjnej**<sup>13</sup>. Zakładam, że każdy z nich przynależy do dających się wyodrębnić domen aktywności państwa i jego administracji. Przywołana dystynkcja nie ma charakteru w pełni rozłącznego, ponieważ w praktyce działania administracji publicznej i wyspecyfikowane powyżej kategorie wiedzy są mocno skorelowane, a często ich granice nakładają się na siebie.

### Wiedza strategiczna

Kreowanie strategicznych warunków dla polityki rozwoju, kreślenie dobrych scenariuszy rozwoju państwa i ich realizacji to działanie lokujące się w sferze ścisłych interakcji pomiędzy urzędnikami i ich politycznymi przełożonymi. Powodzenie tych działań wymaga integrowania komplementarnych zasobów informacji i wiedzy przynależnych zarówno politykom, jak i urzędnikom oraz ekspertom. Twórcze kombinacje i rekombinacje tych zasobów powinny zachodzić w układach otwartych, sieciowych, w procesie wielu interakcji. W Polsce proces ten przebiega w sposób nieodgadniony i jego efekty często

<sup>10</sup> Olsen J.P., „Towards a European Administrative Space?”, *ARENA Working Papers WP 02/26*.

<sup>11</sup> **Wiedza strategiczna** rozumiana jest jako zasób konieczny do refleksyjnego tworzenia meta-reguł programowania zasad i projektowania instrumentów sterowania polityką rozwoju państwa.

<sup>12</sup> **Wiedza taktyczna** rozumiana jest jako zasób znajdujący zastosowanie w tworzeniu polityki i programów publicznych służących realizacji celów polityki rozwoju państwa.

<sup>13</sup> **Wiedza operacyjna** to zasób umożliwiający zarówno efektywną, skuteczną i responsywną realizację polityki i programów publicznych, jak i świadczenie klasycznych usług publicznych.

uznaje się za wątpliwe. Problem deficytu wiedzy, a *de facto* jej wykorzystania, przejawia się na wiele wymownych sposobów.

W działaniach, w przypadku których wiedza strategiczna ma szczególne zastosowanie, wyraźnie dostrzegalny jest brak architektury instytucjonalnej dla jej tworzenia, przepływu i kapitalizowania. To oznacza, że nie ma strategiczno-koordynacyjnego ośrodka odpowiedzialnego za wytyczanie celów polityki rozwoju państwa i koordynację działań prowadzonych w jej ramach. Ta kluczowa kompetencja dzielona jest między resorty, kancelarię prezesa rady ministrów, urzędy centralne i agencje oraz zespoły doradców, w sposób oczywisty prowadząc do niespójności programowej podejmowanych działań i ich zakotwiczeniu w rzeczywistości wertykalnie zorientowanych resortów. Nasuwają się co najmniej dwa pytania. Po pierwsze, gdzie znajduje się w strukturze egzekutywy ośrodek wytyczania strategicznych dla państwa planów rozwojowych? Po drugie, jaki segment egzekutywy i w jaki sposób jest rozliczany za jakość jej implementacji i koordynacji? Odpowiedź wskazująca, że to rząd kolegialnie, a premier personalnie ponosi odpowiedzialność za jakość rządzenia, jest dalece niewystarczająca, w kontekście spełniania standardów dobrego rządzenia, w tym zasady rozliczalności.

Innego rodzaju kwestią jest brak odpowiedzialnego przywództwa politycznego. Jest to szczególnie dotkliwe, kiedy konieczne jest podjęcie wysiłku służącego rozwiązaniu problemu o charakterze strukturalnym. Zwłaszcza, jeżeli jest to obciążone wysokim stopniem niepewności, ryzyka i potencjalnie wysokimi kosztami politycznymi w przypadku niepowodzenia. Chodzi zatem o reformy obiektywnie uzasadnione, ale wyniki kalkulacji politycznych wskazują, że prawdopodobieństwo ich niepowodzenia i koszty polityczne mogą okazać się znaczące. Bez wiedzy, która pokazywałaby konsekwencje tych zaniechań, ich zrozumienie przychodzi politykom niezmiernie trudno.

Paralelny problem braku systemu zarządzania wiedzą strategiczną odnotować można w przypadku samorządu regionalnego. Brak odpowiedniego zaplecza merytorycznego, niewielkie wykorzystanie wiedzy analitycznej i reguł programowania strategicznego w celu projektowania rozstrzygnięć kluczowych dla regionu, to tylko niektóre z przykładów obrazujących istotę problemu.

## **Wiedza taktyczna**

Pomimo wielu pozytywnych zmian, jakie zaszły w ostatnich latach, potencjał administracji publicznej w zakresie kompetencji taktycznych jest niewystarczający. Za niezadowolającą uznać należy jakość procesów kreowania oraz implementacji programów publicznych. Szczególnie wiele problemów pojawia się w sferze procesów koordynacyjnych. Kolejnym deficytem – w zakresie kompetencji taktycznych administracji publicznej – jest niewystarza-

jąca jakość oceny skutków podejmowanych działań oraz niska jakość polityki regulacyjnej. Często działania podejmowane w tym zakresie cechują się małą użytecznością.

Poprawy wymagają również mechanizmy integrowania oceny programów publicznych podejmowanych przez administrację z procesami decyzyjnymi. Nierzadko działaniom tym nadaje się charakter rytualny, a ich wartość poznawcza i implementacyjna jest mocno ograniczona. Implikuje to konieczność doskonalenia systemu ewaluacji polityk i programów publicznych. Zdolność do ich oceny oraz refleksji nad ich racjonalnością ekonomiczną i społeczno-gospodarczą efektywnością to zagadnienie kluczowe dla dobrego rządzenia. To podstawa dla strategicznego namysłu i dobrego działania, sposób na kumulowanie wiedzy o zależnościach przyczynowo-skutkowych i budowanie pamięci instytucjonalnej organizacji. Tak pomyślana ewaluacja łączy dwie funkcje: poznawczą i rozliczeniową. Pierwsza z nich jest podstawą racjonalizacji polityk publicznych, druga natomiast służy do kontrolowania kwestii dotyczących gospodarowania środkami publicznymi i egzekwowania związanej z tym odpowiedzialności. Funkcje te są równie ważne i ze sobą zespolone. Tylko ten sposób ich postrzegania tworzy całościowy – a co za tym idzie – obiektywny obraz efektów interwencji publicznych.

Koniecznością staje się wkomponowanie ewaluacji w proces tworzenia i modyfikacji programów publicznych, doprowadzenie do tego, aby decydenci traktowali badania diagnostyczne i ewaluacyjne jako wartościowe oraz potrzebne źródło informacji o efektach działań przeszłych i prawdopodobnych konsekwencjach tych, które nastąpią. Wymaga to legitymizacji badań ewaluacyjnych przez ich wysoki standard metodologiczny, profesjonalizm oraz obiektywizm. Potrzebne jest także wyposażenie decydentów w instytucjonalną możliwość weryfikacji poziomu owych standardów, a także agregacji, analizy i syntezy wyników ewaluacji różnych projektów i programów oraz przekształcania ich w wiedzę użyteczną w procesie projektowania działań i podejmowania decyzji.

## **Wiedza operacyjna**

Zmiany wymagają mechanizmy zarządzania – bazujące na wiedzy operacyjnej – wykorzystywane w polskiej administracji publicznej. Konieczne jest rozszerzenie zakresu stosowania nowoczesnych systemów i narzędzi zarządzania zasobami ludzkimi i finansowymi (np. zintegrowane zarządzanie strategiczne, budżetowanie zadaniowe itp.). Niezadowalający jest stopień wykorzystania technologii informatycznych, zwłaszcza narzędzi związanych z modelem e-administracji. Istotnym mankamentem jest również brak standaryzacji usług publicznych (dotyczy to głównie administracji samorządowej) oraz systemowych rozwiązań w zakresie pomiaru usług publicznych.

Negatywny wpływ na sytuację ma również brak całościowych opracowań, cechujących się odpowiednimi standardami metodologicznymi, które stanowiłyby podstawę systemowego doskonalenia administracji i podnoszenia jakości rządzenia. Nie ma także danych, które w ujęciu porównawczym pozwalałyby lokować zmiany i tendencje zachodzące w polskiej administracji względem procesów modernizacyjnych zachodzących w innych krajach.

Nierzadko działania podejmowane przez administrację publiczną są zorientowane na wewnętrzne procedury, a sposób postrzegania współpracy z partnerami społecznymi i gospodarczymi ma charakter imperatywny. Wydaje się również, że niski jest stopień kapitalizowania korzyści wynikających z zastosowania mechanizmów współpracy zarówno z organizacjami pozarządowymi, jak i z ośrodkami badawczymi i naukowymi. W tym kontekście warto jednak wspomnieć o ograniczonej zdolności – a być może – i zainteresowaniu organizacji społecznych i uczelni wyższych współpracą z administracją publiczną.

Z perspektywy zagadnienia ZW wymienione powyżej deficyty odczytać można zarówno jako konsekwencję braku umiejętności pozyskiwania i tworzenia wiedzy, jak i niezdolność do wykorzystywania istniejącej. Dowodem jest rozproszenie wiedzy związanej z wynikami ewaluacji polityki i programów publicznych oraz niski poziom ich wykorzystania dla programowania interwencji publicznych.

Wyniki badań ewaluacyjnych stanowią cenny rezerwuar wiedzy, którego wykorzystanie *explicite* prowadziłyby do poprawy jakości rządzenia. Dorobek polskiej praktyki ewaluacyjnej może okazać się szczególnie wartościowy w odniesieniu do wzmocnienia wiedzy taktycznej w polskiej administracji publicznej. Charakter głównego nurtu aktywności ewaluacyjnej lokuje się bowiem w tej sferze wiedzy, a co za tym idzie, wyniki tych badań ewaluacyjnych cechują się wysoką wartością poznawczą i użytecznością.

Inkorporowanie dorobku ewaluacyjnego do praktyki działań instytucji publicznych prowadziłyby do konsolidacji rodzącej się kultury ewaluacyjnej, stanowiącej jeden z podstawowych elementów systemu ZW.

## **ZW w polskiej administracji – istota problemu**

Polska administracja publiczna jest przykładem systemu, w którym występuje nadmiar danych, niepełny zasób informacji i ograniczony potencjał wiedzy, która stanowiłaby podstawę racjonalnych ekonomicznie i społecznie decyzji oraz działań publicznych.

Źródło problemu ZW w administracji publicznej tkwi w jej niezdolności do zrozumienia istoty tej koncepcji i korzyści płynących z jej stosowania. Będąc strukturą wysoce scentralizowaną i hierarchiczną, cechuje się ona konserwatyzmem kulturowym, determinowanym systemem bodźców zniechęca-

jących do działań innowacyjnych. Prymat procedury i racjonalność proceduralna premiujące zachowania zorientowane na wypełnienie procedury, a nie osiągnięcie rezultatu czy rozwiązanie problemu, to istotne demotywatory. Jej cechy konstytutywne i tkwiące w niej immanentne ograniczenia czynią z administracji publicznej strukturę mało plastyczną i raczej reaktywną. Dobrze rozpoznane w socjologii ograniczenia w działaniu dużych sformalizowanych organizacji (proceduralna racjonalność, przemieszczenie celów, wyuczona nieudolność, oligarchizacja i wiele innych) znajdują odzwierciedlenie w administracji publicznej.

Warunkiem koniecznym dla podniesienia wiedzochłonności państwa i jego struktur jest gruntowna modernizacja polskiej administracji publicznej. Podstawowym celem powinna stać się transformacja struktur administracyjnych w system zorientowany na absorpcję wiedzy, prowadząca do efektywnego, skutecznego, responsywnego i interaktywnego sposobu działania. Proces ten miałby charakter dualnego sprzężenia zwrotnego. Z jednej strony podstawą jego przeprowadzenia byłaby wiedza, z drugiej zaś jego efektem byłoby wprowadzenie systemowych rozwiązań w zakresie oparcia działań administracji na wiedzy w większym stopniu niż ma to miejsce obecnie.

Skala wyzwań modernizacyjnych stojących przed polską administracją publiczną jest znacząca. Warunkiem koniecznym powodzenia działań w tej sferze jest ugruntowanie się wśród polityków przekonania, że profesjonalna administracja to zasób kluczowy dla dobrego rządzenia. Proces modernizacji administracji musi być poprzedzony jednak nie tylko debatą publiczną i refleksją na temat pożądanego modelu państwa i jego relacji z obywatelami, ale również uzgodnieniem, czemu administracja ma służyć i w jaki sposób ma działać. Należy zaznaczyć, iż równie ważne jak ustalenie celów działania administracji, jest włączenie w ten proces wszystkich jego kluczowych aktorów, głównie urzędników. Bez ich zaangażowania i zrozumienia celowości wprowadzanych zmian prawdopodobieństwo sukcesu należy uznać za małe.

Wprowadzenia wielu udoskonaleń wymaga obszar zarządzania kadrami w administracji publicznej. Katalog działań, które należałoby podjąć – jest szeroki – począwszy od potrzeby przeprowadzenia rzetelnej analizy wartościowania stanowisk pracy, poprzez wypracowanie motywacyjnego, przejrzystego systemu konkurencyjnego wynagradzania<sup>14</sup> i awansowania, a kończąc na systemie szkoleń i podnoszenia kwalifikacji. Bez kompetentnych i umotywowanych urzędników zarządzanie wiedzą pozostaje szlachetnym, ale romantycznym postulatem.

---

<sup>14</sup> Jednym z istotnych problemów stojących przed polską administracją publiczną jest zjawisko ruchliwości międzysektorowej, przybierające postać rezygnacji z pracy osób o wysokich kwalifikacjach. Potęguje to niezdolność państwa i jego struktur do sterowania ważnymi procesami rozwojowymi.

W ramach politycznie uzgodnionego programu modernizacyjnego konieczne byłoby również podjęcie działań służących wzmocnieniu potencjału analitycznego administracji, jej zdolności do programowania polityk i programów publicznych oraz poprawie zdolności do koordynacji procesów ich implementacji.

Naturalną konsekwencją sygnalizowanych powyżej kierunków działań powinno być przejście od imperatywnych do interaktywnych mechanizmów zarządzania sprawami publicznymi i współdziałania z partnerami zewnętrznymi, opartych o poszanowanie ich autonomii oraz wzrost stopnia odpowiedzialności w relacjach z obywatelami i ich organizacjami.

### Od bezsilności do zarządzania wiedzą

Administracja jako narzędzie państwa w realizacji polityki rozwoju i zaspokajania potrzeb społecznych powinna być jednym z najważniejszych konsumentów wiedzy – wiedzy generowanej w porządku, który znamienici badacze państwa, John L. Campbell i Ove K. Pedersen<sup>15</sup>, określili jako reżim wiedzy, tj. w ośrodkach akademickich, badawczych, uniwersytetach itp. Postrzeganie administracji publicznej jako autonomicznego kreatora wiedzy jest iluzją. Hipotetycznie administracja publiczna mogłaby pełnić rolę systemu generowania wiedzy. Zakładać można, że przynajmniej w części byłaby zdolna rolę taką wypełniać. Jednakże wydaje się, że koszty transakcyjne działania tego systemu przewyższałyby korzyści wynikające z jego funkcjonowania.

Odwołując się do rozważań Schumpetera, można rzec, że podstawową rolą administracji nie jest inwencja (wynajdowanie), ale innowacja (przekładanie wiedzy na praktyczne zastosowanie) i naśladownictwo (pojmwane jako refleksyjne wdrażanie tego, co dobrze gdzieś działa). Adekwatne w tym kontekście byłoby pojęcie refleksyjności instytucjonalnej (zacerpnięte od Giddensa) rozumiane jako rutynowe włączanie nowych informacji lub wiedzy do warunków działania, skutkujące ich rekonfiguracją i reorganizacją<sup>16</sup>.

Aby administracja publiczna mogła być konsumentem wiedzy, musi:

- mieć świadomość własnych ograniczeń (mapa deficytów wiedzy),
- wiedzieć, jakiego typu wiedza jest jej potrzebna,
- wiedzieć, gdzie wiedza jest zlokalizowana „(...) raz dokonanych odkryć (...) «zakryć» na powrót nie można”<sup>17</sup>,

<sup>15</sup> Campbell J.L., Pedersen O.K., *The Varieties of Capitalism and Hybrid Success Denmark in the Global Economy*. Tekst ten prezentowany był w ramach seminarium „Dobre rządzenie” w czerwcu 2007 r., organizowanego przez Uniwersytet Ekonomiczny w Krakowie.

<sup>16</sup> Giddens A., *Nowoczesność i tożsamość*, Wydawnictwo Naukowe PWN, Warszawa 2007, s. 316.

<sup>17</sup> Jarzębski J., *Chaos na cenzurowanym: późna eseistyka Lema*, materiały z VII Krakowskiej Konwencji Metodologicznej, Kraków 2001.

- być w stanie ocenić jakość dostarczonej wiedzy (potrzeba redukcji asymetrii wiedzy w relacjach generator i konsument wiedzy),
- posiadać zdolność do praktycznego wykorzystania przekazanej wiedzy.

Poprawa jakości i stopnia wykorzystywania wiedzy przez administrację publiczną możliwa będzie nie poprzez jednorazowe akcje, ale podjęcie systemowych, długofalowych i instytucjonalnie zakorzenionych działań. Zakładanie, co podaję jako hipotetyczny przykład, że akcyjne szkolenie dla wszystkich pracowników urzędów centralnych z zakresu tworzenia programowania strategicznego w istotny sposób podniesie jego jakość, jest racjonalizacją mitów<sup>18</sup>. Podobnego rodzaju przykładem jest narastające zjawisko powierzania intelektualistom kluczowych stanowisk doradczych w administracji publicznej. Jest to tendencja pozytywna, służąca nasyceniu wiedzą struktur państwowych. Jednak nie można pominąć negatywnych zjawisk, które temu towarzyszą. Ich istotę trafnie oddają śródtytuły jednego z rozdziałów pracy R. K. Mertona, tj.: „Intelektualiści niezależni a intelektualiści biurokratyczni” oraz „Frustracje intelektualisty pozostającego w służbie instytucji państwowej”<sup>19</sup>.

Rozwiązaniem problemu nie jest również mechaniczne rozszerzenie zakresu wykorzystania systemów eksperckich i ich wiedzy przez administrację publiczną. Paradoksalnie bowiem prowadzić to może do dwojakiego rodzaju konsekwencji. Jedną z nich jest defragmentaryzacja i zawężenie rzeczywistości, w której umiejscowiony jest problem stanowiący przedmiot ekspertyzy (co wynika z wysokiego poziomu jej specjalizacji). Dobrze ujął to A. Giddens: „Miarą skuteczności ekspertów w zakresie rozwiązywania problemów staje się często ich umiejętność dostarczania coraz bardziej przejrzyście i precyzyjnych definicji (co z kolei prowadzi do ich dalszej specjalizacji). Jednakże im ostrzej jest zarysowany dany problem, tym bardziej zamazują się sąsiadujące z nim obszary wiedzy, a co za tym idzie, tym trudniej jednostkom, które sięgają po dane rozwiązanie, przewidzieć konsekwencje jego wdrożenia wybiegające poza specyficznie wyróżnioną dziedzinę jego zastosowania. Mimo że ekspertyza jest częścią szerszych systemów abstrakcyjnych, sama dąży do zwiększania ostrości kosztem pola widzenia, przez co może przyczynić się do pojawienia się niezamierzonych konsekwencji, nad którymi nie sposób zapanować – chyba, że w drodze dalszego postępu ekspertyzy odtwarzającego to samo zjawisko”<sup>20</sup>.

<sup>18</sup> Pod pojęciem tym rozumiem postawę opartą na przekonaniu, że podejmowanie znanych i standardowych działań jest rozwiązaniem określonego problemu, a zarazem zwalnia od głębszej refleksji, czy rzeczywiście działania te przynoszą zakładany efekt. To sposób powielania utartych schematów reagowania na problem, na ogół nieskutecznych. Innymi słowy, odwołując się do J. Hausnera, to przykład załatwienia sprawy a nie rozwiązania problemu. Zob. Hausner J., *Pętla rozwoju*, Wydawnictwo Naukowe Scholar, Warszawa 2007.

<sup>19</sup> Merton R.K., *Teoria socjologiczna i struktura społeczna*, Wydawnictwo Naukowe PWN, Warszawa 2002.

<sup>20</sup> Giddens A., op. cit., s. 45.

Problem drugi to zjawisko określane jako efekt kolażu – zestawienie wielu wysoce specjalistycznych ekspertyz generuje poznawczy komunikat o wysokim poziomie heterogeniczności. Jego odczytanie jest skomplikowane, a co za tym idzie, trudno podejmować skuteczne działania. Przykładem może być wielość ekspertyz zamawianych przez urzędy administracji publicznej, a jednocześnie brak zdolności do konceptualnego zintegrowania przekazu z nich wynikającego.

Dla budowania systemowego rozwiązania służącego wzrostowi wiedzy i chłonności administracji publicznej kluczowe jest stworzenie systemu ZW.

Katalog działań na rzecz budowania systemu ZW w polskiej administracji publicznej mógłby przybrać następującą postać<sup>21</sup>:

#### 1. Zasoby ludzkie

- a) zrozumienie przez politycznych zwierzchników administracji publicznej i wysokich urzędników konieczności oparcia działań organizacji publicznych na wiedzy,
- b) kreowanie kultury organizacyjnej zachęcającej do wymiany danych, informacji i wiedzy,
- c) kreowanie liderów inicjujących i stymulujących procesy wymiany wiedzy w organizacji,
- d) wprowadzanie formalnych mechanizmów premiowania i awansowania urzędników aktywnie uczestniczących w procesach wymiany informacji i wiedzy (np. poprzez odpowiedni system rekrutacji, nagród),
- e) tworzenie międzywydziałowych i międzyorganizacyjnych zespołów zadaniowych do poszukiwania innowacyjnych rozwiązań.

#### 2. Proces

- a) identyfikacja – określenie kluczowych z perspektywy zadań realizowanych przez organizację publiczną kompetencji, określenie potencjału wiedzy, który organizacja posiada, zdefiniowanie deficytów wiedzy i sposobów ich zredukowania,
- b) pozyskiwanie wiedzy zarówno z wnętrza organizacji, jak i jej otoczenia, jej formalizacja,
- c) selekcja – ocena przydatności pozyskanej i udokumentowanej wiedzy oraz jej weryfikacja pod kątem jej przydatności,
- d) gromadzenie – kategoryzowanie zgromadzonej wiedzy, jej ujmowanie w formalne standardy organizacyjne, wprowadzanie do pamięci merytorycznej urzędu oraz jej aktualizacja,
- e) dzielenie się wiedzą – klasyfikacja wiedzy oraz jej selekcja i udostępnianie,

---

<sup>21</sup> Zawartość tego katalogu i sposób jego strukturyzacji odzwierciedla podejście do budowania systemu ZW opisane w: Cong X., Pandya K.V., „Issues of Knowledge Management in the Public Sector”, *Electronic Journal of Knowledge Management* 2003, Vol. 1, Issue 2.



- f) stosowanie – wykorzystanie wiedzy w realizacji zadań, np. rozwiązywanie problemów, podejmowanie decyzji, poszukiwanie pomysłów i uczenie się,
  - g) tworzenie – odkrywanie nowej wiedzy poprzez badania, najlepsze praktyki, projekty pilotażowe, analizę danych.
3. Zasoby technologiczne:
- a) określenie odpowiedniego hardware i software dla zarządzania wiedzą w organizacji (dostosowanego do jej specyfiki, procesów w niej zachodzących i realizowanych przez nią zadań),
  - b) tworzenie infrastruktury technologicznej odpowiadającej potrzebom zarządzania wiedzą,
  - c) tworzenie wewnątrzorganizacyjnych kanałów przepływu wiedzy służących budowaniu wzorców współdziałania i dzielenia się wiedzą,
  - d) tworzenie portali wiedzy, wirtualnych platform wiedzy dostępnych dla pracowników organizacji poprzez Intranet, e-mail, grupy dyskusyjne, wideokonferencje,
  - e) przechowywanie zasobów wiedzy w postaci elektronicznej, umożliwiające jej szybki przepływ,
  - f) zapewnienie dostępu do zasobów wiedzy poprzez wykorzystanie technologii ułatwiających kontakty z obywatelami, klientami, dostawcami, partnerami.

Celem przedstawionego katalogu powinno być zapewnienie optymalnych warunków do tworzenia, przesyłania oraz wykorzystania wiedzy w organizacjach i systemach społecznych. Osiągnięcie tego celu nie jest rzeczą łatwą. Immanentne ograniczenia tkwiące w administracji publicznej i środowisko, w którym działa, generują wiele przeciwstawnych – względem tego celu – bodźców. Nie oznacza to jednak, że musi w tym przypadku wystąpić syndrom fatalistycznie pojmowanego determinizmu. Wręcz przeciwnie. Wiele pozytywnych zmian zachodzących w polskiej administracji, w szczególności poziom wykształcenia kadr administracji, pokazuje, że ścieżki zależności i mechanizmy deterministyczne nie mają charakteru absolutnego.

## Zakończenie

Paradygmat ZW zyskuje na znaczeniu. Dostrzega się w nim potencjał, którego wykorzystanie prowadzić może do poprawy jakości rządzenia państwem i zarządzania sprawami publicznymi. Cieszy się on szczególnym zainteresowaniem w tych krajach, które uchodzą za przodujące, jeśli idzie o sposób działania państwa i jego administracji (np. kraje anglosaskie).

Konceptualne oparcie na wiedzy jako podstawowym zasobie dla dobrego rządzenia przesądza o jego intelektualnej atrakcyjności i praktycznej użyteczności. Jego aplikacja do systemów administracji publicznej jest wyrazem

refleksyjnej i racjonalnej adaptacji państwa do wymogów współczesnej rzeczywistości określanej mianem rzeczywistości policentrycznej.

Dostrzegając praktyczne walory ZW, w wielu krajach podejmuje się próby tworzenia systemowych rozwiązań służących jego instytucjonalizacji i zakorzenieniu w praktykach operacyjnych struktur publicznych. Przybierają one postać tzw. systemów ZW. Ich celem jest tworzenie wiedzy, jej systematyzacja, aktualizacja, a nade wszystko jej wykorzystanie w procesach decyzyjnych.

Podejmowane w polskiej administracji publicznej działania na rzecz pełniejszego wkomponowania wiedzy w procesy decyzyjne ocenić należy jako zjawisko ze wszech miar pozytywne. Należy jednak dodać, że wciąż brak jest systemowych rozstrzygnięć, które tworzyłyby warunki dla konsolidacji rozproszonych inicjatyw związanych z ZW i ich ujęcie w ramy rozwiązań ogólnych (tak na poziomie rządu, jak i samorządów). Bez tego wzmacnianie kapitału ludzkiego, ewolucyjna zmiana kapitału kulturowego oraz doskonalenie kapitału technologicznego, koniecznych dla dobrego rządzenia, będzie procesem niezwykle trudnym i długotrwałym.

W opracowaniu działań służących budowaniu kultury ZW kwestią szczególnie istotną jest odpowiednie ujęcie zagadnienia ewaluacji. ZW i ewaluacja mają wspólne korzenie konceptualne. W obu tych przypadkach idzie o to, aby decyzje podejmowane przez państwo i jego struktury zapadały w sposób racjonalny respektujący tak uwarunkowania ekonomiczne, jak i społeczne oczekiwania i preferencje. Ich wspólną cechą jest założenie, że w policentrycznej rzeczywistości oddziaływanie państwa na procesy gospodarcze i społeczne nie jest możliwe bez jego zdolności do refleksji opartej na usystematyzowanym zasobie wiedzy. Takim, który pozwala krytycznie spojrzeć na działania realizowane w przeszłości, celem podejmowania lepszych decyzji w przyszłości. Wiedza pozyskiwana poprzez badania ewaluacyjne powinna być przekazywana do instytucji publicznych poprzez system ZW. Integralność ich relacji polega na tym, że cele ewaluacji *de facto* nie mogą zostać osiągnięte bez systemu ZW, a ten bez strumieni wiedzy pochodzących z badań ewaluacyjnych stanowi konstrukcję pozbawioną istotnych treści.

---

**Stanisław Mazur** – doktor nauk politycznych, absolwent Wydziału Prawa Uniwersytetu Jagiellońskiego (specjalność nauki polityczne). Kierownik Centrum Studiów nad Gospodarką i Administracją Publiczną Akademii Ekonomicznej w Krakowie (1996–1997), dyrektor Małopolskiej Szkoły Administracji Publicznej Akademii Ekonomicznej w Krakowie (1997–2004). Adiunkt w Katedrze Gospodarki i Administracji Publicznej Uniwersytetu Ekonomicznego w Krakowie. Brał udział, jako ekspert i koordynator, w wielu projektach krajowych i międzynarodowych z zakre-

su programowania i analizy polityk i programów publicznych, rozwoju lokalnego i regionalnego, jakości rządzenia, komunikacji i partycypacji społecznej. Napisał kilkudziesiąt publikacji naukowych. Jest autorem i współautorem ponad 100 ekspertyz i opracowań.

## Bibliografia

- Campbell J.L., Pedersen O.K., *The Varieties of Capitalism and Hybrid Success Denmark in the Global Economy*. Tekst ten prezentowany był w ramach seminarium „Dobre rządzenie” w czerwcu 2007 r., organizowanego przez Uniwersytet Ekonomiczny w Krakowie.
- Cong X., Pandya K.V., „Issues of Knowledge Management in the Public Sector”, *Electronic Journal of Knowledge Management* 2003, Vol. 1, Issue 2.
- Giddens A., *Nowoczesność i tożsamość*, Wydawnictwo Naukowe PWN, Warszawa 2007.
- Hausner J., *Pętla rozwoju*, Wydawnictwo Naukowe Scholar, Warszawa 2007.
- Hayek F.A., *Konstytucja wolności*, Wydawnictwo Naukowe PWN, Warszawa 2006.
- Jarzębski J., *Chaos na cenzurowanym: późna eseistyka Lema*, materiały z VII Krakowskiej Konwencji metodologicznej, Kraków 2001.
- Luhmann N., *The Differentiation of Society*, Columbia University Press, New York 1982.
- Mazur S., *Wyzwania stojące przed administracją rządową w latach 2009–2013 a plany działań w ramach PO KL po 2008 roku*, opracowanie przygotowane na zlecenie DSC Kancelaria Premiera, 2008.
- Morgan G., *Obrazy organizacji*, Wydawnictwo Naukowe PWN, Warszawa 1997.
- Merton R.K., *Teoria socjologiczna i struktura społeczna*, Wydawnictwo Naukowe PWN, Warszawa 2002.
- Olsen J.P., „Towards a European Administrative Space?”, *ARENA Working Papers WP 02/26*.
- Polanyi K., *The Great Transformation: The Political and Economic Origins of Our Time* [1944], Beacon Press, Boston 2001.
- Popper, K.R., *Spółczesność otwarte i jego wrogowie*, Wydawnictwo Naukowe PWN, Warszawa 1993.
- Probst G., Raub S., Rombardt K., *Zarządzanie wiedzą w organizacji*, Oficyna Ekonomiczna, Kraków 2002.
- The Learning Government: Introduction and Draft Results of the Survey of Knowledge Management Practices in Ministries/Departments/Agencies of Central Government* – OECD, 2003.

*Leszek Korporowicz*

## **Interakcyjne aspekty procesu ewaluacyjnego: pomiędzy analizą a animowaniem zmian społecznych**

Istnieje wiele sposobów ukazania rozwojowych możliwości, jakie tkwią w realizacji dobrze zaplanowanych i wykorzystanych badań ewaluacyjnych. Niniejszy rozdział zmierza do ukazania mało eksponowanych, a mających dla powodzenia ewaluacji ogromną rolę, interakcyjnych aspektów całego przedsięwzięcia. Są one najczęściej wymieniane jako elementy tzw. warsztatu ewaluacyjnego, choć ich znaczenie jest daleko bardziej istotne i stanowi o wielu innych wymiarach całościowo ujętego procesu ewaluacyjnego. Dla ich zobrazowania w pierwszej części tekstu postawiono kilka tez, które określają właściwą perspektywę postrzegania interakcji ewaluacyjnych, traktowanych jako relacje społeczne, co umożliwiła socjologiczną, aksjologiczną i antropologiczną metarefleksję tego zintegrowanego procesu działań. W drugiej części rozdziału zarysowana jest przyczyna i okres w historii studiów ewaluacyjnych, w jakim narodziła się konieczność analitycznej refleksji nad jakością, znaczeniem i celowością interakcji ewaluacyjnych, w powiązaniu z udanymi próbami ich uspołecznienia i wkomponowania w świadomie inicjowane zmiany w przedmiocie i środowisku ewaluacji. W części trzeciej wskazane są naturalne konsekwencje dostrzegania wartości interakcyjnego traktowania procesu badawczego i uwzględniania ich w nowoczesnych strategiach ewaluacyjnych o charakterze formatywnym. Na zakończenie podniesiony został problem nowych wyzwań ewaluacji w wyłaniającym się z zaawansowanych społeczeństw informacyjnych społeczeństwie wiedzy, w którym interakcje społeczne przybierają nowe formy pośredniej komunikacji elektronicznej, wyznaczają nowe standardy relacji społecznych i modelują realny kształt nowego typu ewaluacji. Praktykowane w niej strategie komunikacyjne określać będą jej funkcje i wymagają rozpoznania nowych zastosowań, perspektyw i problemów, choć ewaluacja, jako forma wiedzy, sama kształtuje charakter nowoczesnego społeczeństwa wiedzy i stanowi jeden z jego wyróżników.

## Warsztat jako interakcja, strategia i jako komunikat

Ewaluacja rozumiana jest na ogół jako stosowany rodzaj badań społecznych, podejmowanych dla określenia wartości konkretnego działania o charakterze celowym, z punktu widzenia przyjętych na rzecz tego badania kryteriów. Warto podkreślić, że nie musi to sprowadzać się jedynie do pomiaru efektywności lub skuteczności ewaluowanych działań. Badanie ewaluacyjne ma zawsze swoją genezę, wyrasta z mniej lub bardziej uświadomionych potrzeb rozwojowych i badawczych organizacji, wchodzi w realny kształt jej kultury organizacyjnej i ewaluacyjnej, angażuje zainteresowane podmioty, a w końcu wchodzi w proces decyzyjny, staje się elementem dyskursu publicznego.

Wszystko, co stanowi więc o warsztacie badań ewaluacyjnych, postrzegać warto w szerszej perspektywie **procesu ewaluacyjnego** jako procesu społecznego o własnej dynamice zmian, aksjologii, rozgrywającego się w gęstej sieci **interakcji** pomiędzy jego indywidualnymi i instytucjonalnymi uczestnikami. Analityczne spojrzenie na problematykę warsztatu ewaluacyjnego jest odniesieniem go do tych właśnie relacji. Stwarzają one konieczność postrzegania ich jako integralnego elementu samego warsztatu, a to znaczy także kompetencji zawodowych zespołu ewaluacyjnego. Aby to przybliżyć, warto postawić kilka wyraźnie sformułowanych tez, które wiążą problematykę metodologii badań ewaluacyjnych, sposobu ich realizacji i wykorzystania, co na ogół nazywane jest łącznie warsztatem ewaluacyjnym, z istotną fazą ich planowania, długiego czasami dyskursu ustalającego ich cele, mniej lub bardziej postrzegane funkcje społeczne, edukacyjne, komunikacyjne, promocyjne, czy te najistotniejsze, czyli rozwojowe.

Po pierwsze: **Czynności i kompetencje określające warsztat ewaluacyjny konkretnego badania są stałym oraz integralnym komponentem długiego i wielofazowego procesu.** Wyznacza on nie tylko charakter procedur badawczych, ale czyni je funkcjonalnym komponentem złożonych relacji w kontekście konkretnego społecznego środowiska ewaluacji. Te istotne fazy całego procesu ewaluacyjnego to:

- a) rozpoznawanie przedmiotu i społecznych podmiotów procesu,
- b) negocjowanie, a następnie wybór celów i funkcji badania ewaluacyjnego na poszczególnych jego etapach i dla konkretnych jego uczestników,
- c) wybór konkretnego lub skonstruowanego na potrzeby procesu ewaluacyjnego modelu badań dopasowanego do oczekiwanych rezultatów i sposobu ich wykorzystania.

Czynności te są już współcześnie tak bardzo eksponowane, iż stanowią one niemal konieczną część samego „warsztatu”, który w strategiach ewaluacji, nastawionych na animowanie ich funkcji demokratycznych i rozwojowych,

stał się rodzajem aksjologii społecznej wielu działań służących stymulowaniu pozytywnych zmian społecznych<sup>1</sup>.

Pozostałe komponenty procesu ewaluacyjnego, które decydują o kluczowych umiejętnościach warsztatowych, to możliwie precyzyjna konstrukcja projektu ewaluacji, obejmującego poza sformułowaniem przedmiotu ewaluacji, także podstawowe, kluczowe pytania, kryteria ewaluacyjne, metody badań, sposób doboru próby, jak również elementy zarządzania procesem. Uzupełniają go fazy generowania i zbierania danych, następnie ich analizy i interpretacje, czynności związane z pisaniem raportów, następnie ich komunikowanie, a więc różne formy prezentacji, a na koniec zasadnicza faza wdrożeniowa i aplikatywna.

Po drugie: **Ujawnienie interakcyjnych aspektów procesu ewaluacyjnego wnosi konieczność określenia roli „aktorów sceny ewaluacyjnej”**, następnie reguł ich interakcji, które mogą być **autorytarne**, gdy ustalane są jedynie przez podmiot zamawiający ewaluację, **technokratyczne**, gdy czynią to wyłącznie sami ewaluatorzy i **populistyczne**, gdy dominującą rolę odgrywają ewaluowani i gdy to oni sterują doбором kryteriów, bazy danych i ich interpretacją. Taka forma interakcji jest wynaturzeniem procesu uspołecznienia ewaluacji i nie stanowi formy jej demokratyzacji, która polega na skutecznym dialogu wszystkich jej uczestników.

Po trzecie: Interakcyjna perspektywa procesu ewaluacyjnego ukazuje **dynamiczny charakter znaczenia**, jakie jego uczestnicy nadają zarówno celom, jak i poszczególnym komponentom, działaniom, pojęciom, ocenom i interpretacjom konkretnych zachowań, stwierdzeń, a nawet danych. Niezwykle inspirująca jest pod tym względem cała tradycja interakcjonizmu symbolicznego, która oświeśla społeczne konstrukcje przyjmowanych znaczeń, ich przewartościowania, dokonywane zarówno przez ewaluatorów, jak i uczestników ewaluowanych „odkryć”, zmianę stanowisk kluczowych postaci procesu jako rezultatu jego dynamiki. Perspektywa ta nakazuje obserwować różnice, które ujawniają się, gdy bliżej wnikamy w świat wartości uczestników, przeobrażenia, jakie dokonują się w trakcie procesu w wyniku zmiany stanowiska innych osób, ale także swoiste pozy, gry, a nawet mistyfikacje czynione dla wykreowania wizerunku własnych działań. Interakcyjność procesu ewaluacyjnego ujawnia się więc nie tylko w analizie zebranych

---

<sup>1</sup> Korporowicz L., „Metodologia badań ewaluacyjnych. O potrzebie integracji i kumulacji doświadczeń”, w: *Ewaluacja ex-ante – podsumowanie doświadczeń administracji polskiej*, Ministerstwo Rozwoju Regionalnego, Warszawa 2007; Olejniczak K., *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INERREG III*, Wydawnictwo Naukowe Scholar, Warszawa 2008; Haber A. (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2007.

danych, ale we wszystkich relacjach wewnątrz i w kontekście bezpośrednio realizowanych badań.

Po czwarte: **Stosowany w badaniach ewaluacyjnych warsztat ma charakter komunikatu** wysyłanego do jego uczestników, niezależnie od deklarowanych celów ewaluacji. Jest on w tym sensie odrębną praktyką interakcyjną, która określa jej rzeczywisty sens, niezależnie także od projektowanych funkcji. To interakcje „aktorów” ewaluacyjnej sceny ostatecznie określą kontrolno-sprawozdawczy, czy wspierająco-rozwojowy efekt ewaluacji. Nie przełamują tego nawet formalne definicje metod badawczych, które są „miękkie” lub „twarde” w zależności od tego, czemu służą i jak są wykorzystywane.

Ów komunikat realizować może bardzo różne strategie interakcyjne bliższe formule i praktyce opisanej jako „Inter – Akcja” lub „re – akcja”. Ta pierwsza podkreśla wzajemność, zwrotność i wymianę (Inter), jak również sprawczość, podmiotowość i otwarcie (Akcja), ta druga odtwórczość i powielanie (re) oraz pasywność i pozorny charakter zaangażowania (akcja). Tak wyglądają rytualne formy ewaluacji przeprowadzane z administracyjnej i proceduralnej konieczności, gdy sprowadzamy komunikacyjne aspekty ewaluacji, także te poza wąsko rozumianym warszatem, do funkcji mało twórczych, technokratycznych przekazów.

Interakcyjne aspekty procesu ewaluacyjnego nie były dostrzegane w teorii i praktyce badań ewaluacyjnych aż do późnych lat 60., w których szerzej spojrzano na nie jako na rodzaj działań społecznych, czasem nawet interwencji społecznej uwikłanej w kontekst, realia i nie zawsze jawne cele towarzyszące ewaluacji.

## Interakcyjna historia studiów ewaluacyjnych

Za ojca ewaluacji uważa się niemal we wszystkich pracach poświęconych historii ewaluacji Ralpa Tylera, który w 1929 roku rozpoczął systematyczne studia nazywane już ewaluacyjnymi w Ohio State University. Niemal od samego początku były one zorientowane na uzyskiwanie mierzalnych rezultatów swych analiz i traktowane niemal jako synonim procedur pomiarowych<sup>2</sup>. Na efekt badań ewaluacyjnych patrzono więc jako na finał kumulujący w sobie wszystkie czynności, które uchwycić mają badany „stan rzeczy”, poza historią jego stwarzania i interpretacji, poza procesem badawczym, tak jakby działania te, jak i sam proces, istniały poza warszatem ewaluacji.

---

<sup>2</sup> Owen J.M., *Program Evaluation. Forms and Approaches*, Sage Publ., London, Thousand Oaks, New Delhi 1999; Norris N., *Understanding Educational Evaluation*, CARE University of East Anglia, Kogan Press, London 1990.

Zadaniem wczesnych studiów ewaluacyjnych było możliwie najbardziej czytelne i zoperacjonalizowane określenie celów ewaluowanego programu, które dają się przełożyć na pożądane zachowania lub umiejętności jego uczestników, a następnie określenie stopnia ich realizacji poprzez porównanie zamiarów z rzeczywistymi efektami. Cała droga dochodzenia do nich, dokonujące się na niej zmiany i interakcyjny kontekst dynamizmu pozostawał nieistotny. Co więcej, takie rozumienie celów ewaluacji nader często sprowadzało cały problem do uchwycenia zmiany jedynie indywidualnych zachowań jednostek. Powodowało to zmarginalizowanie ponadindywidualnych efektów działania programu, że nie wspomnę już o analizie instytucjonalnych, organizacyjnych czy społecznych jego wymiarów.

Nie bez powodów pod koniec lat 60. podniesiono więc problem tak zwanego „redukcjonizmu behawioralnego”, który eliminuje z pola ewaluacyjnych dociekań inaczej określone cele, jak na przykład sferę doświadczeń społecznych uczestników programu, ich rolę w określeniu znaczenia zdobywanych umiejętności, zmiany systemów wartości itd. Należy jednak z całą mocą zauważyć, że już w latach 30. właśnie Tyler był jednym z pierwszych inicjatorów poważnego rozszerzenia funkcji i celów ewaluacji, dostrzegając wiele powstających wówczas pytań i uproszczeń, jak na przykład rola społeczności klasowej i samego programu nauczania w kształtowaniu poziomu sprawności i wiedzy ucznia.

Postulowana w związku z tym przez Tylera korekta strategii i myślenia ewaluacyjnego nie została doceniona na skutek dominacji rozwijających się szybko, choć dość jednostronnie, technik pomiaru psychometrycznego, socjometrycznego i dydaktycznego. Taki rodzaj redukcjonizmu w ewaluacji, szczególnie edukacyjnej, ale również w ewaluacji wielu programów społecznych możemy obserwować niestety także we współczesnej fazie rozwoju wielu polskich programów i szkoleń ewaluacyjnych. Teza ta znajduje ponadto uzasadnienie, gdy weźmiemy pod uwagę zadziwiająco podobne elementy sytuacji społecznej, która rodziła potrzebę ewaluacji zarówno w USA, jak i w Polsce lat 90.

Po pierwsze, wynikała ona z konieczności zahamowania zjawisk korupcji w kręgach decydenckich ze względu na brak zobiektywizowanych, publicznie kontrolowalnych i rzeczywiście jawnych kryteriów udzielania koncesji, zezwoleń lub podejmowanie działalności gospodarczej czy publicznej. Po drugie, w wielu obszarach rzeczywistości, szczególnie w dziedzinie edukacji istniała ogromna różnorodność lokalnie akceptowanych standardów, które trudno było uznać za porównywalne. Komplikowało to z natury rzeczy możliwość prowadzenia ogólnonarodowej polityki edukacyjnej, jak również uniemożliwiało wręcz szereg ponadlokalnych procedur selekcyjnych, egzaminacyjnych lub certyfikacyjnych, które wymagają przyjęcia względnie jednolitych kryteriów postępowania i analizy. Po trzecie, dość powszechnie podzielano przekonanie o niedostosowaniu amerykańskich programów na-



uczania do wymogów szybko rozwijającego się, nowoczesnego społeczeństwa przemysłowego, co także stanowi jeden z głównych powodów polskiej reformy systemu edukacji.

Interesujące jest, że Tyler stając na czele rozwijających się w opisanych powyżej warunkach studiów ewaluacyjnych, dostrzegał jednocześnie kryjące się w nich uproszczenia. Proponował odniesienie do kontekstu ewaluowanych programów, do specyfiki realizujących je grup i do zamierzeń oraz doświadczeń ich realizatorów. Nie ulega wątpliwości, iż były to pierwsze postulaty rozbudowywania społecznej, interakcyjnej bazy ewaluacji, a jak nazwane to zostanie znacznie później, „demokratyzacji” studiów ewaluacyjnych.

O ile jednak dokonania Tylera w dziedzinie badań ewaluacyjnych zostały powszechnie rozpoznane i dobrze opisane, o tyle niemal nieznanym jest fakt prowadzenia w tych samych latach przez Floriana Znanieckiego poważnych studiów edukacyjnych, które miały w istocie charakter ewaluacyjny zarówno ze względu na swą genezę, jak i cele. Został on mianowany w Columbia University dyrektorem programu badawczego realizowanego w latach 1930–1933. Raport z tych badań przez wiele lat zalegał w archiwum Uniwersytetu, chociaż dawał podstawy do zupełnie innego myślenia o ewaluacji<sup>3</sup>. Znaniecki w przeciwieństwie do Tylera wychodził z zupełnie innych założeń metodologicznych, odwołując się do społecznego doświadczenia badanych w sposób, jaki dobrze znany jest socjologom humanistycznym w postaci strategii badań uwzględniających postulat „współczynnika humanistycznego”. Nakazuje on spojrzenie na badaną rzeczywistość nie tylko oczami badacza, ale także samego badanego i to poprzez znaczenia, które konstruuje on w konkretnym kontekście swego życia oraz realiów procesu badawczego.

Występujący już stale we współczesnych badaniach socjologicznych postulat nie jest niestety obecny w większości badań ewaluacyjnych. Zapomina się o tym interakcyjnym w istocie postulacie badawczym, niezwykle istotnym w fazie przyjmowania kryteriów ewaluacyjnych. Nader często są one ustalane jako zewnętrznie przyjmowane standardy określonego typu działań: administracyjne, profesjonalne, technologiczne, a często są wynikiem możliwości przełożenia ich na język pomiaru. Znaniecki wyprzedzał pod tym względem współczesnych, wysuwając tezy, które na gruncie studiów ewaluacyjnych pojawią się dopiero w latach 80., gdy powstanie problem ewaluacyjnego wykluczenia i to nie w prostym sensie niedoreprezentowania w próbie badawczej określonych grup ewaluowanych, ale w znaczeniu

---

<sup>3</sup> Znaniecki F., *Education and Social Change*, E. Hałas, P. Lang (red.), Frankfurt am Main 1998; Patton M.Q., *Qualitative Evaluation and Research Methods*, Sage Publ., Newbury Park, London, New Delhi 1990; Korporowicz L., „Nowe twarze ewaluacji. W kierunku inspiracji, analizy, czy inspiracji”, w: J. Kubin, J. Kwaśniewski (red.), *Socjotechnika; kontrowersje, rozwój, perspektywy*, IPSiR UW, Warszawa 2000, s. 85–95.

„roztopienia”, pominięcia, czy przeinterpretowania ich wizji badanych doświadczeń, relacji, kluczowych odniesień, ośrodków koncentracji wartości, emocji czy zainteresowania.

Zewnętrznie przyjmowane standardy oceny powstają poprzez unifikację wskaźników, które dla celów porównawczych muszą być często intersubiektywnie, a nawet międzyinstytucjonalnie ujednolicone. Strategia ta sprawdza się przy ewaluacji dużych programów, szczególnie tych, które objęte są nadzorem międzynarodowym i koordynowane z uwzględnieniem ścisłej kontroli wykonania. Wprowadza ona jednak wiele spustoszenia, jeśli odniesiona jest do różniących się kulturowo warunków ich realizacji, gdy przedmiot ewaluacji stanowi wyabstrahowaną społecznie sferę działań ocenianą przez równie oderwane kryteria, które nie są poddawane społecznej negocjacji i które w badaniu ewaluacyjnym znaczą coś innego niż w doświadczeniu organizacji i jej pracowników.

Interakcyjny typ studiów ewaluacyjnych bazuje więc nie tyle na parametrycznym, ile opisowym i interpretatywnym sposobie badania tego, co Znaniecki nazywał „faktami kulturowymi”, odróżniając je od „faktów przyrodniczych”. Analiza, którą podjął, koncentrowała się na adekwatności kształcenia kompetencji społecznych i kulturowych młodych Amerykanów w kontekście zupełnie nowych wyzwań stawianych przez szybko modernizujące się społeczeństwo oraz struktury państwowe. Znaniecki podjął się więc wartościowania procesów edukacyjnych, uwzględniając zmiany kulturowe, zmiany porządku społecznego, transformację i wymogi, jakie zmiany te wywoływały w charakterze osobowości społecznej, typie kompetencji kulturowej niezbędnej w procesie skutecznego uczestnictwa w życiu społecznym, gospodarczym i naukowym. Analiza, którą zaproponował, znacznie rozszerzała spektrum kryteriów ewaluacyjnych oraz lokalizowała ten typ analiz na wielu płaszczyznach. Jednocześnie obejmowała zmieniający się charakter wartości społecznych stanowiący naturalną podstawę wszelkich procesów i procedur wartościowania.

Bardzo istotna różnica w rozumieniu przedmiotu ewaluacji pojawiła się pod koniec lat 60., gdy postrzegać zaczęto istotną statyczność modeli ewaluacyjnych nastawionych na jednorazowe uchwycenie relacji pomiędzy nakładami i efektami ewaluowanych programów bez analizy dynamiki, a nawet swoistej „dramaturgii” procesu ich realizacji. Dynamizacja studiów ewaluacyjnych wynikała zarówno z konieczności poszerzenia branych pod uwagę „zmiennych” i zakresu rejestrowanych danych, ale także z potrzeby uchwycenia tego, co dzieje się w trakcie badanego programu, a co może stanowić istotny jego rezultat pomijany w przypadku badania jedynie rezultatów końcowych. Nie ma poza tym powodów, aby pomijać wszystko to, co dzieje się w trakcie realizacji programu, a co znajduje miejsce w doświadczeniu jego uczestników i to w sposób, który wymaga znacznego poszerzenia stosowanych w ewaluacji metod.

W zmianie postrzegania klasycznych modeli ewaluacji bardzo ważną rolę odegrały prace: Lee J. Chronbacha *Course Improvement Through Evaluation* z 1963 roku, Michaela Scrivena *The Methodology of Evaluation* z 1967 roku i Roberta Stake'a *The Countenance of Educational Evaluation* także z 1967 roku. Doprowadziły one do:

- wyraźnego sformułowania postulatu rozszerzenia zakresu danych branych pod uwagę w badaniach ewaluacyjnych z coraz powszechniejszym odwoływaniem się do danych o charakterze jakościowym,
- odejścia od modelu skoncentrowanego wyłącznie na finalnych konsekwencjach analizowanych programów i z większym uwzględnieniem tego, co stanowi o procesie realizacji programu,
- postrzegania owego procesu w jego złożoności, uwikłaniach i kontekście, w relacji do środowiska społecznego,
- wyraźnego podkreślenia roli interakcji społecznych w dynamice procesu badawczego jako konstytutywnych dla zrozumienia jego funkcji,
- unikania alienacji działań ewaluacyjnych w wyniku rozminięcia się z potrzebami zainteresowanych podmiotów instytucjonalnych i indywidualnych,
- uświadomienia potrzeb analiz ewaluacyjnych o charakterze całościowym, wieloaspektowym.

Podobne rozwarstwienia, narastające dylematy i polemiki, które pojawiły się w rozwoju studiów ewaluacyjnych w latach 60. pogłębiły się jeszcze bardziej na początku lat 70. Większość z nich trwa do chwili obecnej. Ich istotą jest spór o uprawniony zakres zainteresowań, funkcji, metod i zastosowań badań ewaluacyjnych. Początek lat 70. to niezwykle interesująca debata nad bilansem osiągnięć, ale i porażek programów ewaluacyjnych realizowanych niejednokrotnie za wielkie pieniądze i w stosunkowo dużej skali. Już wówczas otwarcie stawiano zarzut technokratyczności ewaluacji, zaniedbania problematyki etycznej, zbyt instrumentalnego wkomponowania ewaluacji w procedury menedżerskie, arbitralnego społecznie tworzenia kryteriów ewaluacyjnych, braku postrzegania procesu ewaluacyjnego w jego animacyjnych funkcjach zmian społecznych idących w kierunku uspołecznienia i ewaluacji uczestniczącej, stymulującej procesy uczenia się organizacyjnego, rozwoju umiejętności dyskursu publicznego i wkomponowania w standardy współczesnego społeczeństwa wiedzy.

Niemal przewrót w postrzeganiu funkcji i społecznej roli ewaluacji przyniosła konferencja zorganizowana w grudniu 1972 roku przez Churchill College w Cambridge z udziałem ewaluatorów brytyjskich, amerykańskich, szwedzkich oraz przedstawicieli wielu urzędów i fundacji. Najpełniejszym wyrazem przewartościowań oraz poważnej krytyki, z jaką spotkały się tradycyjne modele ewaluacji, jak również rodzajem manifestu, była prezentowana w czasie konferencji praca M. Parletta i D. Hamiltona *Evaluation as Illumination: A New Approach to the Study of Innovative Pro-*

grammes<sup>4</sup>. Autorzy, podobnie jak wielu dyskutantów, odbierali w istocie wiele administracyjnej władzy ewaluatorom, podnosząc jednocześnie znaczenie ich roli w rozwoju świadomości działań własnych, ale także innych uczestników interakcji ewaluacyjnych. Sformułowano wówczas postulaty zachęcające do:

- dialogu z różnymi audytoriami i uczestnikami procesu ewaluacyjnego,
- „oświeclania” złożoności ludzkich doświadczeń, jakie powstają w trakcie programu i rekonstrukcji ich znaczeń,
- rozumiejącej analizy organizacyjnych, psychologicznych i kulturowych aspektów ewaluowanych działań i ewaluacji jako takiej,
- komunikatywności i interaktywności ze wszystkimi stronami i podmiotami ewaluacji już w fazie jej planowania, w trakcie realizacji, ale także w okresie zastosowań.

Stworzono w ten sposób podwaliny ewaluacji dialogicznej, której społeczna „racjonalność” daleko wybiega poza pomiarową ewaluację lat 30. Michael Q. Patton, jeden z czołowych ewaluatorów współczesności twierdzi, iż stworzony w ten sposób model ewaluacji odchodzi od metodologii hipotetyczno-dedukcyjnej na rzecz „paradygmatu ustawicznych wyborów, podkreślając wielość metod, alternatywnych ujęć i... wkomponowując je w kontekst specyficznych sytuacji i wymogów ewaluacyjnych”<sup>5</sup>. W ten sposób ewaluacja wymaga nie tyle przestrzegania gotowych i niezmiennych reguł, ile elastycznego współgrania z potrzebami, szansami i wyzwaniem konkretnej społeczności. Taką **interakcyjną strategię ewaluacji** rozwijali i praktykowali następnie Robert Stake, Egon G. Guba i Yvonne S. Lincoln, Helen Simons, ewaluatorzy kontynuujący dorobek brytyjskiego Centre for Applied Research in Education w University of East Anglia (Norwich), przez wiele lat kierowanego przez współtwórcę „ewaluacji demokratycznej”, Berry MacDonalda<sup>6</sup>. To tradycja korespondująca z bardzo intensywnie rozwijanymi od lat 70. jakościowymi, interpretacyjnymi badaniami na gruncie nauk społecznych, które ujawniają wagę kultury symbolicznej w różnych obszarach działalności człowieka, w tym także działalności badawczej. Gdyby nakreślić drzewo genealogiczne interakcyjnej koncepcji studiów ewaluacyjnych, to choć nie wyrasta ono z klasycznej teorii interakcjonizmu symbolicznego tzw. Szkoły Chicagowskiej Georga H. Meade’a, Williama I. Thomas’a, Roberta Parka lat 30. oraz tzw. drugiej Szkoły Chicagowskiej już po II wojnie światowej, tworzonej przez Herberta Blumera, Anselma Straussa oraz wielu innych luźno powiązanych

<sup>4</sup> Parlett M., Hamilton D., *Evaluation as Illumination: A New Approach to the Study of Innovative Programmes*, Centre for Research in Educational Sciences, University of Edinburgh, 1972.

<sup>5</sup> Patton M.Q., *Creative Evaluation*, Sage Publ., Beverly Hills 1981, s. 270.

<sup>6</sup> Scriven M., *Evaluation Thesaurus*, Sage Publ., Newbury Park, London, New Delhi 1991; Guba E.G., Lincoln Y.S., *Fourth Generation Evaluation*, Sage Publ., Newbury Park, London, New Delhi 1989.

ze sobą badaczy<sup>7</sup>, to z pewnością drzewa te stykałyby się w wielu miejscach, teoretycznie wzbogacając ewaluację, ale dodając społecznego konkretności interakcjonizmowi.

Aktywność i dokonania tej grupy osób, szybko powstających ośrodków myśli i badań ewaluacyjnych stworzyły z warsztatu ewaluacyjnego, jak i całej ewaluacji działania daleko wykraczające poza doraźną diagnostykę społeczną. To właśnie ujawnienie dynamicznych, interaktywnych komponentów procesu ewaluacyjnego stworzyło z niej element szerszej polityki społecznej, dla której nie jest ona jedynie „narzędziem”, ale strategią i stymulatorem myślenia rozwojowego w obrębie kultur organizacyjnych, w relacjach międzyinstytucjonalnych, a także w sferze programów międzykulturowych, które dopominają się bardzo dialogicznego myślenia o wartościach i wartościowaniu oraz podnoszą bardzo wysoko poprzeczkę świadomości interakcyjnej ewaluacji.

### Dialogiczne wartości ewaluacji formatywnej

Szczególnie istotna rola interakcji ewaluacyjnych, które wymagają od ewaluatorów umiejętności prowadzenia dialogu i negocjacji społecznych, jawi się w **formatywnych** modelach ewaluacji, a więc takich, które wkomponowane są w proces realizacji ewaluowanych programów i których działanie przynosi rezultaty daleko wcześniej przed ich zakończeniem. Ewaluacja formatywna wymaga jednak stosunkowo zaawansowanych umiejętności badawczych i organizacyjnych wynikających z :

- ciągłego dialogu z kierownictwem programu,
- uważnej obserwacji procesu ewaluacyjnego poprzez ustawiczne przywoływanie zasadniczych jego celów, przyjętych kryteriów, poprawności reguł dobierania próby,
- korekty instrumentów badawczych,
- efektywnego sposobu komunikowania się wewnątrz zespołu ewaluacyjnego szczególnie w trakcie zaobserwowanych problemów, konfliktów, dysfunkcji, zmian i modyfikacji,
- efektywnego sposobu komunikowania się z uczestnikami „sceny ewaluacyjnej” w różnych okresach realizacji programu.

Istotna dla ewaluacji formatywnej dyskusja i ciągła weryfikacja kryteriów ewaluacyjnych, stanowiąca jeden z istotnych elementów uspołecznienia procesu ewaluacyjnego, odsłania bardzo poważny wymiar współczesnej ewaluacji, który ignorowany jest przez ujęcia technokratyczne. W sposób ciągły pojawiają się bowiem pytania o zasadnicze cele, przeznaczenie, użytkowników, najbardziej adekwatne formy oraz techniki ewaluacji. Pytania te ujawniają wielość stanowisk, różnice interesów, odmienne wizje jej za-

<sup>7</sup> Hałas E., *Interakcjonizm symboliczny*, Wydawnictwo Naukowe PWN, Warszawa 2006.

stosowań. W ewaluacji technokratycznej odpowiedzi na te pytania padają w wąskim gremium tak zwanych specjalistów od ewaluacji i decydentów. Demokratyczne wizje ewaluacji podkreślają natomiast ten właśnie newralgiczny moment procesu ewaluacyjnego. Stwarza on bowiem szansę na wykorzystanie fazy projektowania i realizacji ewaluacji do pobudzenia dyskursu społecznego pomiędzy zainteresowanymi stronami, ujawnienia swoich stanowisk, artykulacji potrzeb i pośrednio, do ujawnienia szeregu problemów, które powinny być wzięte pod uwagę w samej ewaluacji lub też poza obszarem jej funkcjonowania.

Dyskurs ten zaliczyć trzeba do efektów ewaluacji, co obala powszechnie panujący stereotyp uznający jedynie raport ewaluacyjny za właściwy produkt procedur ewaluacyjnych. Owa polemika mieści się w pośrednich ich celach, częstokroć stanowi przełamanie wielu barier komunikacyjnych i może być impulsem do zmiany kultury organizacyjnej. Z tych to powodów tak często inicjatorzy ewaluacji demokratycznej podkreślali jej potencjalną rolę w zarządzaniu zmianami, w inicjowaniu przeobrażeń świadomości organizacyjnej, w tym też świadomości celów i form działania. Ewaluatorzy muszą traktować powstające wówczas interakcje jako istotną część tak strategii, jak i warsztatu ewaluacyjnego. Stają się wielokrotnie nie tylko analitykami danych, ale także animatorami zmian mentalnych, komunikacyjnych, transformując stosunek do jawności działania, kryteriów jego oceny.

Każda z faz procesu ewaluacyjnego ma swoją społeczną dynamikę, swoich „aktorów”, rozgrywana jest w konkretnym kontekście<sup>8</sup>. Każda z faz generuje także właściwe dla niej problemy etyczne. Są one elementem szeroko pojętej kultury ewaluacyjnej i określają szereg norm, które obowiązywać mają wszystkich uczestników procesu ewaluacyjnego. Uwzględnienie ich jest komponentem społecznego uprawomocnienia działań ewaluacyjnych i powodem wielu nieporozumień, a nawet konfliktów, rozbieżności oczekiwań, odmiennego postrzegania ról i w konsekwencji braku powodzenia wielu przedsięwzięć ewaluacyjnych. Stąd też, szczególnie w warunkach ubogiej kultury ewaluacyjnej i braku wiedzy o jej funkcjach, kluczowe znaczenie ma negocjowanie celów, przebiegu i warunków realizacji ewaluacji. Wyjaśnienie tych kwestii, co z pewnością powinno być świadomie inicjowane przez ewaluatorów, nie tylko dookreśla wzajemne oczekiwania, ale buduje obowiązujące wzajemnie normy, legitymizuje procedury i działania ewaluacyjne, pozwala pokonać restryktywne wyobrażenia o ewaluacji, otwiera możliwości komunikacyjne.

Interakcyjne rozumienie procesu ewaluacyjnego lokalizuje się więc w dobrze znanej już tradycji interpretatywnych badań ewaluacyjnych roz-

---

<sup>8</sup> Patton M.Q., *Qualitative Evaluation and Research Methods*, Sage Publ., London 1990; Aspinwall K., Simkins T., Wilkinson J.F., McAuley M.J., *Managing Evaluation in Education*, Routledge, New York-London 1992.

poczętych, jak było już wcześniej wspomniane, przez Floriana Znanieckiego w latach 30. ubiegłego wieku, ale poddane silnej renowacji, a nawet rozwojowi w latach 70., wkomponowane w realia współczesności przez ewaluatorów „czwartej generacji”. Co odróżnia tego typu studia od interakcjonizmu symbolicznego, to ich zdecydowanie stosowany charakter, podejmowanie praktycznych wyzwań, jakie mieszczą się w tożsamości studiów ewaluacyjnych podejmowanych nie dla jednego, ale co najmniej dla trzech celów: praktycznej korekty realnie podejmowanego działania, nie wykluczając instrumentalnych celów tak podejmowanych badań, dla rozumienia realnej wartości i dynamiki, w jakiej przebiega, oraz dla uwolnienia tkwiących w nim potencjałów.

### **Interakcje ewaluacyjne w społeczeństwie wiedzy**

Zupełnie nowe wyzwania tak przed warsztatem, jak i całym procesem ewaluacyjnym, stają w dobie niezwykle szybkiego rozwoju technologii informacyjno-komunikacyjnych społeczeństwa informacyjnego i jego zaawansowanej postaci, jaką jest społeczeństwo wiedzy<sup>9</sup>. Jest to „typ organizacji społecznej, w której tworzenie i wykorzystanie wiedzy staje się strategicznym czynnikiem jego funkcjonowania i rozwoju, a sama wiedza kluczowym, dynamicznie rosnącym zasobem określającym cywilizacyjny kształt warunków życia, system wartości oraz sposoby uczestnictwa w kulturze symbolicznej i życiu społecznym”.

Wszystkie makrostrukturalne, ekonomiczne, ale i społeczne przyczyny jego rozwoju, takie jak konieczność optymalizacji, koordynacji i rozwoju jakościowego działań, stanowią jednocześnie o koniecznych zmianach w formach, technikach i celach badań ewaluacyjnych. Konieczność uwzględnienia tych zmian coraz bardziej widoczna będzie poprzez rosnącą potrzebę pytania nie tylko o przedmiot i sposób realizacji ewaluowanych działań, ale nade wszystko o ich cel, typ akceptowalnej społecznie wartości, jej kulturowego sensu. Choć logika faz procesu ewaluacyjnego i sposobu uczestniczenia w nim nie zostaje zmieniona, to znaczącej modyfikacji ulegać będą typy interakcji, kształtując nowe kompetencje ewaluatorów. Zmiana generalna wynika z możliwości interakcji zapośredniczonych, które poszerzają zakres uczestników procesu, pod warunkiem jednak, iż posiadają oni stosowne umiejętności komunikacyjne. Dostęp do uczestników ewaluowanych działań może być więc teoretycznie łatwiejszy, w praktyce jednak może być regulowany wieloma czynnikami, które nie zawsze są nam znane i prze-

---

<sup>9</sup> Doktorowicz K., *Europejski model społeczeństwa informacyjnego. Polityczna strategia Unii Europejskiej w kontekście globalnych problemów wieku informacji*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2005.

widywalne, jak na przykład poziom umiejętności korzystania z sieciowych technik komunikacyjnych, zapoznania się z operacyjnymi technikami systemów komunikacji elektronicznej. Wszystko to generuje ukryte linie podziałów, selekcji i marginalizacji społecznych, a więc także uspołecznionych zasobów informacyjnych dostępnych w różnych fazach zbierania i przetwarzania informacji.

Istotnym aspektem interakcyjnym usieciowionego sposobu uczestniczenia w procesie ewaluacyjnym jest możliwość, konieczność, ale i różnego rodzaju problematyczność korzystania z istniejących już zasobów informacyjnych, ważnych dla ustalenia celów, kluczowych pytań, kryteriów, próby badawczej, sposobu weryfikacji danych oraz innych istotnych czynności ewaluacyjnych. „Społeczeństwo sieci” to zwielokrotnione systemy wielopoziomowych interakcji, przepływów informacji, odnajdywania dróg wzajemnego oddziaływania, dopełniania, ale i weryfikacji, odniesień i porównań. To właśnie społeczeństwo sieci czyni z interakcji nową jakość współczesnej cywilizacji, zmusza do poszerzenia i wzbogacenia horyzontów postrzegania świata, innych ludzi, działań, jak i ich rezultatów<sup>10</sup>.

Ewaluatorzy społeczeństwa informacyjnego i społeczeństwa wiedzy są w istocie animatorami takich przepływów, ukierunkowują je w celu wygenerowania wiedzy, a następnie wprowadzają ją w obszar organizacyjnego doświadczenia<sup>11</sup>. Nowym składnikiem ich kompetencji staje się jednak konieczna biegłość w dziedzinie technologii informacyjno-komunikacyjnej, która generuje drogi „przepływów”, ich sprawność, szerokość, dostępność, ale i celowość, trafność, rzetelność. Ogromne zasoby informacyjne stają się z jednej strony bardzo bogatym źródłem, z jakiego korzystać może zespół ewaluacyjny przystępujący do projektowania celu, funkcji i sposobu badania, z drugiej strony wprowadzać może swoisty chaos informacyjny, wymagający rozbudowanych umiejętności kierowania informacją dla potrzeb projektu, a następnie realizacji i zastosowania badań ewaluacyjnych. Warto zaznaczyć, iż wszystkie czynności związane z pozyskiwaniem, przekazywaniem i przetwarzaniem informacji prowadzić mają do uzyskania jakościowo odmiennej **wiedzy o wartościach** ewaluowanego programu, która jest jednym, choć nie jedynym, z ważnych produktów procesu ewaluacyjnego. To, w jaki sposób informacje przetworzone zostają w wiedzę, zależy od przyjętego modelu ewaluacji, jej funkcji i społecznego zakorzenienia, od tego, jakie ma wewnętrzne

<sup>10</sup> Castells M., *Społeczeństwo sieci*, Wydawnictwo Naukowe PWN, Warszawa 2007; Mattelart A., *Społeczeństwo informacji*, tłum. Pomorski J.M., Universitas, Kraków 2004; Batorski D., Marody M., Nowak A. (red.), *Społeczna przestrzeń Internetu*, Academica, Warszawa 2006; Truszkowska M., Wojtkowiak J., *Społeczeństwo informacyjne*, Wydawnictwo Uniwersyteu Gdańskiego, Gdańsk 2008; Bard A., Soderqvist J., *Netokracja. Nowa elita władzy i życie po kapitalizmie*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2006.

<sup>11</sup> Korporowicz L., „Ewaluacja i społeczeństwo”, w: E. Hałas (red.), *Rozumienie zmian społecznych*, Towarzystwo Naukowe KUL, Lublin 2001.



możliwości generowania społecznych doświadczeń rozwoju, uczenia się i stawania impulsem przekształceń.

**Ewaluacja staje się jednym z wyróżników społeczeństwa wiedzy**, jako że to ona właśnie, wiedza (a ewaluacja prowadzi do zdobycia pewnego typu wiedzy), uczyniona została strategicznym jego zasobem decydującym o możliwościach i kierunkach rozwoju. Ewaluacja, jako wiedza, jest jego stymulatorem, a czasami bardzo istotnym warunkiem. Ze względu na fakt, że społeczeństwo wiedzy jest czymś innym niż tylko gospodarką opartą na wiedzy, nowego znaczenia nabierają zapośredniczone przez technologie informacyjno-komunikacyjne interakcje, zarówno te pomiędzy uczestnikami procesu, pomiędzy nimi a kontekstem ich działania, wewnątrz zespołu ewaluacyjnego, a w końcu wśród odbiorców ewaluacji, dla których ma być ona dostępnym i inspirującym źródłem wiedzy i refleksji. Społeczeństwo wiedzy wzmacniać będzie intensywność, jak i zakres różnego typu przepływów, oddziaływań i komunikacji, coraz bardziej potrzebując jednocześnie ich selekcji, ukierunkowania, a to znaczy identyfikacji ich wartości, zarówno dla szerszych celów społecznych, ale także dla jakości przeprowadzanych w nim ewaluacji.

Bardzo istotne cechy współczesnego społeczeństwa wiedzy, które z pewnością wyznaczać będą nowe standardy ewaluacji, wymienia Lech W. Zacher w bardzo analitycznej pracy *Transformacje społeczeństw: od informacji do wiedzy*. Pisze on, iż: „...społeczeństwo wiedzy – w ramach permanentnej wszech- i samo-edukacji – powinno umieszczać w swoich systemach edukacyjnych «pakiety nowej wiedzy» – służącej jego umacnianiu i rozwojowi. Zapewne niełatwo będzie sprecyzować co powinno wchodzić w skład owej nowej wiedzy, ale przecież można dać przykłady nowych obszarów wiedzy (...):

- podejście systemowe, holizm;
- teoria złożoności (wielkich systemów);
- teoria chaosu (czy raczej koncepcje chaotyczności);
- teoria sieci (wszelkiego rodzaju i zasięgu);
- analiza ryzyka i niepewności;
- koncepcja, metody i procedury tzw. analizy skutków (*impact assessment*) oraz jej odmian typu ewaluacja techniki (*technology assessment*), ewaluacja skutków środowiskowych (*environmental impact assessment*), ewaluacja skutków społecznych (*social impact assessment*) itp.;
- wielowymiarowa perspektywa oglądu rzeczywistości; multi, inter- i transdyscyplinarność;
- modelowanie, symulacja, prognozowanie...;
- metody poszukiwania synergii procesów, działań itp.”<sup>12</sup>.

<sup>12</sup> Zacher L.W., *Transformacje społeczeństw. Od informacji do wiedzy*, Wydawnictwo C.H.Beck, Warszawa 2007, s. 232.

Pomijając w tym momencie zagadnienie relacji ewaluacji i pomiaru (*assessment*), których z wielu względów nie warto ze sobą utożsamiać, powyższa charakterystyka wskazuje, jak dalece współczesne myślenie o ewaluacji odbiegać musi od prostego odwzorowania rzeczywistości. Staje się to jeszcze bardziej widoczne w kontekście innych przywołanych przez Zachera cech charakteryzujących przechodzenie od społeczeństwa informacyjnego do społeczeństwa wiedzy, jak np.:

- zdolność do nietendencyjnej, wielokryterialnej oceny,
- zdolność do ogarnięcia wielu zmiennych,
- zdolność do naprzemiennie stosowanej analizy i syntezy,
- umiejętność myślenia asocjacyjnego i równoległego oraz „wiązania spraw”, czyli identyfikacji sprzężeń,
- zdolność do myślenia o przyszłości,
- umiejętność posługiwania się intuicją, wyobraźnią, fantazją jako przydatnymi sposobami dociekań otwierającymi nowe horyzonty<sup>13</sup>.

Z ewaluacyjnego punktu widzenia istotna jest też rosnąca w społeczeństwach wiedzy waga **kultury symbolicznej**, której znaczenie odkrywane jest w miarę intensyfikowania różnorodnych zjawisk wielokulturowości personelu i „kapitału kulturowego” organizacji, wagi procesów komunikacji międzykulturowej w wymiarze interpersonalnym, międzygrupowym oraz pomiędzy organizacjami. Znaczący dla procesu ewaluacyjnego jest także samorealizacyjny, nieinstrumentalny wymiar wiedzy jako wartości w ponowoczesnej charakterystyce kultury pracy, co bezpośrednio wpływa na dobór kryteriów ewaluacyjnych, społeczną percepcję interakcji ewaluacyjnych jako metody pozyskiwania i wartościowania wiedzy, następnie na sposób dzielenia się zdobytą i zweryfikowaną wiedzą. Wyzwania, jakie stają przed ewaluacją w społeczeństwie selektywnego przetwarzania informacji, a więc w systematycznie konstruowanym społeczeństwie wiedzy, ująć można w postaci zestawienia (tab. 1).

Ewaluacja w społeczeństwie wiedzy staje się funkcjonalnym elementem charakterystycznym dla ogólnego wzrostu znaczenia refleksyjnego i analitycznego sposobu realizacji wielu działań. Implikacją interakcyjną tej zmiany jest znaczący krok ku różnym metodom wymiany, interpretacji oraz przetwarzania informacji w formie bezpośredniej i zapośredniczonej, aż do tworzenia i dzielenia się powstałą w ten sposób wiedzą. Sama wiedza ma jednak również coraz bardziej dialogiczny i rozwojowy charakter oddalając się od koncepcji jednoznacznej wyroczni i statycznego obrazu na rzecz dynamicznego procesu animującego procesy uczenia się, wymiany znaczeń i uwewnętrznienia ich celu.

<sup>13</sup> Ibidem, s. 231.

Tabela 1

Wyłaniające się cechy społeczeństwa wiedzy	Odpowiadające im cechy ewaluacji
<b>A. Cechy technologiczne:</b>	
Wysoki poziom użycia zaawansowanych technologii informacyjno-komunikacyjnych, w których pozyskane informacje niemal natychmiast stają się elementem komunikacyjnych przekazów	Użycie elektronicznych technologii informacyjno-komunikacyjnych staje się jedną z podstawowych cech warsztatu ewaluacyjnego i kompetencji zawodowych ewaluatora
Szeroki dostęp i ustawiczne aktualizowanie zasobów informacyjnych w wyniku rosnącej swobody sieciowego przepływu informacji	Wszystkie fazy i czynności planowania procesu ewaluacyjnego, jak również czynności czysto badawcze wykorzystują, ale i analizują konsekwencje sieciowego przepływu informacji
Postępujące możliwości opanowania nadmiaru i chaosu informacji poprzez technologiczne warunki ich przetwarzania, selekcji i porządkowania	Fundamentalne znaczenie jasno określonych i poddających się empirycznej operacjonalizacji kryteriów ewaluacji jako generatorów wiedzy
<b>B. Cechy strukturalne</b>	
Relacjonalne sposoby tworzenia wiedzy ogarniającej wiele perspektyw i zmiennych, ich konteksty, „wiązanie spraw”, „identyfikację sprzężeń”	Uwrażliwienie na wielowymiarowy i wielofunkcyjny charakter procesu ewaluacyjnego o wielorakich możliwościach wykorzystania w różnych jego fazach, w relacji do różnych jego uczestników
Coraz szersze stosowanie wiedzy w procesach podejmowania decyzji poprzez modelowanie, prognozowanie, symulowanie i projektowanie rzeczywistości	Rosnąca rola formatywnych strategii procesu ewaluacyjnego i jego obecności w kulturze organizacyjnej, kulturze pracy i procesach decyzyjnych konkretnych przedsięwzięć
Dynamizowanie oglądu rzeczywistości, problemów i rozwiązań ujmowanych prospektywnie, rozwojowo, nastawionych na racjonalizowanie zmian i projektowanie transformacji	Rosnąca rola demokratycznych strategii procesu ewaluacyjnego traktowanych jako „zaproszenie do rozwoju” w planowaniu strategicznym organizacji
Poszukiwanie synergii procesów, łączenia zasobów, animowanie wartości dodanych	Rosnąca rola dialogicznych, partycypacyjnych strategii i technik warsztatu ewaluacyjnego w jego funkcjach „refleksyjnej praktyki”, jak również uwrażliwienie na synergetyczne kryteria ewaluacji

Tabela 1 – cd.

Wyłaniające się cechy społeczeństwa wiedzy	Odpowiadające im cechy ewaluacji
<b>C. Cechy interakcyjne</b>	
Decentralizacja ośrodków tworzenia i upowszechniania wiedzy powiązana z procesem decentralizacji władzy i uczestniczenia w poziomych systemach komunikacji	Rosnąca rola ewaluacji jako stymulatora procesu uczenia się poprzez stymulowanie umiejętności stałego identyfikowania, interpretowania i korzystania tak przez ewaluatorów, jak i ewaluowanych z nowych źródeł wiedzy jako formy dyskursu publicznego
Medialne zapośredniczenie procesów przekazu i selekcji wiedzy	Rosnąca rola medialnie zapośredniczonych form komunikowania w każdej fazie procesu ewaluacyjnego wkomponowanych w kompetencje ewaluatorów, ale i w analityczny obszar refleksji nad konsekwencjami posługiwania się „elektronicznymi węzłami komunikacji”
Międzykulturowa i transkulturowa perspektywa wartościowania procesu tworzenia, dzielenia się i stosowania wiedzy	Rosnące znaczenie ewaluacji międzykulturowej

## Zakończenie

Integralnie pojęty proces ewaluacyjny spełnia wiele, wzajemnie powiązanych i współzależnych funkcji. Jego cele realizują się nie tylko w postaci finalnych ustaleń oraz skutecznych ich aplikacji, ale rozkładają się na różne fazy poprzez różnorodne interakcje, w jakie wchodzi jego uczestnicy. Ich świadomość oraz przypisywane im cele stanowią integralny element warsztatu ewaluacyjnego, jak również współokreślają leżącą u ich podstaw społeczną aksjologię ewaluacji. Znaczenie interakcyjnych aspektów ewaluacji najwyraźniej widać poprzez odwołanie się do perspektywy interakcjonizmu symbolicznego oraz koncepcji współczynnika humanistycznego dobrze znanego z prac Floriana Znanieckiego, który realizował go w swoich studiach z pogranicza socjologii edukacji, zmian społecznych i ewaluacji już na początku lat 30. na Uniwersytecie Columbia, dając inspirujące nawet współcześnie wzorce ewaluacji humanistycznej, zdominowane przed laty przez orientacje behawioralne, a następnie wąsko utylitarne, pomiarowe cele ewaluacji. Jej dalsza historia udowodniła jednak wagę, jak i możliwości integralnie traktowanego procesu badawczego, który ukazał ewaluatorom lat 80. twórcze, sprawcze

i demokratyzujące funkcje interakcji ewaluacyjnych jako instrumentu, ale i wartości animujących rozwojowy charakter zmian w kulturze pracy, świadomości, jakości i społecznej celowości realizowanych działań, wybiegając poza kontrolne i sprawozdawcze cele ewaluacji. Współcześnie oraz w przyszłości nowe wyzwania stawia przed nią zaawansowane społeczeństwo informacyjne oraz nowoczesne społeczeństwo wiedzy. Upowszechnienie dostępu do informacji, a także chaos informacyjny, swoboda przepływu informacji, jak również zapośredniczenie oraz uzależnienie od technologii informacyjno-komunikacyjnych stwarzają nowe realia warsztatu ewaluacyjnego, potrzebę rozpoznania nowych kompetencji ewaluatorów, nowych możliwości, ale i zagrożeń. Ewaluacja wchodzi w nową fazę rozwoju społeczeństwa otwartego, społeczeństwa, które z wiedzy tworzy jeden z podstawowych swoich zasobów funkcjonalnych i aksjologicznych, i które tworzy je w warunkach globalnej sieci powiązań informacyjnych, nowych systemów zapośredniczonych interakcji. Systemy te nadają informacji i wiedzy fundamentalne znaczenie w kreowaniu decyzji niemal we wszystkich sferach życia, stymulują postawy **innowacyjne, synergetyczne, refleksyjne, projektujące, nastawione na samorealizację i kooperację**, otwarte na wartości i współlistnienie różnych kultur w dynamicznych procesach komunikacji międzykulturowej. Te ostatnie nakazują szczególnie rozpoznanie antropologicznych podstaw interakcji ewaluacyjnych w stopniu proporcjonalnym do ich umiędzynarodowienia, przenikania granic państwowych i kulturowych, a to oznacza także dotykanie problemów różnorodności i zmienności świata wartości. Od wrażliwości i wyobraźni ewaluatorów zależeć będzie, czy rozpoznają te wartości nie tylko w środowisku, lecz również we własnym systemie działania, w aktualizowanych kompetencjach, postawach i własnym warsztacie pracy.

---

**Leszek Korporowicz** – doktor habilitowany, profesor Uniwersytetu Warszawskiego, prorektor Szkoły Wyższej im. Bogdana Jańskiego, socjolog, współzałożyciel i pierwszy prezes Polskiego Towarzystwa Ewaluacyjnego. Były członek, a obecnie ekspert Państwowej Komisji Akredytacyjnej, inicjator i konsultant wielu badań, studiów podyplomowych, w tym Podyplomowego Studium Ewaluacji Programów Społecznych w Instytucie Socjologii UW, organizator i koordynator międzynarodowych szkoleń ewaluacyjnych, od 15 lat ekspert wielu komisji ministerialnych ds. ewaluacji, w tym Ministerstwa Rozwoju Regionalnego. Jest autorem licznych publikacji z zakresu metodologii i społecznych funkcji ewaluacji, szczególnie ewaluacji jako zmiany społecznej i procesu demokratyzacji.

## Bibliografia

- Aspinwall K., Simkins T., Wilkinson J.F., McAuley M.J., *Managing Evaluation in Education*, Routledge, New York–London 1992.
- Bard A., Soderqvist J., *Netokracja. Nowa elita władzy i życie po kapitalizmie*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2006.
- Batorski D., Marody M., Nowak A. (red.), *Spoleczna przestrzeń Internetu*, Academica, Warszawa 2006.
- Castells M., *Spoleczeństwo sieci*, Wydawnictwo Naukowe PWN, Warszawa 2007.
- Doktorowicz K., *Europejski model społeczeństwa informacyjnego. Polityczna strategia Unii Europejskiej w kontekście globalnych problemów wieku informacji*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2005.
- Guba E.G., Lincoln Y.S., *Fourth Generation Evaluation*, Sage Publ., Newbury Park, London, New Delhi 1989.
- Haber A. (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2007.
- Hałas E., *Interakcjonizm symboliczny*, Wydawnictwo Naukowe PWN, Warszawa 2006.
- Korporowicz L., „Ewaluacja i społeczeństwo”, w: E. Hałas (red.), *Rozumienie zmian społecznych*, Towarzystwo Naukowe KUL, Lublin 2001.
- Korporowicz L., „Metodologia badań ewaluacyjnych. O potrzebie integracji i kumulacji doświadczeń”, w: *Ewaluacja ex-ante – podsumowanie doświadczeń administracji polskiej*, Ministerstwo Rozwoju Regionalnego, Warszawa 2007.
- Korporowicz L., „Nowe twarze ewaluacji. W kierunku inspiracji, analizy, czy inspiracji”, w: J. Kubin, J. Kwaśniewski (red.), *Socjotechnika; kontrowersje, rozwój, perspektywy*, IPSiR UW, Warszawa 2000, s. 85–95.
- Mattelart A., *Spoleczeństwo informacji*, tłum. Pomorski J.M., Universitas, Kraków 2004.
- Norris N., *Understanding Educational Evaluation*, CARE University of East Anglia, Kogan Press, London 1990.
- Olejniczak K., *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INTERREG III*, Wydawnictwo Naukowe Scholar, Warszawa 2008.
- Owen J.M., *Program Evaluation. Forms and Approaches*, Sage Publ., London, Thousand Oaks, New Delhi 1999.
- Parlett M., Hamilton D., *Evaluation as Illumination: A New Approach to the Study of Innovative Programmes*, Centre for Research in Educational Sciences, University of Edinburgh, 1972.
- Patton M.Q., *Creative Evaluation*, Sage Publ., Beverly Hills 1981.
- Patton M.Q., *Qualitative Evaluation and Research Methods*, Sage Publ., Newbury Park, London, New Delhi 1990.
- Scriven M., *Evaluation Thesaurus*, Sage Publ., Newbury Park, London, New Delhi 1991.
- Truszkowska M., Wojtkowiak J., *Spoleczeństwo informacyjne*, Wydawnictwo Uniwersyteu Gdańskiego, Gdańsk 2008.
- Zacher L.W., *Transformacje społeczeństw. Od informacji do wiedzy*, Wydawnictwo C.H.Beck, Warszawa 2007.
- Znaniecki F., *Education and Social Change*, E. Hałas, P. Lang (red.), Frankfurt am Main 1998.



*Agnieszka Haber, Maciej Szalaj*

## **Kapitał społeczny – warunek i efekt udanych ewaluacji**

Kapitał społeczny w ciągu ostatnich trzech dekad należy do najpopularniejszych pojęć powstałych w obszarze nauk społecznych. Oprócz licznych odniesień, jakie znajdujemy w literaturze fachowej, z pojęciem tym stykamy się niejednokrotnie w debacie publicznej krążącej wokół zagadnień aktywności obywatelskiej, sprawności instytucji politycznych. Kapitał społeczny stanowi niejednokrotnie swoiste „koło ratunkowe”, pomocne wielu publicystom i innym komentatorom dnia powszedniego, gdy stają przed koniecznością wyjaśnienia pierwotnej przyczyny zapóźnień cywilizacyjnych danego regionu, porażek projektów reformatorskich czy też patologii występujących w przestrzeni publicznej. Można również śmiało przyjąć, iż kapitał społeczny zadomowił się na poziomie języka potocznego, zyskuje szczególnie podatny grunt w żargonie adeptów niezliczonych uczelni, wtajemniczanych w arкана socjologii, politologii, czy też innych, nieco bardziej efemerycznych dyscyplin naukowych.

Obecność refleksji skoncentrowanej wokół kapitału społecznego widoczna jest od pewnego czasu także w obszarze praktyki i teorii badań ewaluacyjnych, co znalazło swoje najbardziej wyraźne odzwierciedlenie w tematyce najbliższej konferencji Europejskiego Towarzystwa Ewaluacyjnego. Według organizatorów opierać się ona będzie na wieloaspektowym ujęciu relacji pomiędzy ewaluacją a zróżnicowanymi formami kapitału społecznego. W tym kontekście, niniejszy rozdział stanowi głos w aktualnej dyskusji angażującej międzynarodową społeczność praktyków i odbiorców ewaluacji.

Niekwestionowana „moda na kapitał społeczny” nie jest jednak główną przesłanką, dla której tekst ten znalazł miejsce w tomie *Środowisko i warsztat ewaluacji*. Perspektywa kapitału społecznego, pod warunkiem jej selektywnego zastosowania, wnosi bowiem dużą wartość poznawczą w zakresie opisu organizacyjnego (instytucjonalnego) środowiska badań ewaluacyjnych. Pozwala ona wyeksponować pomijane zazwyczaj uwarunkowania właściwego osadzenia ewaluacji w praktyce instytucji publicznych i ich społecznych partnerów, które rozstrzygają z kolei o użyteczności, czyli zasadności, wszelkich przedsięwzięć spod szyldu „ewaluacji”.



W pierwszej części rozdziału przedstawiony zostanie zarys genezy pojęcia kapitału społecznego oraz różnorodność jego definicji i zastosowań. Wskazana zostanie także propozycja selektywnego rozumienia tego terminu, ograniczonego do kluczowych aspektów organizacyjnych uwarunkowań ewaluacji, ze szczególnym uwzględnieniem kontekstu polskiej kultury instytucjonalnej.

W części drugiej omówimy podwójną rolę kapitału społecznego w obszarze badań ewaluacyjnych. Wskażemy na znaczenia kapitału społecznego zarówno jako jednego z warunków realizacji dobrych badań ewaluacyjnych, jak i wytworu ewaluacji zorganizowanych i realizowanych zgodnie z uznanymi standardami profesjonalnymi.

Część trzecia natomiast posłuży jako podsumowanie, w którym zarysowane zostanie znaczenie i specyfika kapitału społecznego na tle innego rodzaju kapitałów i zasobów dostępnych instytucjom publicznym. Wskażemy także, w jaki sposób poszczególne formy kapitału mogą być „inwestowane” w tworzenie struktur i kultury ewaluacyjnej, a tym samym w dobre, wysokiej jakości i użyteczne z punktu widzenia szerokiego spektrum odbiorców, badania ewaluacyjne.

## **Kapitał społeczny – geneza, definicje, funkcje**

Pojęcie kapitału społecznego wprowadzone zostało po raz pierwszy na początku lat 70. przez Pierre’a Bourdieu, jednego z najbardziej znanych socjologów europejskich minionego stulecia. W kilka lat później rozważania nad kapitałem społecznym podjęli Amerykanie, przyjmując nieco odmienne niż Bourdieu założenia wyjściowe. Ukształtowały się tym samym dwa główne i dominujące do dziś podejścia. Ich wspólne elementy są bezpośrednią pochodną przyjętej analogii wobec innych, „klasycznych” form kapitału (materialnego, finansowego) i koncentrują się na zdolności do inwestowania w relacje społeczne, co przynieść może osobliwe zyski, jakich nie dostarczyłaby żadna inna forma kapitału.

Bourdieu, w odróżnieniu od autorów amerykańskich, podkreśla zarówno pozytywne, jak i negatywne aspekty inwestowania kapitału społecznego, który definiuje w kategoriach neutralnych jako „zbiór rzeczywistych i potencjalnych zasobów, jakie związane są z posiadaniem trwałej sieci mniej lub bardziej zinstytucjonalizowanych związków wspartych na wzajemnej znajomości i uznaniu – lub inaczej mówiąc z członkostwem w grupie – która dostarcza każdemu ze swych członków wsparcia w postaci kapitału posiadanego przez kolektyw, wiarygodności, która daje im dostęp do kredytu w najszerszym sensie słowa”<sup>1</sup>.

<sup>1</sup> Bourdieu P., „The Forms of Capital”, w: J.G. Richardson (red.), *Handbook of Theory and Research for the Social Education*, Greenwood Pres, New York 1986, s. 248.

Kapitał społeczny, według podejścia Bourdieu, stanowi atrybut jednostek, jednakże poprzez agregację może wywierać silny wpływ na ogół stosunków społecznych, zwłaszcza na kształt społecznej struktury i hierarchii. Przykładem wskazywanym przez Bourdieu był, oparty na kapitale społecznym, system stanowy z instytucją tytułów szlacheckich jako mechanizmem zabezpieczeń uprzywilejowanej pozycji jednej z grup społecznych.

Według tradycji amerykańskiej kapitał społeczny postrzegany jest natomiast jako atrybut przynależny całym społecznościom, co pośrednio przełożyło się na wyraźnie pozytywne konotacje związane z tym terminem. Ten zasób, będący w dyspozycji szeroko rozumianej społeczności (zazwyczaj bez rozróżnienia na istniejące w jej obrębie podgrupy, koalicje, czy koterie), przynosi bowiem ogólnospołeczne zyski. James Coleman definiuje kapitał społeczny jako: „dobro społeczne (...) grupę podmiotów społecznych (*entities*), które mają dwa wspólne elementy: są częścią jakiejś społecznej struktury i ułatwiają wspólne działania – zarówno jednostek, jak i całych instytucji w ramach tej struktury. Tak jak inne formy kapitału, kapitał społeczny jest produktywny (*productive*) – umożliwia realizację pewnych celów, których osiągnięcie nie było by możliwe w sytuacji jego braku (...)”<sup>2</sup>.

Kontynuatorem rozważań nad kapitałem społecznym był Francis Fukuyama. Główny punkt jego analiz to wpływ kapitału społecznego na rozwój, dzięki redukcji kosztów transakcyjnych i tym samym zwiększonej wydajności. Według niego: „kapitał społeczny jest istotny dla efektywnego funkcjonowania nowoczesnej gospodarki i jest warunkiem *sine qua non* stabilnej demokracji liberalnej”<sup>3</sup>.

Ciesząca się dziś niezwykle popularnością koncepcja odwołuje się przede wszystkim do dwóch ostatnich ze wspomnianych powyżej podejść. Zyskała ona szeroki rozgłos na początku lat 90. wraz z ukazaniem się prac Roberta Putnama, będącego kontynuatorem myśli Colemana.

Putnam dokonał empirycznej ilustracji funkcjonowania kapitału społecznego na przykładzie współczesnych Włoch oraz Stanów Zjednoczonych. Główne tezy płynące z tekstów tego autora ogniskują się wokół konstatacji, iż o rozwoju społeczno-gospodarczym oraz sprawności funkcjonowania instytucji demokratycznych decydują nie tylko czynniki ekonomiczne oraz formalno-prawne. Owym brakującym, zapomnianym czynnikiem, jest według Putnama kapitał społeczny definiowany jako pojęcie odnoszące się: „do takich cech organizacji społeczeństwa, jak zaufanie, normy i powiązania, które mogą zwiększyć sprawność społeczeństwa ułatwiając skoordynowane działania”. Wychodząc z takiego założenia, Putnam przywołuje

<sup>2</sup> Coleman J.S., „Social Capital in the Creation of Human Capital”, *American Journal of Sociology*, nr 94, s. 98.

<sup>3</sup> Fukuyama F., „Social capital, civil society and development”, *Third World Quarterly* 2001, nr 1, s. 7.

element wspólny wszystkim definicjom kapitału społecznego, pisząc: „tak jak i inne postaci kapitału, kapitał społeczny jest produktywny, umożliwia bowiem osiągnięcie pewnych celów, których nie dałoby się osiągnąć, gdyby go zabrakło”<sup>4</sup>.

Koncepcja kapitału społecznego, rozumianego jako unikalny zasób społeczności, stanowiła atrakcyjną odpowiedź na wyzwania poznawcze związane z rozpadem dwubiegunowego układu sił światowych oraz turbulencjami obserwowanymi w krajach, które w myśl założeń wielu ekspertów powinny gładko zaadaptować model ustrojowy typowy dla rozwiniętych krajów Zachodu. W efekcie, kapitał społeczny włączony został do głównych nurtów badań zleczanych przez kluczowe instytucje międzynarodowe. Ich wyniki stały się podstawą dla argumentów formułowanych na rzecz podejmowania określonych decyzji i przedsięwzięć.

Poprzez raporty sygnowane logo czcigodnych instytucji kapitał społeczny przeniknął do języka manifestów i programów, obrastając, naturalną kolejną rzeczą, pewnym patosem. Zatarciu uległa również precyzja odnośnie rozumienia samego pojęcia oraz jego powiązania z kluczowymi dla współczesnego świata procesami. Kapitał społeczny określony został nawet z pewną ironią „lekarstwem na wszelkie bolączki” (*something of cure-all*) nowoczesności. Aby nie zatracić jego podstawowych wartości i jednocześnie wyjaśnić pojęcie kapitału społecznego, konieczne jest oczyszczenie interpretacji ze zbędnego patosu, czym też zajmiemy się w dalszej części tekstu.

## Kapitał społeczny w procesach badań ewaluacyjnych

Powiązanie kapitału społecznego z problematyką ewaluacji wymaga dużej dbałości o precyzję stosowanych pojęć. Pole naszych zainteresowań ograniczymy zatem do występowania kapitału społecznego w obrębie instytucji publicznych, prowadzących ewaluację, oraz partnerów, będących odbiorcami wyników badań ewaluacyjnych. Przyjmiemy również założenie, iż kapitał społeczny stanowi właściwość ponadjednostkową, a więc przynależną konkretnym zbiorowościom – na potrzeby tych rozważań – rozumianą zarówno jako kadry określonych instytucji lub inne grupy zaangażowane w tworzenie oraz realizację danych przedsięwzięć publicznych, jak i społeczeństwo. Nasze rozumienie kapitału społecznego naznaczone jest wyraźnie pozytywnym wartościowaniem, jakkolwiek nie negujemy występowania w interesującym nas obszarze zjawiska negatywnego kapitału społecznego (np. mechanizmy korupcyjne, tzw. „układy”, zawłaszczanie państwa, nietransparentne relacje

<sup>4</sup> Putnam R.D., *Demokracja w działaniu: Tradycje obywatelskie we współczesnych Włoszech*, Znak, Kraków 1995.

publiczno-prywatne, wykluczenia itd.)<sup>5</sup>. Wyjście z tak nakreślonych założeń, pozwala na postawienie tezy o dwóch funkcjach, jakie kapitał społeczny pełni w kontekście badań ewaluacyjnych, będąc jednocześnie koniecznym warunkiem wyjściowym i wytworem udanej ewaluacji.

### **Kapitał społeczny jako wyjściowy warunek udanej ewaluacji**

Udana ewaluacja powinna spełniać przynajmniej kilka kryteriów. Po pierwsze, badanie ewaluacyjne spełniać musi określone wymogi jakościowe – jeśli chodzi o rzetelność i sposób gromadzenia wiedzy. Po drugie, konieczna jest odpowiednia aranżacja całego procesu ewaluacji, tak by był on wkomponowany w zaadaptowane praktyki działania danej instytucji. Z tego punktu widzenia o kapitale społecznym, w kontekście warunków wyjściowych dla realizacji ewaluacji, możemy mówić w dwóch wymiarach ze względu:

- 1) na projektowanie, realizację oraz zarządzanie procesem ewaluacji,
- 2) na instytucjonalną percepcję produktów procesu ewaluacji.

Każdy projekt ewaluacyjny kreuje unikalne połączenia i relacje. Nowo tworzone więzi modyfikują istniejący kapitał społeczny i stwarzają szanse dla wykorzystywania zasobów i ich wymiany pomiędzy jednostkami wewnątrzorganizacyjnymi (patrz rysunek 1). Powstają mniej lub bardziej dynamiczne sieci relacji i kontaktów z udziałem wielu uczestników, odgrywając rolę w strategii realizacji projektów ewaluacyjnych, jak również w późniejszym etapie odbioru i wykorzystania jej wyników. Wykorzystane w projektach ewaluacyjnych sieci oparte na współpracy<sup>6</sup> pozwalają nie tylko na szybkie zidentyfikowanie potrzeb informacyjnych, dopasowanych do rzeczywistych potrzeb instytucji, ale również sprawne zaprojektowanie (konceptualizację), przygotowanie (operacjonalizację) i realizację ewaluacji.

Fukuyama wskazywał, że zdolność organizacji do wykorzystywania szans rozwoju tkwi w kapitale społecznym, w tym w umiejętnej współpracy wewnątrz organizacji i w ramach środowiska zewnętrznego w celu realizacji wspólnych interesów. Taka zdolność łączenia się w grupy, stanowi znaczną część kapitału społecznego, z którego wywodzi się w dużej mierze zaufanie między uczestnikami organizacji, które jest istotną i wymierną wartością ekonomiczną<sup>7</sup>.

<sup>5</sup> Zob. więcej Kazimierczak T., „Kapitał społeczny a rozwój społeczno-ekonomiczny – przegląd pojęć”, w: T. Kazimierczak, M. Rymśza (red.), *Kapitał społeczny. Ekonomia społeczna*, Instytut Spraw Publicznych, Warszawa 2007, s. 54–55.

<sup>6</sup> Każdy uczestnik poprzez wniesienie nowych kontaktów, wiedzy, umiejętności, kompetencji rozwija sieć, tym samym jeszcze bardziej katalizując i rozwijając zasoby i możliwości organizacyjne danego projektu ewaluacyjnego.

<sup>7</sup> Fukuyama F., *Zaufanie. Kapitał społeczny a droga do dobrobytu*, PWN, Warszawa 1997.

W działaniach polskich instytucji publicznych ewaluacja stanowi jednak częstokroć praktykę obcą, narzuconą formalnym wymogiem związanym z procedurami wydatkowania funduszy unijnych. Jedną z kluczowych barier utrudniających instytucjonalne osadzenie ewaluacji w kontekście rodzimej administracji publicznej jest postrzeganie badań ewaluacyjnych jako czynności kontrolnej, mającej na celu rozliczenie osób odpowiedzialnych za realizację badanego obszaru działań publicznych<sup>8</sup>. Rozliczanie, w przytaczanym przez nas sensie, wiąże się jednak z powodowanym nieufnością paraliżem instytucji, który uniemożliwia podjęcie pomiędzy poszczególnymi osobami czy komórkami kooperacji ukierunkowanej na czerpanie korzyści z prowadzonych przez daną instytucję badań ewaluacyjnych.



Rysunek 1. Aktorzy ewaluacji w organizacji

Zasygnalizowany powyżej problem jest odzwierciedleniem szerszego zjawiska cechującego polski kontekst funkcjonowania systemu instytucji publicznych. Deficyty zaufania stanowią bowiem jedną ze sztandarowych cech polskiego społeczeństwa<sup>9</sup>. Literatura przedmiotu wielokrotnie wskazuje natomiast, iż z poziomem zaufania ściśle powiązana jest jakość instytucji publicznych. Prawidłowość ta znajduje również odniesienie do ewaluacji, które wpisane w „nieprzyjazny”, nasycony nieufnością kontekst stają się przedsięwzięciami z góry skazanymi na porażkę. W najlepszym razie napotykają na poważne bariery, w pierwszej kolejności – wewnątrz rodzimej organizacji. Instytucja zlecająca ewaluację nie wyraża bowiem wówczas rzeczywistego zainteresowanie jej wynikami, a ewaluatorzy prowadzący badanie natrafiają na swoisty mur nieufności, uniemożliwiający im odpowiednie zaplanowanie badania (np. poprzez poznanie zawiłości kontekstu badanego programu, zdiagnozowanie potrzeb informacyjnych instytucji). Sukcesem w rozumieniu tak funkcjonującej instytucji jest wówczas formalne zrealizowanie badania ewaluacyjnego, przy ograniczonej do minimum popularyza-

<sup>8</sup> Należy przy tym zaznaczyć, iż mówimy w tym miejscu o rozliczeniu w znaczeniu bardziej zbliżonym do znanego z krajowej debaty publicznej, niżli wobec anglojęzycznego terminu *accountability*, który tłumaczyć należałoby jako odpowiedzialność społeczną urzędników publicznych. Egzekwowanie społecznej odpowiedzialności stanowi bowiem immanentny element każdego systemu demokratycznego, będąc także jedną z pierwotnych i do dziś często wykorzystywanych funkcji ewaluacji.

<sup>9</sup> Zobacz więcej:

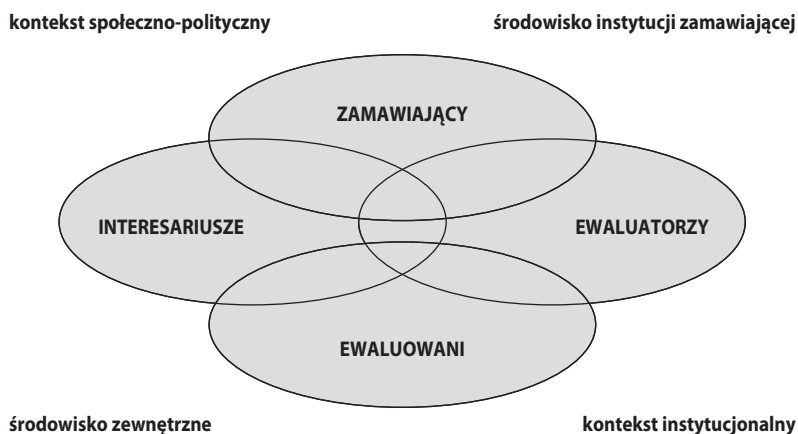
- 1) European Social Survey, <http://www.ads.org.pl/pobieranie-zbioru-danych.php?id=15>;
- 2) Kapitał intelektualny i społeczny w Polsce – raport doradców premiera, [http://pliki.innowacyjnosc.gpw.pl/Kapital\\_Intelektualny\\_Polski.pdf](http://pliki.innowacyjnosc.gpw.pl/Kapital_Intelektualny_Polski.pdf);
- 3) World Values Survey, <http://www.worldvaluessurvey.org/>.

cji jego wyników, nie mówiąc już o praktycznym wykorzystaniu płynących z badania rekomendacji.

Pojęcie kapitału społecznego dostarcza w tym przypadku ciekawej perspektywy analitycznej ze względu na akcentowaną wagę relacji zaufania w jego tworzeniu, podtrzymywaniu i pomnażaniu kapitału. Wykorzystując tę opcję, stwierdzić można, iż kapitał społeczny (zwłaszcza w postaci relacji opartych na zaufaniu) stanowi zasób niezbędny do prowadzeniu udanych ewaluacji. Jego brak bowiem nawet przy najwyższych kompetencjach ewaluatorów oraz wystarczających budżetach zawsze będzie stanowił ograniczenie, którego nie sposób pokonać, wzmacniając inne dostępne zasoby.

Słabość relacji zaufania świadczy bez wątpienia o braku niezbędnych warunków dla prowadzenia udanych ewaluacji, jednakże nie jest argumentem na rzecz zaniechania wszelkich badań ewaluacyjnych. Według retoryki chętnie stosowanej przez popularyzatorów ewaluacja stanowić ma dla instytucji publicznych oraz instytucji społeczeństwa obywatelskiego „zaproszenie do rozwoju”. Doświadczenia krajów, dla których badania ewaluacyjne stanowią naturalny element zarządzania sprawami publicznymi, wskazują, iż hasło to znajduje swoje empiryczne potwierdzenie. Ich adaptacja do krajowych warunków jest możliwa, przy założeniu, iż niedobory kapitału społecznego zostaną uzupełnione, co jest wyzwaniem bez wątpienia trudnym, lecz osiągalnym.

W celu przekształcenia kapitału społecznego zarówno w atrybut podmiotów zajmujących się ewaluacją, jak i grunt dla efektywnej ewaluacji, konieczny jest zwrot w myśleniu. Zwrot, który nauczy organizacje tego, jak pracować nad relacjami społecznymi występującymi w złożonym otoczeniu organizacji i tym samym otoczeniu projektów ewaluacyjnych.



Rysunek 2. Środowisko i kontekst ewaluacji

Źródło: Opracowanie własne.

Podstawą tego myślenia powinny być kluczowe czynniki pomagające organizacjom zrozumieć skomplikowane otoczenie, a w konsekwencji wyciągnąć największe możliwe korzyści z rozwiniętego kapitału społecznego<sup>10</sup>:

- **Wielowymiarowość spojrzenia.** Organizacje powinny być beneficjentami swoich zachowań organizacyjnych, a jednocześnie rozumieć założenia, wartości i idee, które podtrzymują i rozwijają owe zachowania.
- **Dynamiczne związki.** Tworzenie map relacji zachodzących pomiędzy uczestnikami organizacji pozwala na dostrzeżenie wyłaniających się trendów oraz na zrozumienie istniejących związków, kontaktów i współzależności w organizacji.
- **Wspólnie podzielane wzorce.** Radzenie sobie ze złożonością jest łatwiejsze w chwili, gdy uczestnicy organizacji dzielą sposoby myślenia, zachowania czy współdziałania.
- **Praca grupowa.** Jednostki pozostawione samym sobie nie są w stanie pojąć złożoności czy całości systemu. Jego zrozumienie wymaga bowiem grupowego wysiłku. Uczucie się w „ekonomii wiedzy” jest procesem polegającym na współdziałaniu i ogromnych nakładach kapitału społecznego.

Akumulacja kapitału społecznego, w interesującej nas postaci, wymaga długofalowej i systematycznej pracy nad relacjami wewnątrzinstytucjonalnymi. Istotna rola przypada na tym polu jednostkom ewaluacyjnym ulokowanym obecnie w strukturach większości instytucji polskiej administracji szczebla centralnego i regionalnego. Działania tych komórek powinny mieć charakter szeroko zakrojonej animacji, mającej na celu budowanie trwałych relacji wewnątrz instytucji oraz z kluczowymi partnerami zewnętrznymi.

Ujmując rzecz w kategoriach nieco bardziej praktycznych, budowanie relacji zaufania wobec ewaluacji powinno obejmować cały katalog działań, które mogą potencjalnie wnieść wkład w „odczarowywanie” ewaluacji, czyli przewyżczenie poglądu o jej czysto kontrolnym, rozliczeniowym charakterze.

Gromadzenie kapitału społecznego na rzecz badań ewaluacyjnych wymaga podjęcia wieloaspektowych działań na różnych poziomach instytucjonalnej hierarchii, m.in. w takich obszarach, jak budowa wspomnianego już wcześniej zaufania oraz tworzenie kultury opartej na wiedzy, współdziałaniu, aktywności, innowacyjności, eksperymentowaniu, uczeniu się na błędach i akceptowaniu porażki. Do przykładowych, najprostszych działań zaliczyć należy:

- szkolenia/opracowania z zakresu podstaw ewaluacji, adresowane odrębnie dla osób zarządzających daną instytucją oraz dla innych osób kluczowych

<sup>10</sup> Allee V., „Return on Knowledge”, *Executive Excellence* 2000, nr 9. Provo, Birmingham 2000, s. 2, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, [http://www.zti.com.pl/instytut/pp/referaty/ref42\\_full.html](http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html).

będących rzeczywistymi lub potencjalnymi współpracownikami komórki ewaluacyjnej;

- rozwijanie kompetencji po stronie pracowników komórek ewaluacyjnych, w takich obszarach, jak: 1) weryfikacja rzetelności prac ewaluatorów zewnętrznych oraz projektowanie i metodologia badań ewaluacyjnych, 2) kompetencje interpersonalne oraz menedżerskie niezbędne do prawidłowego zarządzania złożonym procesem ewaluacji oraz animacji procesów: a) tworzących trwałe relacje zaufania, b) komunikacji, c) budowania sieci współpracy;
- konstruowanie dynamicznych sieci współpracy z instytucjami naukowymi, komórkami ewaluacyjnym oraz stowarzyszeniami branżowymi, zorientowanych na podobną grupę problemów instytucjonalnych lub ten sam obszar zagadnień badawczych;
- budowanie dynamicznych sieci współpracy wewnątrz instytucji: tworzenie relacji zaufania poprzez angażowanie osób odpowiadających za oceniany obszar w procesy konsultowania oraz dystrybucji wyników ewaluacji;
- tworzenie platform wymiany informacji i doświadczeń w obszarach projektów ewaluacyjnych (konferencje, spotkania, seminaria branżowe, ale również strony internetowe, internetowe bazy projektów itp.);
- poświęcenie wystarczającej ilości czasu na wymianę informacji, ułatwiających precyzyjnie sformułowanie i odebranie sprzężeń zwrotnych od każdego z partnerów procesu ewaluacji (działania w tym obszarze pozwalają zmniejszyć niezrozumienie, niewłaściwą interpretację, zły odbiór przekazu oraz wzmocnić współpracę).

Zarysowany powyżej katalog działań jest jedynie zestawem przykładowych inicjatyw służących budowaniu relacji zaufania umożliwiających akumulację, niezbędnego dla udanych ewaluacji, kapitału społecznego. Paleta działań wpisujących się w ten obszar pozostaje bardzo szeroka, a ich dobór warunkowy powinien być każdorazowo poprzedzony pogłębioną analizą specyfiki danego przypadku. Podkreślenia wymaga jednak, iż budowanie relacji wewnątrz instytucji zaliczyć należy do pełnoprawnych działań na rzecz tworzenia tak często przywoływanego „potencjału ewaluacyjnego” administracji publicznej. Podobieństwo niektórych działań do technik typowych dla praktyk marketingowych, zarządzania zasobami ludzkimi czy PR (*public relations*) nie jest w żadnym stopniu deprecjonujące, przy zastrzeżeniu, iż zachowana zostanie wierność kanonom profesjonalizmu. Badania ewaluacyjne nie tracą wówczas nic ze swej wiarygodności, zyskują natomiast bardzo wiele, jeśli chodzi o ich użyteczność i dostosowanie do potrzeb odbiorców.



## Kapitał społeczny jako wytwór udanej ewaluacji

Kapitał społeczny rozumiany jako warunek konieczny dla udanej ewaluacji jest zagadnieniem mniej znanym i z tego względu poświęciliśmy mu nieco więcej uwagi. Zestawienie pojęcia kapitału społecznego z ewaluacją koncentruje się bowiem na drugiej ze wskazanych wcześniej ról, czyli pomnażaniu kapitału społecznego poprzez wyniki ewaluacji. Jest to nurt rozważań wywodzący się z popularnego konceptu ewaluacji demokratycznej (partycypacyjnej). Wskazywaliśmy uprzednio, iż angażowanie partnerów instytucjonalnych w procesy ewaluacyjne jest praktyką służącą budowaniu relacji zaufania, niezbędnych jako warunek wyjściowy.

W drugiej z omawianych przez nas ról kapitał społeczny znajduje zastosowania w środowisku cechującym się wystarczającymi zasobami kapitału społecznego na etapie „wyjściowym”. Demokracja partycypacyjna stanowi bowiem formę otwarcia instytucji na partnerów zewnętrznych poprzez budowanie zaufania, współpracę oraz komunikację. W modelowym ujęciu partnerzy zewnętrzni zapraszani są do aktywnego udziału w procesie ewaluacji.

Bezpośrednie zaangażowanie partnerów społecznych w procesy zarządzania przedsięwzięciami publicznymi, a ewaluacja stanowi składową tych procesów, przekłada się na wzrost kapitału społecznego poprzez wzmocnienie znaczenia instytucji społeczeństwa obywatelskiego. Jak wskazuje Robert Picciotto, ewaluacja „pozwała przewyciężyć asymetrię między obywatelami jako mocodawcami i państwem jako agentem (...), a także pomaga rozwiązywać dylematy kolektywnego działania i rozwijać współpracę między obywatelami i władzą”<sup>11</sup>.

Wymiar, w którym ewaluacja przyczynia się do akumulacji kapitału społecznego, wiąże się z wątkiem ewaluacji demokratycznej, tj. demokratyzującymi funkcjami ewaluacji. Kapitał społeczny jest wówczas wzmacniany poprzez dostarczanie informacji o efektach polityk oraz wydatkowanych funduszach publicznych. Włączenie wyników ewaluacji w debatę publiczną aktywizuje wówczas zainteresowanie społeczeństwa do podejmowania wspólnych działań ukierunkowanych na realizację interesów ogółu. Jak zauważa Picciotto: „... ilekroć rynek nie jest skuteczny w alokacji rzadko występujących dóbr (a w przypadku funduszy publicznych zasadniczo mamy do czynienia z taką sytuacją – przyp. autorów), z hierarchią związane jest uczestnictwo. W takiej sytuacji ewaluacja wymusza transparentność i minimalizuje «jazdę na gapę». Współuczestnictwo w ewaluacji jest związane z decentralizacją i prywatyzacją. Powstanie w skali globalnej aktywnego społeczeństwa obywatelskiego wykreowało nowy typ ewaluacji bliski funk-

<sup>11</sup> Picciotto R., „Ekonomia ewaluacji”, *Państwo i Rynek* 2006, nr 2; zob. również Kapitał intelektualny i społeczny w Polsce – raport doradców premiera: [http://pliki.innowacyjnosc.gpw.pl/Kapital\\_Intelektualny\\_Polski.pdf](http://pliki.innowacyjnosc.gpw.pl/Kapital_Intelektualny_Polski.pdf).

cji społecznego adwokata. (...) Otwarcie decydentów polityki na opinie pracowników, klientów sprawia, że łatwiejsze jest porozumienie, co do przesłanek i słabych stron programów publicznych. Niezależna ewaluacja pomaga zdemaskować fałszywy rozgłos grup interesu bądź umożliwi poparcie ich racji”<sup>12</sup>. W ten sposób poprzez dostęp do informacji, przejrzystość działań i jawne efekty polityk publicznych, aktywne społeczeństwo może redefiniować swoje poglądy, lepiej rozumieć istotę problemów i tym samym aktywnie oddziaływać na politykę publiczną.

Głęboko zakorzeniona w społecznym kapitale ewaluacja jest wiedzą o wiedzy. Promując transparentność i odpowiedzialność, przyczynia się ona do wzrostu jakości publicznych usług i stanowi przeciwwagę dla wysoko pojętych interesów. Nic tak efektywnie nie osłabia roszczeń polityków i biurokratów jak faktyczna ocena rezultatów podejmowanych przez nich działań. Ewaluacja przynosi zatem korzyści w postaci poprawy efektywności wydatkowania i wrażliwości organizacji w zakresie reagowania na potrzeby klientów<sup>13</sup>.

## **Kapitał społeczny jako atrybut budowania potencjału ewaluacyjnego**

Idea kapitału społecznego nabiera szczególnego znaczenia w tworzącej się nowej wizji sprawnej, dynamicznej administracji publicznej, nastawionej na efektywne rozwiązywanie problemów publicznych, gdzie podstawowym zasobem strategicznym jest wiedza, a o wartości instytucji decyduje jej kapitał intelektualny. Idea ta staje się kluczem do rozwiązywania problemów zarówno wewnątrzorganizacyjnych, jak i więzi ze „światem zewnętrznym”. W. Dyduch, M. Szczepankiewicz, E. Szczepankiewicz wskazują, że podobnie jak inne formy kapitału, kapitał społeczny służy podwyższeniu efektywności funkcjonowania i rozwoju organizacji głównie poprzez ułatwianie współdziałania pomiędzy uczestnikami<sup>14</sup>.

W przypadku ewaluacji koncepcja ta jest szczególnie istotna. Przede wszystkim dlatego, że obecnie na polskim rynku, mamy do czynienia: 1) ze stosunkowo młodym, cały czas dynamicznie kształtującym się rynkiem popytu/podaży badań ewaluacyjnych, 2) relatywnie nowymi strukturami insty-

<sup>12</sup> Picciotto R., op. cit.

<sup>13</sup> Ibidem.

<sup>14</sup> Dyduch W., Szczepankiewicz M., Szczepankiewicz E., „Kapitał społeczny podstawą zdobywania przewagi konkurencyjnej w Nowej Ekonomii”, w: M. Moszkowicz (red.), *Strategie i konkurencyjność przedsiębiorstw po dziesięciu latach transformacji*, Politechnika Wrocławska, Polanica Zdrój 2001, s. 193–202.

tucjonalnymi, powołanymi do zadań związanych z oceną interwencji publicznych, 3) deficytami zasobów bazowych (ludzkich, finansowych, informacyjnych, zaplecza strukturalnego, kapitału intelektualnego, kapitału kulturowego itd.), a co za tym idzie 4) niskim potencjałem ewaluacyjnym.

Przyjęte tutaj ujęcie kapitału społecznego oznacza traktowanie go jako komponentu umiejętności współdziałania i współpracy jednostek ludzkich w ramach grup społecznych, organizacji i instytucji społecznych różnego typu (nie tylko gospodarczych) dla realizacji wspólnych celów. Umiejętne wykorzystanie kapitału społecznego, czyli wiedzy i umiejętności zawartej w jednostkach, szerszych i węższych grupach społecznych, jest źródłem przyszłej satysfakcji, wzmacniania efektywności działań organizacji w obszarze ewaluacji oraz katalizatorem rozwoju, innowacyjności zróżnicowanych podejść ewaluacyjnych.

Literatura przedmiotu wyraźnie wskazuje, że organizacje dysponujące dużą ilością kapitału społecznego są bardziej skłonne do przedsiębiorczego podejmowania ryzyka i wprowadzania nowych działalności niż organizacje, które ograniczają spontaniczne tworzenie się grup. Dzieje się tak, gdyż kapitał społeczny jest formą struktury społecznej w przedsiębiorstwie i wzmacnia proaktywne zachowanie się ludzi wewnątrz tej struktury<sup>15</sup>. Rozumienie kapitału społecznego jako zaangażowania się jednostek w budowanie sieci powiązań stanowi przesłankę do wyjaśnienia skłonności do innowacji<sup>16</sup>. Powiązania społeczne są bowiem przyczyną powstania wzajemności działań w sieci, a ta z kolei jest podłożem do rozwoju zaufania, zwiększonej chęci podejmowania ryzyka i działań innowacyjnych. Sieci ułatwiają koordynację i komunikację oraz współpracę, a także wzmacniają reputację uczestników, co pozwala na grupowe rozwiązywanie problemów<sup>17</sup>. Sukces organizacji osadzonych w otoczeniu społecznym zależy od umiejętnie zbudowanej sieci kontaktów społecznych, która wspiera podejmowane innowacyjne działania<sup>18</sup>.

Orientacja na szybkie budowanie zdolności i umiejętności, umożliwiających realizację postawionych przed ewaluacją zadań, wymaga metodycznego budowania zaplecza ewaluacyjnego. W obecnych warunkach wystarczającego kapitału finansowego i rzeczowego oraz deficytów rynkowych i instytucjonalnych w zakresie wiedzy, informacji, zasobów ludzkich, wydaje się, że jest

<sup>15</sup> Bratnicki M., „Dylematy i pułapki współczesnego zarządzania”, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, [http://www.zti.com.pl/institut/pp/referaty/ref42\\_full.html](http://www.zti.com.pl/institut/pp/referaty/ref42_full.html).

<sup>16</sup> Bullen P., Onyx J., *Measuring Social Capital in Five Communities in NSW*. Internet, materiały ze strony <http://www.mapl.com.au>, w: W. Dyduch, op. cit.

<sup>17</sup> Sirianni C., Friedland L., *Social Capital and Civic Innovation: Learning and Capacity Building from the 1960s to the 1990s*, Internet, materiały ze strony: [http://www.cpn.org/sections/new\\_citizenship/theory.html](http://www.cpn.org/sections/new_citizenship/theory.html), w: W. Dyduch, op. cit.

<sup>18</sup> Steier L., Greenwood R., „Entrepreneurship and the Evolution of Angel Financial Networks”, *Organization Studies* 2000, nr 21, EGOS, s.165, w: W. Dyduch, op. cit.

to możliwe jedynie poprzez<sup>19</sup> – uczestnictwo w sieciach powiązań instytucjonalnych oraz projektowych opartych na wzajemności, współdziałaniu, współpracy, zaufaniu, bazujących na kapitale społecznym.

Nie ulega wątpliwości, że stanowi on podstawę korzyści i efektywności przedsiębiorczych organizacji, między innymi dlatego, że pomaga zmniejszyć koszty poszukiwania i przepływu informacji. Sieci kontaktów społecznych decydują o tym, kto dowiaduje się pierwszy, np. o nowościach metodologicznych i technicznych, nowych zasobach źródeł wiedzy, działaniach wykonawców, nowych podmiotach, prowadzonych projektach badawczych i analitycznych, kluczowych ekspertach w danej dziedzinie itp. Ponadto dzięki możliwości szybkiego budowania nowych i efektywnych połączeń, pozwalają maksymalizować wykorzystanie dostępnych zasobów (przez jak największą liczbę uczestników sieci), rozwijać innowacyjne podejścia i rozwiązania oraz obniżyć koszty transakcyjne (koordynacji i kooperacji).

Wprawdzie mieszanka sieci, wzajemności, zaufania, norm jest bez wątpienia szansą dla stworzenia sprawnie działającego „środowiska ewaluacyjnego”, dysponującego wspólnie podzielanymi i wypracowywanymi zasobami (w tym najważniejszym – informacją, wiedzą i zapleczem intelektualnym), do którego każdy uczestnik sieci będzie miał dostęp, jednak stworzenie nowych połączeń sieciowych nie jest zadaniem prostym. Przede wszystkim dlatego, że jednostki organizacyjne nie zawsze są gotowe do tego, by stać się uczestnikiem sieci bądź nie zawsze chcą dzielić się swoimi zasobami, które mogą stanowić źródło ich przewagi.

Budowanie trwałych relacji jest zatem zależne od zmiany myślenia oraz aktywności kluczowych liderów – świadomych animatorów szeregu systematycznych działań na rzecz budowania szeroko pojętego potencjału ewaluacyjnego.

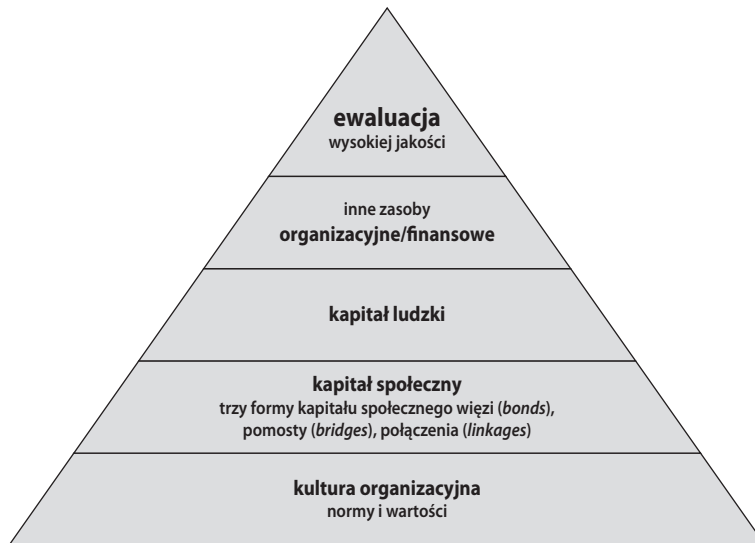
## Zakończenie

Doświadczone podmioty zlecające oraz realizujące ewaluację wiedzą, że sukces ewaluacji, a więc jej celowe, rzetelne, sprawne i efektywne przeprowadzenie i wykorzystanie jest uzależnione przede wszystkim od zasobów niematerialnych.

Kapitał społeczny stanowi bez wątpienia jeden z fundamentalnych zasobów, koniecznych z jednej strony do wykonania dobrej ewaluacji, z drugiej strony do „konsumpcji wyników”. Biorąc jednak, w świetle kolejnych edycji badań systemu wartości społecznych w Polsce, pod uwagę deficyty – występujące w takich obszarach, jak zaufanie, prospołeczna aktywność obywateli, umiejętność i chęć kooperacji itp., a więc deficyt kapitału społecznego – czer-

<sup>19</sup> Porównaj: Bullen P., Onyx J., op. cit.

panie z niego obecnie jest dla procesów ewaluacji mocno ograniczone. Sytuacja ta stanowi wyzwanie nie tylko dla ewaluatorów, instytucji publicznych, organizacji pozarządowych, ale również dla samego społeczeństwa.



Rysunek 3. Podstawy ewaluacji wysokiej jakości.

Źródło: Opracowanie własne.

Mamy nadzieję, że powyższe omówienie ewaluacji w kontekście kapitału społecznego ułatwi Czytelnikowi lepsze poruszanie się w złożonych relacjach i otoczeniu projektów ewaluacyjnych. Zdajemy sobie w pełni sprawę, że niniejszy tekst może być jedynie wstępem do dalszej dyskusji. Dyskusji, która już niebawem w środowisku ewaluacyjnym rozpocznie się na nowo przy okazji tegorocznej konferencji Europejskiego Towarzystwa Ewaluacyjnego w Lizbonie.

---

**Agnieszka Haber** – absolwentka Instytutu Socjologii oraz Wydziału Stosowanych Nauk Społecznych Uniwersytetu Warszawskiego. Od 2004 roku kieruje Sekcją Ewaluacji w Polskiej Agencji Rozwoju Przedsiębiorczości. Jest odpowiedzialna za realizację ewaluacji programów przedakcesyjnych oraz strukturalnych w obszarach rozwoju sektora MSP, Rozwój Zasobów Ludzkich oraz projektów infrastrukturalnych wdrażanych przez Agencję. Problematyką badań społecznych zajmuje się od ponad 9 lat, zaś obszarem ewaluacji i zarządzania interwencjami publicznymi od 6 lat. Od

roku 2005 organizator międzynarodowej cyklicznej Konferencji Ewaluacyjnej odbywającej się pod auspicjami PARP i MRR. Od roku 2005 uczestnik grup ds. ewaluacji kolejnych programów: EFS, PO IG, PO KL, PO RPW. Od 4 lat prowadzi wykłady i warsztaty z ewaluacji programów o charakterze społeczno-gospodarczym. Jest autorką szeregu publikacji i opracowań z dziedziny ewaluacji programów sektora publicznego.

**Maciej Szałaj** – absolwent Instytutu Socjologii Uniwersytetu Warszawskiego. Pracuje w Sekcji Ewaluacji PARP. Problematyką badań ewaluacyjnych zajmuje się od 5 lat. Specjalizuje się w ewaluacji programów ukierunkowanych na rozwój kadr gospodarki oraz wspierania potencjału adaptacyjnego przedsiębiorstw. Kierownik projektów badawczych oraz współautor ich koncepcji i metodologii. Posiada doświadczenie w ewaluacji programów przedakcesyjnych i funduszy strukturalnych. Członek Grupy Sterującej Ewaluacją SPO RZL oraz Grupy ds. Ewaluacji Priorytetu I i II PO KL. Obecnie zaangażowany w opracowywanie kompleksowego systemu pomiaru efektów Działania 2.1, 2.2 PO KL, mającego stanowić podstawę dla przyszłych ewaluacji tego obszaru. Autor opracowań i artykułów poświęconych efektom przedsięwzięć publicznych finansowanych ze środków wspólnotowych oraz zagadnieniom teorii i praktyki badań ewaluacyjnych. Od czerwca 2008 r. członek Zarządu Polskiego Towarzystwa Ewaluacyjnego.

## Bibliografia

- Allee V., „Return on Knowledge”, *Executive Excellence* 2000, nr 9. Provo, Birmingham 2000, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, [http://www.zti.com.pl/instytut/pp/referaty/ref42\\_full.html](http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html).
- Bourdieu P., „The Forms of Capital”, w: J.G. Richardson (red.), *Handbook of Theory and Research for the Social Education*, Greenwood Pres, New York 1986.
- Bratnicki M., „Dylematy i pułapki współczesnego zarządzania”, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, [http://www.zti.com.pl/instytut/pp/referaty/ref42\\_full.html](http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html).
- Bullen P., Onyx J., *Measuring Social Capital in Five Communities in NSW*. Internet, materiały ze strony <http://www.mapl.com.au>, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, [http://www.zti.com.pl/instytut/pp/referaty/ref42\\_full.html](http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html).
- Coleman J.S., „Social Capital in the Creation of Human Capital”, *American Journal of Sociology*, nr 94.
- Dyduch W., Szczepankiewicz M., Szczepankiewicz E., „Kapitał społeczny podstawą zdobywania przewagi konkurencyjnej w Nowej Ekonomii”, w: M. Moszkowicz (red.), *Strategie i konkurencyjność przedsiębiorstw po dziesięciu latach transformacji*, Politechnika Wroclawska, Polanica Zdrój 2001.
- European Social Survey, <http://www.ads.org.pl/pobieranie-zbioru-danych.php?id=15>.
- Fukuyama F., „Social capital, civil society and development”, *Third World Quarterly* 2001, nr 1.
- Fukuyama F., *Zaufanie. Kapitał społeczny a droga do dobrobytu*, PWN, Warszawa 1997.
- Kapitał intelektualny i społeczny w Polsce – raport doradców premiera, [http://pliki.innowacyjnosc.gpw.pl/Kapital\\_Intelektualny\\_Polski.pdf](http://pliki.innowacyjnosc.gpw.pl/Kapital_Intelektualny_Polski.pdf).

- Kazimierczak T., „Kapitał społeczny a rozwój społeczno-ekonomiczny – przegląd pojęć”, w: T. Kazimierczak, M. Rymśza (red.), *Kapitał społeczny. Ekonomia społeczna*, Instytut Spraw Publicznych, Warszawa 2007.
- Picciotto R., „Ekonomia ewaluacji”, *Państwo i Rynek* 2006, nr 2.
- Putnam R.D., *Demokracja w działaniu: Tradycje obywatelskie we współczesnych Włoszech*, Znak, Kraków 1995.
- Sirianni C., Friedland L., *Social Capital and Civic Innovation: Learning and Capacity Building from the 1960s to the 1990s*, Internet, materiały ze strony: [http://www.cpn.org/sections/new\\_citizenship/theory.html](http://www.cpn.org/sections/new_citizenship/theory.html), w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, [http://www.zti.com.pl/instytut/pp/referaty/ref42\\_full.html](http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html).
- Steier L., Greenwood R., „Entrepreneurship and the Evolution of Angel Financial Networks”, *Organization Studies* 2000, nr 21, EGOS, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, [http://www.zti.com.pl/instytut/pp/referaty/ref42\\_full.html](http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html).
- World Values Survey, <http://www.worldvaluessurvey.org/>.

Część II

**Wybrane zagadnienia  
warsztatu badań ewaluacyjnych**





*Aleksander Surdej*

## **Metoda analizy kosztów i korzyści: imperatyw efektywności w politykach publicznych**

### **Uwagi wstępne**

Pojęcie analizy kosztów i korzyści szybko wchodzi do języka polskiej administracji i polityki. Dzieje się tak z co najmniej dwóch powodów. Po pierwsze, pod wpływem Komisji Europejskiej, która wymaga, aby analiza kosztów i korzyści w rozmaitych wersjach poprzedzała wybór projektów zakwalifikowanych do finansowania ze środków funduszy strukturalnych lub *ex post* identyfikowała skutki zrealizowanych przedsięwzięć. Po drugie, w okresie presji na jakość rządzenia i obniżenie wydatków rządowych analiza kosztów i korzyści jawi się jako naturalna podstawa decyzji publicznych. Zalecenie, aby korzyści przewyższały koszty, wydaje się zasadą nie do podważenia, tak na gruncie zdrowego rozsądku, jak i w politycznej retoryce.

Szybkość, z jaką metoda analizy kosztów i korzyści upowszechnia się w Polsce, rodzi obawę, że jest ona stosowana w sposób zrytualizowany – raczej dla uzasadnienia dokonywanych wyborów niż dla ich rzetelnego przygotowania. Podejrzenie to bierze się z doświadczeń państw, w których metoda analizy kosztów i korzyści została wprowadzona do praktyki rządzenia już kilkadziesiąt lat temu, a jednak do dzisiaj trwają spory o sposób jej wdrożenia i konsekwencje.

W opracowaniu tym spróbuję dokonać krytycznej rekonstrukcji podstawowych elementów metody analizy kosztów i korzyści nie po to, aby podważać zasadność jej stosowania, lecz by zidentyfikować warunki służące lepszemu jej wykorzystaniu w decyzjach i politykach publicznych.

### **Rozwój historyczny i rozwój podstaw teoretycznych**

Początki analizy kosztów i korzyści sięgają połowy XIX wieku i można je znaleźć w pracach francuskiego inżyniera i ekonomisty Julesa Dupuit (1844, 1853). Pisząc o projektach infrastrukturalnych, takich jak budowa dróg i mostów, twierdził, że ich powstanie ma sens wtedy, gdy płynące z nich korzyści

przewyższają koszty. Owa teza brzmiała oryginalnie w zderzeniu z dominującymi poglądami tamtej epoki, a i dzisiaj nie jest powszechnie akceptowana.

W praktyce politycznej zasada porównywania kosztów i korzyści pojawiła się po raz pierwszy w Stanach Zjednoczonych, gdzie w 1936 roku w Ustawie o kontroli przeciwpowodziowej (*Flood Control Act*) Kongres zalecił, aby projekty dotyczące gospodarowania wodą realizować ze środków publicznych jedynie wtedy, gdy „korzyści przekraczają szacowane koszty”. Prace nad sformalizowaniem metody analizy kosztów i korzyści kontynuowano w Stanach Zjednoczonych po II wojnie światowej w związku z przedsięwzięciami w zakresie gospodarki wodnej. W 1950 roku został opublikowany pierwszy przewodnik do oceny kosztów i korzyści przy projektach wodnych. Pod koniec lat 50. wskazówki do oceny kosztów i korzyści przy projektach zbrojeniowych opracowali analitycy Rand Corporation. W latach 60. pojawiły się zastosowania analizy kosztów i korzyści do oceny programów społecznych, a w latach 70. do oceny przedsięwzięć w zakresie ochrony środowiska. W latach 80. z inicjatywy Ronalda Reagana wprowadzono analizę kosztów i korzyści do oceny polityk regulacyjnych i od tamtego czasu można mówić, że ta analiza funkcjonuje zarówno w odniesieniu do oceny poszczególnych projektów finansowanych ze środków publicznych, jak i publicznych polityk.

W Europie pierwsze zastosowania analizy kosztów i korzyści pojawiły się w latach 60. – w Wielkiej Brytanii w odniesieniu do inwestycji transportowych (budowa dróg i lotnisk). Za Wielką Brytanią nie podążał jednak żaden inny kraj europejski i było tak aż do lat 90., kiedy to za sprawą Komisji Europejskiej analiza kosztów i korzyści stała się wyrazem dążenia do wzrostu efektywności programów publicznych i poprawy jakości rządzenia. W przeciwieństwie do Stanów Zjednoczonych w większości państw członkowskich UE analiza kosztów i korzyści nie jest dobrze utrwalonym narzędziem rządzenia, a ogólniejsze przyczyny tego stanu rzeczy z pewnością mają związek z czynnikami kulturowymi oraz tradycją filozoficzną i prawną.

W tle upowszechniania analizy kosztów i korzyści trwały prace teoretyków ekonomii, którzy kładli podstawy pod dzisiejsze rozumienie zalet i wad tej metody. Najważniejszym osiągnięciem w rozwoju podstaw teoretycznych okazało się sformułowanie przez Vilfredo Pareto zasady efektywności alokacji (optimum Pareto), zgodnie z którą alokacja jest efektywna, jeśli nie można poprawić położenia żadnego elementu bez pogorszenia położenia jakiegoś innego elementu. Zasada ta pozostaje normatywnym punktem odniesienia dla analizy polityki publicznej, chociaż jest krytykowana za nierealistycznie wysokie wymagania. Na ogół każdy projekt publiczny wiąże się z korzyściami dla jednych i stratami dla drugich. Jeśli władze publiczne nie mogą zbudować drogi bez wykupienia działek i wysiedlenia części ludności, to rygorystyczne stosowanie zasady efektywności według Pareto musi prowadzić do zaniechania realizacji projektu, nawet gdyby relacja liczby beneficjentów do liczby ponoszących straty była nadzwyczajnie wysoka. Dopiero wprowadzona przez

Nicholasa Kaldora<sup>1</sup> modyfikacja zasady optimum Pareto umożliwiła wzrost jej praktycznego zastosowania. Nowe kryterium efektywności, znane obecnie jako kryterium Kaldora–Hicksa, głosi, że przedsięwzięcie publiczne jest uzasadnione wtedy, gdy tworzone przezeń korzyści społeczne przewyższają społeczne koszty. W rezultacie pojawia się możliwość, że odnoszący korzyści zrekompensują straty ponoszących koszty i wciąż pozostanie im pewna nadwyżka. Należy zwrócić uwagę, że kryterium Kaldora–Hicksa mówi o „potencjalnej kompensacie”, a nie o jej faktycznym zaistnieniu. Istotnie, jeśli interesują nas wyłącznie skutki przedsięwzięcia, to takie ograniczenie jest najzupełniej uzasadnione. Kryterium Kaldora–Hicksa w pełni zadowala tych, którzy analizę kosztów i korzyści postrzegają wyłącznie jako narzędzie poszukiwania wzrostu efektywności projektów i polityk publicznych.

### **Analiza kosztów i korzyści a poszukiwania teoretycznych podstaw określania społecznego dobrobytu**

Przywołane nazwiska teoretyków ekonomii wskazują, że analiza kosztów i korzyści jest ściśle związana z ekonomią neoklasyczną i tym jej nurtem badawczym, który nazywany jest ekonomiką dobrobytu<sup>2</sup>. Niektóre z założeń ekonomii neoklasycznej nie budzą większych kontrowersji, inne natomiast są przedmiotem dyskusji. Powszechna zgoda istnieje co do tego, że to, co określamy korzyściami i kosztami społecznymi, sprowadza się do kosztów i korzyści poszczególnych jednostek. Analiza skutków projektu lub polityki powinna więc brać pod uwagę skutki dla wielu (potencjalnie wszystkich) jednostek, których mogą one dotyczyć. Warto zauważyć, że zasada ta jest bardzo demokratyczna, gdyż nakazuje, aby analiza brała pod uwagę interesy jednostek, które w procesie politycznym mogą być niedostatecznie reprezentowane. Od stwierdzenia o redukowalności kosztów i korzyści społecznych do kosztów i korzyści dla jednostek tylko krok do pytania, jak je identyfikować i porównywać. Rodzi się więc problem oszacowania tak kosztów, jak i w szczególności korzyści, oraz zagadnienie możliwości i podstaw międzyjednostkowego porównania. Analiza neoklasyczna posługuje się pojęciem użyteczności na określenie korzyści odczuwanych przez jednostkę. Użyteczność może być pochodną uzyskanego dochodu pieniężnego, jak i innych czynników (w tym wrażeń estetycznych czy odczuć etycznych). Założenia o charakterze relacji pomiędzy użytecznością a dochodem i innymi czynnikami są wciąż kontrowersyjne. Jeśli przyjmiemy, w ślad za Alfredem Marshalllem,

<sup>1</sup> Kaldor N., „Welfare propositions of economics and interpersonal comparisons of utility”, *Economic Journal* 1939, vol. 49.

<sup>2</sup> Blaug M., *Teoria Ekonomii*, PWN, Warszawa 1994.

założenie o malejącej krańcowej użyteczności dochodu (wzrost dochodu o jedną jednostkę w przypadku osoby zamożnej prowadzi do mniejszego wzrostu użyteczności niż w przypadku wzrostu dochodu osoby biednej), to analiza kosztów i korzyści pokazuje, że wzrostowi społecznego dobrobytu może służyć redystrybucja dochodu. Założenie Marshalla wprowadza jednakże element wartościujący i jako takie zostało usunięte ze współczesnej analizy ekonomicznej, która zaleca pozostawienie kwestii redystrybucyjnej decyzjom politycznym.

Problem redystrybucji nie jest jednak tak łatwy do pominięcia, jak mogłoby się wydawać niektórym ekonomistom. Gdy bowiem współcześnie analitycy podsumowują łączne korzyści z realizacji jakiegoś projektu publicznego, to odwołują się do miary zwanej gotowością do zapłaty (*willingness to pay*). „Analityk pragnie poznać maksymalną wielkość, którą jednostki byłyby skłonne zapłacić za te skutki polityki, które postrzegają, jako korzystne oraz minimalną wielkość, którą jednostki byłyby skłonne zaakceptować jako rekompensatę za te skutki polityki, które postrzegają jako szkodliwe”<sup>3</sup>. Gotowość do zapłaty nie jest jednak niezależna od poziomu dochodu. Jeśli więc publiczny decydent ma wybrać projekt, którego łączna wartość określona jest sumą gotowości do zapłaty, to częściej podejmowane mogą być projekty preferowane przez ludzi zamożnych. Korzyści społeczne mierzone gotowością do zapłaty zależą bowiem od rozkładu dochodu w społeczeństwie.

O ile problem zróżnicowanego skutku projektów publicznych pozostaje sednem politycznych rozgrywek, o tyle mylne jest przypisywanie pojęciu użyteczności egoistycznych implikacji. Użyteczność jest zakorzeniona w odczuciach jednostki, lecz pojęcie to jest z pewnością do pogodzenia z występowaniem altruistycznej motywacji, zważania na innych, w tym na przyszłe pokolenia i wartości przyrody.

## Podstawowe elementy metody analizy kosztów i korzyści

Metoda analizy kosztów i korzyści jest stosowana do analizy problemów publicznych, czyli dotyczących zbiorowości. Jej celem jest odpowiedź na pytanie, na co przeznaczyć środki publiczne, tak aby zrealizować cele obrane przez zbiorowość i generalnie zwiększyć jej dobrobyt. Decyzje zbiorowości są podobne do decyzji jednostek w tym względzie, że ich prakseologicznym warunkiem koniecznym jest warunek alokacyjnej efektywności. Nie jest to warunek jedyny, gdyż w przypadku decyzji dotyczących problemów zbiorowości uwzględniane są inne cele, takie jak zgodność z zasadą sprawiedliwości, a także możliwość uzyskania politycznego wsparcia.

<sup>3</sup> Haveman R.H., Weimer D.L., „Cost Benefit Analysis”, *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier 2001.

Punktem wyjścia praktycznej analizy kosztów i korzyści zarówno w odniesieniu do projektów, jak i polityk publicznych, jest pytanie, czy realizacja projektu lub wdrożenie polityki przyczyni się do wzrostu dobrobytu zbiorowości. Porównanie obecnej sytuacji zbiorowości (opcja, którą można nazwać „zachowanie *status quo*”) z prawdopodobną sytuacją, która zaistniałaby w wyniku realizacji projektu i polityki, przenosi ciężar dowodu na autorów zamierzenia. Korzyści oraz koszty projektu publicznego są korzyściami i kosztami prognozowanymi. Zestawienie ich oraz analiza staje się w tym przypadku *ex ante* analizą kosztów i korzyści. Wydaje się, że metoda analizy kosztów i korzyści jest niekiedy krytykowana właśnie za wymuszenie starannego rachunku oczekiwanych kosztów i korzyści. Zagrożenie w związku z tym widzą ci, którzy dzięki politycznemu umocowaniu lub umiejętnej presji lobbystycznej są w stanie „uruchomić” wydatki publiczne na projekt lub politykę, mimo że staranna *ex ante* analiza dowodzi (lub mogłaby dowieść), że prywatne korzyści są uzyskiwane kosztem pogorszenia dobrobytu większej zbiorowości. To, co jest wadą z perspektywy partykularnych grup interesu, może być jednak korzyścią z punktu widzenia interesu ogólnego.

### **Czym są społeczne korzyści i koszty i jak mogą być mierzone?**

Zasadą decyzyjną metody analizy kosztów i korzyści jest fakt, że projekt lub polityka jest rekomendowana do realizacji czy wdrożenia wtedy, gdy korzyści przewyższają koszty. Być może najmniej rozumianym założeniem omawianej metody jest to, że koszty i korzyści są pojmowane bardzo szeroko – jako każdy typ użyteczności lub kosztów i każdy wzrost lub spadek dobrobytu.

Centralną pozycją kosztów są najczęściej wydatki z budżetu publicznego wyrażone finansowo. Główną kategorią korzyści nie muszą, i najczęściej nie są, zyski określone w pieniądzu. Rodzi to problem współmierności kosztów i korzyści, który stanie się lepiej zrozumiały, gdy przedstawimy podstawowe kategorie korzyści stosowane w analizie.

Pierwszą kategorią korzyści są skutki obserwowane jako zmiany zachowań ludzi, nazywane inaczej „aktywnym użytkowaniem”. Jeśli więc w wyniku realizacji pewnego projektu zachowane zostaną wartości przyrodnicze jakiegoś obszaru, to część korzyści społecznych przejawia się jako wartość dla osób, które ten teren odwiedzają i podziwiają jego walory przyrodnicze. Ta kategoria korzyści jest zakorzeniona w indywidualnych odczuciach osób, toteż korzyści mogą być mierzone poprzez nakładanie opłat na zwiedzających i w ten sposób identyfikowanie ich gotowości do zapłaty.

Korzyść społeczna może jednak także wynikać z doceniania pewnych stanów rzeczy, które powstają lub są utrzymywane dzięki realizacji projektów czy polityk publicznych, chociaż wartościowanie to nie ma związku z aktywnym użytkowaniem. Kategoria ta nazywana jest „pasywnym użytkowaniem”

i dzieli się na dwie subkategorie korzyści. Pierwszą z nich jest „korzyść z posiadania opcji”, która polega na tym, że ceni się samą ewentualność aktywnego skorzystania z powstałej możliwości. W zakresie ochrony przyrody korzyść z posiadania opcji, to przywiązywanie wagi do możliwości odwiedzenia tego obiektu w przyszłości.

Inny rodzaj korzyści wiąże się z odczuwaniem satysfakcji z samego istnienia jakiegoś obiektu lub utrzymania stanu środowiska naturalnego, chociaż usatysfakcjonowane osoby nie przewidują możliwości aktywnego skorzystania z cenionego dobra. Korzyść tę nazwać można „korzyścią z istnienia” lub „korzyścią z zachowania istnienia”.

Podstawowa trudność pomiaru wiąże się z niemożnością wyrażenia wszystkich korzyści w postaci wielkości pieniężnych. Wskazane wyżej teoretyczne kategorie korzyści, jakkolwiek byłyby przekonujące, muszą zostać sprowadzone do porównywalnej miary. Trudność ta, co należy podkreślić, nie jest wynikiem niematerialności doznań, lecz braku podstaw dla rynkowej wyceny użyteczności. Spróbujmy lepiej wyjaśnić ten problem.

Jeśli w wyniku realizacji projektu publicznego powstają mieszkania, które są za darmo przydzielane potrzebującym, to korzyści społeczne są łatwo mierzalne i wyrażone w pieniądzu, jako suma cen możliwych do uzyskania na rynku za lokale o podobnej jakości, wielkości i położeniu. Jeśli jednak wynikiem projektu lub polityki publicznej są unikalne i niewystępujące na rynku stany lub obiekty, to pojawiają się trudności z wyceną korzyści, które tworzą. Analitycy chcieliby poznać gotowość do zapłaty przez wszystkich obywateli, lecz nie są w stanie takiej informacji w sposób prosty uzyskać, gdyż nie mogą zaobserwować fizycznych czynności zapłaty.

Praktyczne stosowanie metody analizy kosztów i korzyści zależy więc od rozwoju metod, jakie umożliwiają wycenę korzyści z dóbr (w szerokim rozumieniu), które nie są przedmiotem wyceny rynkowej. Jedną z metod rozwijanych w celu wyceny wartości pasywnego użytkowania obiektów przyrodniczych jest sondaż na reprezentatywnej próbie, w którym ludzie pytani są o to, ile skłonni byliby zapłacić, aby obiekt przyrody został zachowany w nienaruszonym stanie. Metoda ta zwana wartościowaniem kontyngentnym (*contingent valuation*) jest jednak krytykowana, gdyż zakłada, że respondent posiada dużą wiedzę na temat ocenianego obiektu, a dodatkowo przypuszczać można, że może on strategicznie podawać fałszywą wycenę. W tym drugim przypadku chodzi o to, że badacz nie wie, czy wysoka wycena skutków projektu i polityki nie jest próbą strategicznego wpływu na decyzję publiczną: werbalne deklaracje wyrażonej w pieniądzu chęci do zapłaty mogą znacznie się różnić od faktycznej gotowości do zapłaty i rzeczywistych korzyści. Uzyskanie przybliżonej zobiektywizowanej wyceny w oparciu o metodę wartościowania kontyngentnego wymaga więc metodologicznego wyrafinowania i czasu na staranne przeprowadzenie analizy.

Inną pośrednią metodą szacowania korzyści jest rekonstrukcja krzywych podaży i popytu oraz wyznaczonych przez nie cen równowagi w sytuacji dóbr, które wprawdzie nie mają swojego rynku, lecz wycenianie ich przez ludzi może być pośrednio wyprowadzone z analizy podaży i popytu na dobra powiązane. I tak, przykładowo, nie ma rynku na hałas lub czyste powietrze, lecz z odmiennych cen mieszkań znajdujących się na obszarach różniących się natężeniem hałasu lub stopniem czystości powietrza wyprowadzić można ceny widmowe (*shadow prices*) hałasu lub czystego powietrza.

Estymacja cen widmowych dokonywana za pomocą tzw. metody hedonicznej jest szczególnie ważna w próbach wyliczenia wartości życia jako parametru dla optymalizacji wielu projektów i polityk publicznych, w tym ochrony środowiska i polityki zdrowotnej. W analizie polityk publicznych wartość życia nie jest kategorią etyczną, lecz parametrem konstruowanym na potrzeby decyzji publicznych poprzez analizę gotowości do zapłaty. Prowadzona jest zgodnie z zasadą, że „najlepszą miarą wartości dobra dla ludzi jest to, ile są gotowi za nie zapłacić”, czy też na podstawie zaspokojonych lub ujawnionych preferencji (*revealed preferences*), w tym poprzez skłonność do płacenia za zmniejszenia ryzyka dla życia i zdrowia (np. poprzez zakup nowych opon samochodowych lub wybór lepiej płatnej pracy, lecz wiążącej się ze zwiększonym ryzykiem zachorowania lub śmierci)<sup>4</sup>.

Wcześniej wspomniałem, że wyliczenie kosztów jest zadaniem łatwiejszym, gdy ograniczamy się do nakładów finansowanych z budżetu. Nie są to oczywiście jedyne koszty, które należy brać pod uwagę. W przypadku śmierci, uszkodzeń ciała oraz wzrostu zachorowań powinny być uwzględniane koszty utraty życia lub pogorszenia jego jakości. Ponadto analityk powinien być świadomy konieczności uwzględnienia tzw. kosztów okazji (*opportunity costs*). Dla realizacji projektów lub polityk publicznych są ponoszone nakłady, a ponieważ zasoby stojące do dyspozycji zbiorowości są ograniczone, to analityk powinien dodatkowo uwzględnić „koszty okazji danego użycia zasobów, którym jest wartość najlepszej dostępnej alternatywy”<sup>5</sup>. Uwzględnianie kosztów okazji jest szczególnie istotne przy analizie decyzji, których konsekwencje nie podlegają wycenie i weryfikacji rynkowej.

Podsumowując tę część rozważań, warto zauważyć, że poszukiwania sposobów rozwiązywania problemu oceny korzyści dóbr (wartości) niewymienianych na rynku prowadzą do metodologicznej konwergencji klasycznych metod badań społecznych (w tym badań sondażowych) oraz mikroekonomii. O ile tradycyjna mikroekonomia negowała wartość werbalnych de-

<sup>4</sup> Bez wnikania w inne filozoficzne i metodologiczne problemy szacowania wartości dla życia warto zauważyć, że potrzeba szacowania takiego parametru nie może być nawet postawiona w polskiej debacie publicznej, choć decydenci systemu ochrony zdrowia wiedzą, że *implicite* posługują się nim w decyzjach dotyczących leczenia różnych kategorii i grup chorób.

<sup>5</sup> Bednarski M., Wilkin J., *Ekonomia dla prawników*, PWN, Warszawa 2003, s. 84.



klaracji, ograniczając analizy do obserwacji aktów kupna i sprzedaży oraz zaagregowanych skutków w postaci zmian cen, to nowoczesna mikroekonomia próbuje zrekonstruować zachowania na podstawie odpowiedzi na pytania ankietowe oraz obserwacji zachowań w społecznych mikroeksperymentach.

### **Analiza kosztów i korzyści w perspektywie międzypokoleniowej**

Trudności z identyfikacją i wyrażeniem w pieniądzu wszystkich kategorii kosztów i korzyści społecznych nie są ostatnią przeszkodą dla poprawnego przeprowadzenia analizy. Koszty i korzyści, co oczywiste, rozkładają się w czasie. Analityk musi ten fakt uwzględnić, a doprowadzenie do porównywalności wartości pieniężnych skutków (kosztów i korzyści), które występują w różnych okresach czasu jest nazywane dyskontowaniem. Na ogół potrzeba dyskontowania nie jest kwestionowana, gdyż zdecydowana większość badaczy zgadza się ze stwierdzeniem, że ludzie cenią przyszłe korzyści mniej niż równe im co do pieniężnej wysokości korzyści teraźniejsze (zgodnie z założeniem malejącej krańcowej użyteczności). Problemem jest wybór wysokości stopy dyskontowej. Jeśli wybrana zostanie wysoka stopa dyskontowa, oznacza ona preferencje dla efektów krótkookresowych, swobodną „niecierpliwość społeczną”, jeśli zaś niska stopa dyskontowa, to oceniający decyduje się na przydanie wyższej wartości projektom o oddalonych w czasie korzyściach.

Stopy dyskontowe nie są całkowicie dowolne. Punktem wyjścia są stopy dyskontowe obserwowane na rynkach kapitałowych. Zwiększanie wartości teraźniejszej netto projektu przy zastosowaniu rynkowej stopy dyskontowej prowadzi do maksymalizacji efektywności tylko wtedy, gdy rynki kapitałowe są bardzo konkurencyjne, a decyzje oszczędzających i inwestujących nie są zniekształcone przez podatki nakładane na oszczędności i dochody z inwestycji. Podatki „wbijają klin” w krańcową ocenę oszczędzających i inwestorów, a analityk nie może w jednoznaczny sposób odpowiedzieć, czy do oceny projektu lub polityki powinien przyjąć stopę zwrotu dla inwestora (pożyczającego) przed opodatkowaniem, czy też stopę zwrotu po opodatkowaniu dla oszczędzającego (udzielającego pożyczki).

Wartość teraźniejsza przyszłych kosztów i korzyści zależy także od trafności przewidywań inflacyjnych. Chociaż rynkowa stopa dyskontowa odzwierciedla oczekiwania inflacyjne, to są obawy co do przyszłego przebiegu inflacji.

Wątpliwości co do wielkości inflacji nie są jedyną przeszkodą, którą powinien uwzględnić analityk. Niepewność może mieć głębszy charakter i dotyczyć realizacji i efektów projektu lub przebiegu i skutków polityki. Technicznym rozwiązaniem problemu jest zastosowanie w analizie ocze-

kiwanych wyników projektu lub polityki, czyli skutków uwzględniających prawdopodobieństwo ich wystąpienia. To rozsądne, zdawałoby się, rozwiązanie niesie ze sobą jednakże poważną trudność w postaci możliwości pojawienia się znacznych rozbieżności przy oszacowaniu korzyści netto, gdyż małe modyfikacje parametrów mogą skutkować znacznymi różnicami w uzyskiwanych wartościach – szczególnie wtedy, gdy brane pod uwagę są wyniki odległe w czasie.

Gdy od problemów określenia rynkowej stopy dyskontowej przechodzimy do postulowanej społecznej stopy dyskontowej, czyli takiej, którą zbiorowość powinna zastosować do oceny teraźniejszej wartości przyszłych skutków, to powstają dwie nowe komplikacje. Po pierwsze, należy uzasadnić, czy społeczna stopa dyskontowa powinna się różnić od rynkowej stopy dyskontowej. Po drugie, należy rozstrzygnąć, jaka ma być podstawa oceny skutków, które przechodzą z pokolenia na pokolenie.

Wskazywane są trzy przyczyny dla stosowania społecznej, odbiegającej od rynkowej, stopy dyskontowej: *troska o przyszłe pokolenia* – według tego uzasadnienia rząd jest stróżem interesów przyszłych pokoleń. Jednakże założenie takie wymaga, aby publicznych administratorów cechował znaczny stopień wszechwiedzy i skłonność do dbania o dobro wspólne; *paternalizm* – zakładający, że ludzie są krótkowzroczni w swoich działaniach i nie są w stanie docenić przyszłych korzyści. Według Arthura Pigou: „rząd powinien więc stosować taką stopę dyskontową, którą stosowaliby ludzie, gdyby byli wiedzieli, co jest [co będzie w przyszłości] dla nich dobre”<sup>6</sup>. Argument ten prowadzi jednak do pytania, czy preferencje publiczne (tj. polityków) mogą być narzucane jednostkom; *niesprawności rynku* – w tym problemy płynące z nieistnienia rynków na przyszłe dobra. W tym przypadku zapytać można, czy społeczne skutki niesprawności interwencji rządu nie przewyższają skutków niesprawności rynku?

Długotrwała debata teoretyczna prowadzi do konkluzji, że nie ma mocnych podstaw dla tworzenia społecznej stopy dyskontowej – jeśli to pojęcie jest stosowane, to ze względu na obszar decyzyjny (decyzje dotyczące zbiorowości), a nie ze względu na faktyczną merytorycznie uzasadnioną potrzebę. Z kolei, jeśli chodzi o tzw. międzypokoleniowe preferencje czasowe, to różnice stanowisk są zakorzenione w odmiennych, lecz z osobna dobrze uzasadnionych stanowiskach. Zgodnie z pierwszym korzyści dla przyszłych pokoleń powinny ważyć mniej przy podejmowaniu bieżących decyzji niż korzyści dla obecnych pokoleń, gdyż przyszłe pokolenia będą od nich bogatsze (tak było w świecie przez ostatnie mniej więcej 200 lat). Natomiast zgodnie z drugim stanowiskiem przypisywanie mniejszej wagi korzyściom możliwym do uzyskania przez przyszłe pokolenia jest „etycznie nieuzasadnione”. Jak sądzi Hal R. Varian, zrównoważenie korzyści i kosztów różnych poko-

<sup>6</sup> Pigou A., *The Economics of Welfare*, MacMillan, Londyn 1920.

leń prowadzi w emblematycznym przypadku analizy kosztów „globalnego ocieplenia” do konkluzji zalecającej poniesienie znacznych kosztów dzisiaj w celu ochrony poziomu życia przyszłych pokoleń (zerowa stopa dyskontowa sprawia, że dolar kosztów rocznie daje w ciągu miliona lat łączne koszty równe milionowi dolarów)<sup>7</sup>. Zastosowanie dyskontowania przy użyciu nawet niskiej stopy procentowej, takiej jak 3 procent, w stosunku do skutków odczuwanych w długim okresie prowadzi do znacznych różnic w wielkości wartości teraźniejszej i w konsekwencji w ich ocenie. Techniczną odpowiedzią na ten problem jest przeprowadzenie analizy wrażliwości, czyli analizy podatności wyników na zmiany parametrów, w tym przypadku na zmiany stopy dyskontowej, przyjętych do oszacowania kosztów i korzyści<sup>8</sup>. Ostatecznie jednak decyzja o zastosowaniu danej stopy dyskontowej pozostaje decyzją etyczną, a narzędzia techniczne umożliwiają analitykowi jedynie zawężenie pola „etycznej dowolności”.

Poprawny pomiar kosztów i korzyści, dyskontowanie oraz agregacja są podstawą do zastosowania typowego dla metody kosztów i korzyści kryterium decyzyjnego, które może zostać sformułowane następująco: należy wybrać ten projekt lub tę politykę, która przyniesie największe łączne korzyści netto. W pewnych warunkach kryterium to prowadzi to decyzji tożsamej z tą, która zostałaby podjęta przy zastosowaniu kryterium stosunku kosztów do korzyści. Jednakże, kryterium różnicy netto korzyści i kosztów pozwala wziąć pod uwagę skalę projektu lub rozmach polityki, podczas gdy kryterium relacji korzyści do kosztów może prowadzić do wyboru małych projektów, których łączne skutki są małe. Zakładamy więc, że analiza kosztów i korzyści pozwala na dokładną ocenę skali korzyści i kosztów społecznych – potencjalnie, gdyż owa analiza nie przestaje budzić kontrowersji wśród filozofów, prawników i ekonomistów. Najłagodniejszy i najsłabszy z podnoszonych argumentów krytycznych odmawia analizie kosztów i korzyści uniwersalizmu i wskazuje na moralne granice takiej analizy. Najmocniejszy natomiast wprost ją odrzuca jako kryterium decyzyjne i postuluje, co czyni Henry S. Richardson, zastąpienie jej metodą „inteligentnej delibracji”, czyli „przeglądu przyjętych do realizacji celów w świetle napływającej nowej informacji”<sup>9</sup>.

<sup>7</sup> Varian R.H., „Recalculating the Costs of Global Climate Change”, *New York Times*, 14.12.2006.

<sup>8</sup> Analiza wrażliwości jest procedurą służącą badaniu wrażliwości wyników analizy kosztów i korzyści na zmiany prawdopodobieństwa wystąpienia danego poziomu kosztów i korzyści. Porównując dwie lub więcej alternatyw, analityk może wprowadzić zmiany założeń co do kosztów i korzyści. Wnioski powinny uwzględniać wrażliwość wyników analizy na zmiany założeń o kosztach i korzyściach.

<sup>9</sup> Richardson S.H., „The Stupidity of Cost-Benefit Standard”, w: M.D. Adler, E.A. Posner, *Cost-Benefit Analysis. Legal, Economic and Philosophical Perspectives*, University of Chicago Press, Chicago-London 2001.

## Cele i skutki proceduralizacji metody analizy kosztów i korzyści

Nawet najostrzejsza krytyka analizy kosztów i korzyści traci wiele na znaczeniu, gdy stawiane jest pytanie o wskazanie lepszej metody analizy czy lepszego kryterium decyzyjnego. Nie powinno więc dziwić, że nierozwiązane kontrowersje dotyczące pomiaru i monetyzacji skutków, ich ważenia i porównywalności, a także realizmu zasady kompensacji schodzą w cień wobec pytań o to, czy krytycy zalecają rezygnację z kwantyfikacji, porównywania skutków i uwzględniania konsekwencji dla całej zbiorowości? Ponieważ trudno jest znaleźć teoretyka, który postulowałby rezygnację z dążenia do wzrostu analitycznej precyzji, wciąż trwa ekspansja praktycznych zastosowań jakiejś formy analizy kosztów i korzyści. Napisałem „jakiejś formy”, gdyż przy kierunkowej akceptacji dążenia do pełnej kwantyfikacji i wyrażania w pieniądzu skutków decyzji publicznych, analiza ta może i często jest prowadzona w sposób różniący się świadomością założeń metodologicznych i rygoryzmem badawczym.

Dlatego niektórzy badacze, w tym Amartya Sen, skłonni są pisać o „dyscyplinie analizy kosztów i korzyści”, a nie o metodzie<sup>10</sup>. Elementami tej dyscypliny<sup>11</sup> są:

- a) wspólne podstawowe zasady, w tym zasada jawności wartościowania, konsekwencjonalizm etyczny, czyli pogląd, że wartość czynów zależy od ich konsekwencji oraz rachunkowość addytywna, czyli założenie, że koszty i korzyści są określane i mierzone we wspólnej przestrzeni, mogą więc być dodawane i odejmowane;
- b) wymogi strukturalne, w tym założenie o kompletności ewaluacji (możliwości identyfikacji wszystkich konsekwencji), założenie pełnej informacji oraz wiedzy o dystrybucji prawdopodobieństwa oraz
- c) wartościowanie odnoszone do procesów rynkowych, poprzez wykorzystywanie gotowości do zapłaty i zasady potencjalnej kompensacji.

Zdaniem Sena system analizy kosztów i korzyści jest natomiast indyferentny wobec motywacji podmiotów działających (ocena motywacji jest ważniejsza w decyzjach osobistych niż przy wyborze publicznym), praw jednostki oraz wartości indywidualnej wolności.

W ramach tak rozumianej dyscypliny analizy kosztów i korzyści mieszczą się praktyczne metody i procedury, które służą do przygotowywania decyzji publicznych.

Wspomnieliśmy już, że analiza kosztów i korzyści dotyczy skutków działań (projektów, polityk) podejmowanych w imieniu i dla zbiorowości. Jej podstawową zaletą jest to, że poprzez *ex ante* analizę wydatków publicznych

<sup>10</sup> Sen A., „The Discipline of Cost-Benefit Analysis”, w: M.D. Adler, E.A. Posner, op. cit.

<sup>11</sup> Według Władysława Kopalińskiego w języku polskim dyscyplina to także ustalony porządek, podporządkowanie się wewnętrznemu regulaminowi.

z punktu widzenia ich celowości i efektywności kosztowej może zapobiegać podejmowaniu błędnych lub zbyt kosztownych programów.

## Pragmatyzm analizy kosztów i korzyści

Analiza kosztów i korzyści jest instytucjonalizowana jako element polityki publicznej w krajach o pragmatycznej (a nie ideologicznej) orientacji rządów. Jest upowszechniana jak metoda poprawy jakości interwencji rządu w krajach anglosaskich – USA, Kanadzie czy Wielkiej Brytanii. Nie ma zastosowania w krajach realizujących cele ideologiczne, takie jak budowa pozycji międzynarodowej czy międzynarodowa ekspansja państwa. Trudno znaleźć jej zastosowanie w krajach autorytarnych, gdzie w decyzjach publicznych nie jest uwzględniany głos poszczególnych obywateli.

Powtórzmy raz jeszcze: w standardowych sytuacjach wymóg dokonywania kalkulacji potencjalnych skutków projektów i polityk publicznych dla dobrobytu społecznego nie powinien budzić kontrowersji. Rozpowszechnienie tego typu analiz zdecydowanie poparła grupa wpływowych ekonomistów pod kierunkiem K. Arrowa<sup>12</sup>. Stwierdzili oni, że przy podejmowaniu decyzji publicznych nie powinno się ograniczać analizy ekonomicznych kosztów i korzyści różnych typów przygotowywanych działań, a agendy rządowe powinny przeprowadzać takie analizy przy ustanawianiu priorytetów wszelkich swoich działań. Zalecali ponadto, aby jak najczęściej analiza pochodziła ze źródeł zewnętrznych w stosunku do urzędów. Postulowali także, aby analiza ekonomiczna uwzględniała takie kategorie, jak: społeczna stopa dyskontowa, wartość redukcji ryzyka przedwczesnej śmierci i wypadków oraz wartości związane z poprawą stanu zdrowia.

Przyjmując pragmatyczne podejście do analizy kosztów i korzyści, zarysujmy jej podstawowe etapy. W pierwszym etapie analizy kosztów i korzyści przygotowujący projekt czy proponujący politykę urząd jest zobowiązany do *jasnego zdefiniowania problemu* i pokazania, że projekt, czy polityka są rzeczywiście potrzebne. Następnie urząd powinien zbadać *alternatywne* sposoby rozwiązania problemu, w tym takie sposoby, które nie wiążą się z ponoszeniem wydatków. W kolejnym etapie jest miejsce na *analizy kwantyfikujące korzyści i koszty* związane z realizacją każdej z potencjalnych alternatyw. Potem, powinna zostać przeprowadzona *konsultacja* z instytucjami i grupami, na które projekt lub polityka może oddziaływać w celu weryfikacji jej empirycznych podstaw, technicznej wykonalności i społecznego poparcia.

---

<sup>12</sup> Arrow K.J., Cropper M.L., Eads G.C., Hahn R.W., Lave L.B., Noll R.G., Portney P. R., Russel M., Schmalensee R., Smith V.K., Stavins R.N., „Is there a role for benefit-cost analysis in environmental, health and safety regulation”, *Science*, 12.04.1996.

W pragmatycznym ujęciu analiza kosztów i korzyści jest zbiorem praktycznych procedur służących podejmowaniu decyzji co do wydatków publicznych. Zastosowanie procedur w celu identyfikacji i klasyfikacji wartości i preferencji, które leżą u podstaw danego programu politycznego, jest szczególnie istotne dla wyboru kryterium decyzyjnego (np. efektywność, skuteczność, sprawiedliwość) w przypadku decyzji dotyczących problemów zbiorowości.

Tabela. Główne zadania w analizie typu koszty–korzyści

Zadanie	Opis
Strukturyzacja problemu publicznego	Sformułowanie problemu poprzez określenie wstępnych celów, alternatyw, kryteriów, grup zainteresowanych, kosztów i korzyści
Uszczegółowienie celów	Przekształcenie celów ogólnych w czasowo określone i mierzalne cele
Określenie alternatyw	Wybór najkorzystniejszych alternatywnych polityk z szerszego zbioru potencjalnych rozwiązań danego problemu
Poszukiwanie informacji, analiza i interpretacja	Znalezienie, analiza i interpretacja informacji potrzebnych do przewidzenia skutków danych alternatyw
Identyfikacja zainteresowanych grup i beneficjentów danego programu	Wyliczenie wszystkich grup, których dotyczy dany program publiczny (np. regulacja) lub celowe zaniechanie działań publicznych, oraz tych, którzy odniosą korzyści z działania lub zaniechania
Oszacowanie kosztów i korzyści	Oszacowanie w jednostkach pieniężnych szczegółowych kosztów i korzyści każdej z alternatyw we wszystkich kategoriach kosztów
Dyskontowanie kosztów i korzyści	Sprowadzenie pieniężnych kosztów i korzyści do ich wartości obecnej przy użyciu stóp dyskontowych
Oszacowanie ryzyka i niepewności	Zastosowanie analizy wrażliwości do oszacowania prawdopodobieństwa wystąpienia przyszłych kosztów i korzyści
Wybór kryterium decyzyjnego	Wybór kryterium decyzyjnego spośród następujących alternatyw: poprawa w sensie Pareto, maksymalizacja korzyści netto
Zalecenia	Wybór alternatywy, która jest najbardziej prawdopodobna, z uwzględnieniem konkurencyjnych założeń etycznych i hipotez przyczynowych

Źródło: Opracowanie na podstawie: Dunn W., *Public Policy Analysis. An Introduction*, wyd. 3, Prentice Hall 2003.

## Analiza kosztów korzyści w perspektywie polityki publicznej

Znaczenie jednolitych reguł i procedur oraz jawnych i jednolitych parametrów decydowania o wydatkach jest niezwykle ważne, gdyż współczesne społeczeństwa domagają się ograniczenia dyskrecjonalności w działaniach polityków i rządów. Chcą, aby decyzje polityczne (decyzje rządów) nie były wynikiem materialnych interesów samych polityków (rządzących), ani też

interesów partyjnych. Warto podkreślić, że postulat ograniczenia dyskrecjonalności jest ściśle związany z żądaniem maksymalizacji dobrobytu szerokich grup wyborców, a nie tylko z postulatem demokratycznej kontroli. Związek ten dobrze wyjaśnia wskazana przez Josepha Stiglitz'a analogia pomiędzy relacją między menedżerami przedsiębiorstwa a jego właścicielami i relacją między rządzącymi a obywatelami. W obu przypadkach proceduralne ograniczenia dyskrecjonalności mają na celu zmniejszenie skłonności menedżerów i rządzących do działań służących własnym interesom i stworzenie bodźców do działań na rzecz zwiększania dobrobytu udziałowców i interesariuszy (*stakeholders*)<sup>13</sup>.

Ograniczenia proceduralne nie są więc celem same w sobie, nie są li tylko sposobem realizacji wyobrażeń etycznych, co do sposobów urzeczywistnienia ideałów sprawiedliwego i bezstronnego rządu. Są także, a może przede wszystkim, narzędziem służącym maksymalizacji dobrobytu społecznego.

Przeciwieństwem dyskrecjonalności jest działanie według reguł. Może się wydawać paradoksalne, że okoliczności działań dyskrecjonalnych są eliminowane (lub ograniczane), chociaż w pewnych sytuacjach działania dyskrecjonalne mogłyby lepiej służyć maksymalizacji dobrobytu społecznego niż działania według reguł. Wyjaśnienie tego problemu wymaga wprowadzenia rozróżnienia pomiędzy efektami działań rządu w krótkim a skutkami długim okresie. Badania nowoczesnej ekonomii politycznej dowodzą, że działanie według reguł przynosi korzyści w dłuższym okresie właśnie dzięki ograniczeniu dyskrecjonalności. Jest tak pomimo tego, że w danym momencie i przypadku bardziej korzystna mogłaby być decyzja dyskrecjonalna – decyzja podjęta z naruszeniem proceduralnych ograniczeń.

Podstawową zaletą reguł jest to, że zmniejszają one niepewność co do interpretacji norm ustanawianych przez władze publiczne. Znaczenie reguł jest szczególnie istotne w otoczeniu cechującym się dużą złożonością, w systemie składającym się z dużej liczby aktorów i wyższej częstotliwości działań.

W badaniach ekonomii politycznej – w badaniach prowadzonych w perspektywie metodologicznego indywidualizmu i przyjmujących założenie o racjonalności działań jednostki, znaczenie reguł nie było do niedawna dostatecznie zauważane. Sytuacja ta zaczęła się zmieniać od początku lat 60., dzięki pracom Jamesa M. Buchanana i Jamesa S. Colemana. Analizy te wskazały na dwa źródła społecznej użyteczności norm i reguł. Po pierwsze, są one narzędziem rekompensującym „niedostateczną racjonalność” jednostki, jej ograniczenia poznawcze oraz koszty i niekompletność informacji. Po drugie, normy i reguły mogą pomagać w przewyżnianiu swoistej „nadmiernej racjonalności” jednostki (w znaczeniu racjonalności instrumentalnej), która jej

<sup>13</sup> Stiglitz J., „Principal and agent (II)”, w: J. Eatwell, M. Milgate, P. Newman (red.), *The New Palgrave: A Dictionary of Economics*, t. 3, Macmillan Press, London 1987.

utrudnia realizację długookresowych celów, skłaniając ją do oportunistycznych zachowań.

Istnienie reguł umożliwia jednostce uzyskanie informacji pozwalających jej sformułować poprawne oczekiwania dotyczące zachowań innych ludzi i przedsiębiorstw. Większość reguł ma charakter samopodtrzymujący, czyli nie wymaga istnienia zewnętrznego organu sankcjonującego ich łamanie. Dzieje się tak wtedy, gdy przestrzeganie reguł jest korzystne dla wszystkich. Dostarczając informacji, reguły wspierają procesy rynkowe, czasami jednak mogą zastępować transakcje rynkowe. Twierdzenie to stanie się bardziej zrozumiałe, gdy zostanie wyjaśnione na następującym przykładzie.

Można sobie wyobrazić, że zamiast zachowań według reguły „przejazd na czerwonym świetle jest zabroniony, przejazd na zielonym świetle jest dozwolony”, każdym skrzyżowaniem zarządzać będzie osoba, które będzie przydzielać prawo do przejazdu przez skrzyżowanie poprzez alokację na zasadzie aukcji. Jednakże taka hipotetyczna możliwość wiąże się z wysokimi kosztami transakcyjnymi i reguły przejazdu okazują się bardziej efektywne niż czyste instytucje rynkowe w rozwiązywaniu problemu koordynacji przejazdów przez skrzyżowania.

Reguły zawierają bodźce do określonego działania. Analiza reguł wymaga, aby zidentyfikowane zostały zawarte w nich bodźce. Od czasów prac Adama Smitha wiadomo, że pewne reguły gospodarki rynkowej, takie jak prawo własności czy prawne instrumenty egzekwowania dobrowolnych zobowiązań, służą harmonizacji interesów częściowych w interes ogólny. Tworzący, w tym samym, co Adam Smith, okresie, Alexander Hamilton wskazał na etyczne znaczenie reguł, pisząc, że „reguły mogą sprawić, że interesy zbiegają się z powinnością”. Według Geoffreya Brennana i Alana Hamlina stanowisko Hamiltona oznacza uznanie tego, że „instytucje rynkowe nie tylko dostarczają tego, co powinno zostać dostarczone, lecz również mówią nam, czym to jest”<sup>14</sup>.

Doświadczenie wskazuje również na możliwość występowania prawidłowości, zgodnie z którą ludzie nie tylko odpowiadają na bodźce, ale, w pewien sposób, instytucje „dobierają sobie ludzi”, którzy odpowiadają na ich bodźce. Dla lepszego zilustrowania tej prawidłowości warto posłużyć się przykładem. Z dużym prawdopodobieństwem stwierdzić można, że ludzie unikający ryzyka nie będą inwestować na giełdzie, lecz będą kupować obligacje skarbowe, że częściej będą administratorami w instytucjach państwowych niż menedżerami w przedsiębiorstwach rynkowych. Uogólniając, w ujęciu statycznym, stwierdzić można, że instytucje o pewnych wymaganiach i ludzie o pewnych dyspozycjach dopasowują się wzajemnie. W dynamicznym ujęciu można powiedzieć, że ludzie aspirujący do pewnych stanowisk starają się wykształcić dyspozycje umożliwiające im skuteczne funkcjonowanie w takich instytu-

<sup>14</sup> Brennan G., Hamlin A., „Constitutional Economics”, w: P. Newman (red.), *The New Palgrave Dictionary of Economics and the Law*, Macmillan, London 1998.



cjach. Optymalna selekcja ludzi do instytucji wymaga rozwiązania problemu „fałszywych sygnałów”, czyli efektywnego eliminowania jednostek, które udają, że posiadają wymagane przez nie cechy.

Gry koordynacyjne są grami, w którym istnieje zbieżność pomiędzy korzyścią zbiorową a korzyścią indywidualną, trudność polega jedynie na „skoordynowaniu zachowań jednostek w kierunku pewnego wspólnego wzoru”<sup>15</sup>. Przykładem problemu koordynacji jest unikanie kolizji drogowych w sytuacji dowolności wyboru strony drogi, po której poruszają się pojazdy. Z punktu widzenia jakości ruchu drogowego nie jest ważne, czy pojazdy poruszają się prawą czy lewą stroną drogi, ważne jest, aby poruszały się tą samą stroną drogi.

W funkcji rozwiązywania problemów koordynacyjnych reguły służą dostarczaniu informacji o tym, co jest kolektywnie optymalne i jak tę kolektywną optymalność osiągnąć. W życiu zbiorowości reguły, wskazał na to Andrew Schotter, mogą służyć zapobieganiu powtarzalnemu występowaniu zachowań, które prowadzą do zbiorowej nieefektywności<sup>16</sup>.

## Zakończenie

Dotychczasowe wyniki poszukiwań obiektywnej i naukowej podstawy podejmowania decyzji publicznych nie są zadowalające. Należy się zgodzić ze stwierdzeniem dwójki klasyków analizy polityk publicznych, którzy mówią, że chociaż „filozofowie i ekonomiści próbowali od dwóch stuleci opracować jednoznaczne procedury służące do pomiaru i łączenia dobrobytu dwóch lub więcej jednostek w celu zmierzenia łącznego dobrobytu społecznego, to ich dążenie okazało się takim sukcesem, jakim były próby alchemików przekształcenia ołowiu w złoto. Sporadyczne iskierki nadziei wygasły, nie tylko nie powstała żadna wykonalna procedura, nie ma także takiej na horyzoncie”<sup>17</sup>.

Z kolei według innego pesymistycznego stwierdzenia nawet „amerykańska praktyka analizy kosztów i korzyści jest rodzajem ozdoby dodawanej do już podjętych planów, które przygotowane zostały bez większego odniesienia do kryteriów efektywności ekonomicznej”<sup>18</sup>.

Nie trzeba być jednak całkowitym pesymistą. Chociaż decyzja o rozpoczęciu wojny w Iraku nie zapadła po zastosowaniu analizy kosztów i korzyści, to jej niecałkowicie *ex post* skutki zostały zmierzone. Jakkolwiek jest prawdą, że wiele projektów publicznych ma na celu głównie budowanie poparcia

---

<sup>15</sup> Problemy koordynacji omawia m.in. praca Besanko D., Dranove D., Shanley M., *Economics of Strategy*, John Wiley & Sons 2000).

<sup>16</sup> Schotter A., *The Economic Theory of Social Institutions*, Cambridge University Press, 1981.

<sup>17</sup> Stockey E., Zeckhauser R., *A Primer for Policy Analysis*, Norton, New York 1978, s. 276.

<sup>18</sup> Marglin S., *Public Investment Criteria: Benefit-Cost Analysis for Planned Economic Growth*, Mass.: MIT Press, Cambridge 1967.

politycznego, to można mieć nadzieję, że dzięki analizie kosztów i korzyści poparcie to może być budowane przy mniejszym marnotrawstwie środków publicznych.

Innymi słowy analiza kosztów i korzyści, tam gdzie jest stosowana i upowszechniana, służy poprawie społecznego dobrobytu nie dlatego, że zapewnia niezawodną ocenę zmian wywoływanych przez programy publiczne, ale dlatego, że dostarcza języka racjonalnej debaty publicznej, umożliwiając zrozumienie i sformułowanie natury sporu politycznego oraz dlatego, że wymusza ujawnianie informacji, na podstawie których są podejmowane decyzje władz publicznych. Myśl tę najlepiej oddają słowa Richarda A. Posnera, który pragmatyczną obronę analizy kosztów i korzyści, tj. niegłoszącą tezy, że analiza kosztów i korzyści jest jedyną zasadą decyzyjną właściwą dla polityk publicznych, zakończył stwierdzeniem, że „analiza kosztów i korzyści nie wymaga żadnego głębszego i bardziej rygorystycznego uzasadnienia niż to, że dzięki niej uzyskujemy wyniki, które lubimy”<sup>19</sup>. Te preferowane społecznie wyniki nazwać można triadą społecznego dobrobytu: więcej konkurencji na rynku, więcej rynku w działaniach państwa i więcej dobrowolnych zbiorowych działań jednostek zamiast biurokratycznej intruzji państwa.

---

**Aleksander Surdej** – doktor habilitowany nauk ekonomicznych, kierownik Katedry Studiów Europejskich i profesor Uniwersytetu Ekonomicznego w Krakowie. Swoje zainteresowania badawcze koncentruje na analizie polityki regulacyjnej, w tym szczególnie analizie polityki regulacji społecznych. Tym zagadnieniom poświęcona jest jego monografia: *Determinanty regulacji administracyjnoprawnych w oddziaływaniu państwa na gospodarkę* (Kraków 2006). W swoich publikacjach, artykułach, rozdziałach w opracowaniach zbiorowych oraz raportach dla Banku Światowego i Komisji Europejskiej publikowanych w języku polskim, włoskim i angielskim analizuje inne polityki publiczne okresu transformacji, w tym politykę państwa wobec małych i średnich przedsiębiorstw, wspieranie rozwoju przedsiębiorczości (*Supporting Youth Entrepreneurship: the Case of Poland*, Palgrave 2008) oraz politykę regulacji ryzyka na przykładzie bezpieczeństwa żywności (*Enlarging the EU Food Safety Regime: Selected Problems in Adjusting the Polish Food Safety Regime to EU Food Safety Requirements*, Florencia 2004). Uczestniczył w wielu międzynarodowych projektach badawczych, był stypendystą Holenderskiej Akademii Nauk, Uniwersytetu Europejskiego we Florencji oraz Uniwersytetu Narodów Zjednoczonych. Członek Rady Naukowej Instytutu Organizacji i Zarządzania w Przemysle „Orgmasz”. Absolwent UJ, Johns Hopkins University oraz Uniwersytetu Europejskiego we Florencji. Redaktor naczelny i wydawca pisma internetowego *Panstwo i rynek* ([www.pir.org.pl](http://www.pir.org.pl)).

---

<sup>19</sup> Posner R., „CBA: Definition, Justification and Comment”, w: M.D. Adler, E.A. Posner, op.cit.

## Bibliografia

- Arrow K.J., Cropper M.L., Eads G.C., Hahn R.W., Lave L.B., Noll R.G., Portney P. R., Russel M., Schmalansee R., Smith V.K., Stavins R.N., „Is there a role for benefit-cost analysis in environmental, health and safety regulation”, *Science*, 12.04.1996.
- Bednarski M., Wilkin J., *Ekonomia dla prawników*, PWN, Warszawa 2003.
- Besanko D., Dranove D., Shanley M., *Economics of Strategy*, John Wiley & Sons 2000.
- Blaug M., *Teoria Ekonomii*, PWN, Warszawa 1994.
- Boardman A.E., Greenberg D.H., Vining A.R., Weimer D.L., *Cost-Benefit Analysis: Concepts and Practice*, Upper Saddle River, Prentice-Hall 2001.
- Brennan G. Hamlin A., *Constitutional Economics*, w: P. Newman (red.), *The New Palgrave Dictionary of Economics and the Law*, Macmillan, London 1998.
- Coleman J.S., „Norms as Social Capital”, w: G. Radnitzky, P. Bernholz (red.), *Economic Imperialism. The Economic Approach Applied Outside the Field of Economics*, Paragon House, New York.
- Dunn W., *Public Policy Analysis. An Introduction*, wyd. 3, Prentice Hall 2003.
- Dupuit J., *On the measurement of the utility of public works* (tłumaczenie na język angielski w: *International Economic Papers* 1952, nr 2).
- Dupuit J., „On utility and its measure – on public utility”, *Journal des Economistes*, nr 36.
- Haveman R.H., Weimer D.L., „Cost Benefit Analysis”, *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier 2001.
- Kaldor N. „Welfare propositions of economics and interpersonal comparisons of utility”, *Economic Journal* 1939, vol. 49.
- Marglin S., *Public Investment Criteria: Benefit-Cost Analysis for Planned Economic Growth*, Mass.: MIT Press, Cambridge 1967.
- Persky J., „Retrospectives: Cost-benefit analysis and the classical creed”, *The Journal of Economic Perspectives* 2001, vol. 15, nr 4.
- Pigou A., *The Economics of Welfare*, MacMillan, Londyn 1920.
- Posner R., „CBA: Definition, Justification and Comment”, w: M.D. Adler, E.A. Posner, *Cost-Benefit Analysis. Legal, Economic and Philosophical Perspectives*, University of Chicago Press, Chicago–London 2001.
- Richardson S.H., „The Stupidity of Cost-Benefit Standard”, w: M.D. Adler, E.A. Posner, *Cost-Benefit Analysis. Legal, Economic and Philosophical Perspectives*, University of Chicago Press, Chicago–London 2001.
- Sen A., „The Discipline of Cost-Benefit Analysis”, w: M.D. Adler, E.A. Posner, *Cost-Benefit Analysis. Legal, Economic and Philosophical Perspectives*, University of Chicago Press, Chicago–London 2001.
- Schotter A., *The Economic Theory of Social Institutions*, Cambridge University Press 1981.
- Stiglitz J., „Principal and agent (II)”, w: J. Eatwell, M. Milgate, P. Newman (red.), *The New Palgrave: A Dictionary of Economics*, t. 3, Macmillan Press, London 1987.
- Stockey E., Zeckhauser R., *A Primer for Policy Analysis*, Norton, New York 1978.
- Varian R.H., „Recalculating the Costs of Global Climate Change”, *New York Times*, 14.12.2006.

Cezary Trutkowski

## O uwarunkowaniach stosowania metod jakościowych w badaniach społecznych i ewaluacyjnych

Pod koniec lat 90. ubiegłego wieku (a zatem nie więcej niż 10 lat temu!), na zjeździe Amerykańskiego Stowarzyszenia Ewaluacyjnego, ówczesny prezydent tej organizacji, William Sadish, podsumowując rozwój badań ewaluacyjnych stwierdził: „założę się, że gdyby poprosić członków naszego stowarzyszenia, aby wskazali jedno zagadnienie, najczęściej dyskutowane i najbardziej dzielące [ewaluatorów] w trakcie ostatnich 20 lat, to wskazałoby debaty ilościowo-jakościowe. Elementem tych debat z pewnością była metodologia, jako że wielu ewaluatorów musiało dopiero zapoznać się z metodami jakościowymi, o których przecież wiedzieli dość niewiele”<sup>1</sup>.

Dziś, pod koniec pierwszej dekady XXI wieku debaty te nie są już tak gorące, a kwestia stosowania metod ilościowych i jakościowych nie kreśli wśród badaczy zdecydowanych linii podziałów teoretycznych. Część z nich uznała bowiem dyskusję wokół metod za jałową i nieprowadzącą do żadnych powszechnie akceptowanych ustaleń. Badacze ci zajęli się na ogół uprawianiem swej dyscypliny, wykorzystując te metody, dla których osobiście znajdują lepsze uzasadnienie, w których zostali lepiej wyszkoleni i w których stosowaniu mają większe doświadczenie. Z dystansem przy tym obserwują działania kolegów po drugiej stronie dawnego metodologicznego sporu. Inni próbują znaleźć drogę do integracji podejścia ilościowego i jakościowego, wskazując na wynikające z niej możliwości pełniejszego spojrzenia na badany problem i podkreślając pozytywny efekt triangulacji wynikający ze stosowania metod charakteryzujących się różnym rodzajem błędu pomiaru<sup>2</sup>. Wreszcie trzecią grupę stanowią badacze społeczni świadomi różnic paradygmatycznych, leżących u podłoża obu perspektyw, i przez to odrzu-

---

<sup>1</sup> Sadish W.R., „Evaluation theory is who we are”, *American Journal of Evaluation* 1998, vol. 19, nr 1, s. 3.

<sup>2</sup> Jasiński M., Kowalski M. „Falszywa sprzeczność: metodologia jakościowa czy ilościowa?”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.

cający prostą ideę triangulacji, lecz pragmatycznie wykorzystujący w praktyce badawczej oba podejścia w odniesieniu do różnych, odpowiadających im zjawisk.

Nie zamierzam uczestniczyć w zarysowanej wyżej debacie. Chcę jednak zastanowić się nad naturą i celem stosowania badań jakościowych, praktycznymi uwarunkowaniami ich wykorzystywania oraz konsekwencjami, jakie to rodzi dla ewaluacji. Istnieje bowiem istotna różnica pomiędzy badaniami społecznymi a badaniami ewaluacyjnymi, która sprawia, że przynajmniej w odniesieniu do tych ostatnich, debata ilościowo-jakościowa traci na znaczeniu.

### Specyfika metod jakościowych

Jak zauważa Babbie w swoim, klasycznym już, podręczniku: „jakościowe badania terenowe umożliwiają badaczom obserwowanie życia społecznego w jego naturalnym środowisku: idź tam, gdzie toczy się proces i obserwuj go”. Tym samym metody jakościowe „są szczególnie odpowiednie do poznawania tych postaw i zachowań, które najlepiej można zrozumieć w ich naturalnym otoczeniu”<sup>3</sup>. I to jest właśnie pierwsza, zasadnicza cecha konstytutywna badań jakościowych – pozwalają one na uchwycenie procesów społecznych w trakcie ich trwania i objęcie obserwacją szerokiego spektrum zjawisk składających się na te procesy bez konieczności predefiniowania jednostek obserwacji, jak to ma miejsce w przypadku np. badań sondażowych.

Zakładając pewne uproszczenie, można także przyjąć, że metody jakościowe to takie sposoby postępowania badawczego, które dają rezultat w postaci nienumerycznych, niekwantyfikowalnych danych (przybierających na ogół postać tekstu – np. opisy obserwacji, transkrypcje wywiadów, opinie ekspertów itp.)<sup>4</sup>.

Narracyjna postać danych jakościowych w założeniu umożliwia głębsze i pełniejsze poznanie badanego problemu<sup>5</sup> (m.in. dzięki aktualizacji przekazu), a proces zbierania tych danych zakłada bezpośredni udział badacza w naturalnie zachodzących procesach społecznych. Zatem drugą cechą wyróżniającą jakościowe metody badawcze jest format uzyskiwanych danych – na ogół są to dane w takiej, czy innej formie tekstowej.

Jak zauważa przy tym Bamberger: „celem analiz opisowych jest zrozumienie unikalnych charakterystyk określonego kontekstu (np. społeczności,

<sup>3</sup> Babbie E., *Badania społeczne w praktyce*, Wydawnictwo Naukowe PWN, Warszawa 2003, s. 308–310.

<sup>4</sup> Por. Schwandt T.A., *Qualitative inquiry*, Sage, Thousand Oaks, CA 1997.

<sup>5</sup> Miller G., Dingwall R. *Analyzing Social Settings*, Wadsworth Publishing, Belmont, CA 1997.

organizacji, wydarzenia), rodziny lub jednostki”<sup>6</sup>. Oznacza to, że badania jakościowe różnią się od badań ilościowych także zakresem analizowanych zjawisk – zainteresowanie badaczy skupia się tu przede wszystkim na poziomie lokalnym (mikro)<sup>7</sup>, a celem badań jest zrozumienie procesów zachodzących na poziomie jednostkowym bądź grupowym, a nie jakakolwiek agregacja. Uzyskiwane dane mają w związku z tym charakter kontekstowy i pozostają silnie uzależnione od lokalnych sposobów interpretacji. Należy przy tym pamiętać, że podejście jakościowe stawia sobie za cel holistyczne rozumienie badanego zjawiska<sup>8</sup>, co podkreśla Babbie, zauważając, że „przez podążanie do zjawiska społecznego poddawanego badaniu i obserwowanie go na tyle całościowo, na ile to jest możliwe, badacz ma szansę uzyskać głębsze i pełniejsze jego rozumienie”<sup>9</sup>.

Kolejną cechą wyróżniającą badania jakościowe jest specyfika procesu gromadzenia danych. Proces ten jest na ogół znacznie mniej ustrukturyzowany niż ma to miejsce w przypadku badań ilościowych. W zasadzie nie istnieją sztywne kanony dotyczące rozplanowania badań jakościowych, zakresu i czasu ich realizacji, doboru respondentów, czy nawet sposobu formułowania pytań. Można przyjąć, że jak w przypadku badań ilościowych, minimalną gwarancją sukcesu jest trzymanie się procedur od dawna utrwalonych w teorii i praktyce badawczej (np. zasady doboru próby, reguły formułowania pytań, standardy prowadzenia badań w terenie, czy wzory analiz statystycznych), tak w badaniach jakościowych jedyną gwarancją może być profesjonalizm, doświadczenie i wiedza metodologiczna badacza. Podkreślam, że elastyczność metodologii jakościowej nie oznacza w żadnym razie ekspresji anarchizmu metodologicznego<sup>10</sup>. Po pierwsze, dobra praktyka podejścia jakościowego oznacza systematyczność realizacji badań – prowadzenie ich zgodnie z wcześniej ustalonymi i wyeksplikowanymi założeniami dotyczącymi konkretnego projektu badawczego (począwszy od jednoznacznego sformułowania pytań badawczych, poprzez opracowanie i przetestowanie narzędzi, określenie czasowego i przestrzennego zakresu prac terenowych, wypracowanie i skuteczne wdrożenie zasad doboru respondentów, uzgodnienie form utrwalania danych oraz zasad analizy i prezentacji wyników, a skończywszy na różnych sposobach weryfikacji wypracowanych wniosków<sup>11</sup>). Po drugie, istnieje sze-

<sup>6</sup> Bamberger M., „Opportunities and challenges for integrating quantitative and qualitative research”, w: M. Bamberger (red.), *Integrating quantitative and qualitative research in development projects*, World Bank, Washington, DC 2000, s. 15.

<sup>7</sup> Ibidem.

<sup>8</sup> Zob. Chung K., „Issues and approaches in the use of integrated methods”, w: M. Bamberger, op. cit.

<sup>9</sup> Babbie E., op. cit., s. 309.

<sup>10</sup> Szerzej na ten temat: Górniak J., Worek B., Krupnik S. „Zastosowanie podejścia badań jakościowych w ewaluacji *ex-post*”, w: A. Haber (red.), *Ewaluacja ex-post...*, op. cit.

<sup>11</sup> Ibidem.

reg ustalonych i dość powszechnie akceptowanych reguł prowadzenia badań jakościowych<sup>12</sup> – np. prowadzenia grup fokusowych<sup>13</sup>, wywiadów pogłębionych<sup>14</sup> czy realizacji innych technik<sup>15</sup>.

Powyższe zestawienie cech konstytutywnych badań jakościowych pozwala wskazać ich mocne i słabe strony z punktu widzenia praktyki badawczej. Bez wątplenia, jak zauważa Babbie, są one „szczególnie efektywne, gdy stosuje się je do badania subtelnych niuansów w postawach i zachowaniach oraz do badania procesów społecznych w czasie. Dlatego główna mocna strona tej metody leży w głębi rozumienia, które umożliwia”<sup>16</sup>. Inne mocne strony to elastyczność, na którą pozwalają, i większa trafność ustaleń niż w przypadku badań ilościowych (uzyskiwana właśnie dzięki głębi rozumienia i holistycznemu podejściu do przedmiotu badania). Z kolei wśród słabych stron badań jakościowych najczęściej wymienia się brak możliwości generalizacji wyników<sup>17</sup> i problemy z rzetelnością – zbytni subiektywizm wypracowywanych wniosków i ich uzależnienie od osobistego osądu i perspektywy badacza<sup>18</sup>.

## Konsekwencje perspektywy badawczej – metody kontekstowe i pozakontekstowe

Jak zauważa Alvin Gouldner: „stosowanie takich a nie innych technik badawczych zakłada istnienie określonych założeń na temat człowieka i społeczeństwa”<sup>19</sup>. Tym samym, uznanie zasadności wykorzystywania określonych metod wymaga uwzględnienia i weryfikacji założeń dotyczących natury ba-

---

<sup>12</sup> Bliższe informacje w wydanych ostatnio: Babbie E., *Badania społeczne w praktyce*, op.cit.; Silverman D., *Interpretacja danych jakościowych*, PWN 2008; Rubacha K. *Metodologia badań nad edukacją*, Wydawnictwa Akademickie i Profesjonalne 2008.

<sup>13</sup> Zob. np. Maison D., *Zogniskowane wywiady grupowe. Jakościowa metoda badań marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 2001; Morgan D.L., *Successful focus groups. Advancing the state of the art.*, Sage 1993; Morgan D.L., „Focus groups”, *Annual Review of Sociology* 1996, vol. 22.

<sup>14</sup> Helling I.K., „Metoda badań biograficznych” w: J. Włodarek, M. Ziółkowski, *Metoda biograficzna w socjologii*, PWN, Warszawa 1990; Rubin H.J., Rubin, I.S., *Qualitative interviewing: The art of hearing data*, Sage Publications, Thousand Oaks, CA 2004.

<sup>15</sup> Antoszkiewicz J., *Metody heurystyczne. Twórcze rozwiązywanie problemów*; Państwowe Wydawnictwo Ekonomiczne, Warszawa 1990; Straś-Romanowska M. (red.), *Metody jakościowe w psychologii współczesnej*, Wydawnictwo Uniwersytetu Wrocławskiego 2000.

<sup>16</sup> Babbie E., op. cit., s. 334.

<sup>17</sup> Chung K., op. cit., s. 42.

<sup>18</sup> Babbie E., op. cit., s. 335.

<sup>19</sup> Gouldner A., „Teoretyczny kontekst socjologii”, w: E. Mokrzycki, *Kryzys i schizma*, PIW, Warszawa 1984, s. 89.

danej rzeczywistości. Jeśli przedmiotem naszego zainteresowania są zjawiska zachodzące na poziomie lokalnym (mikro), np. w obrębie jakiejś grupy społecznej, instytucji, społeczności itp., to powinniśmy do ich badania wykorzystywać metody, które umożliwią nam uchwycenie i uwzględnienie lokalnego kontekstu.

Kontekst ten możemy zdefiniować za Teunem van Dijkem jako „ustrukturyzowany zestaw wszystkich właściwości sytuacji społecznej, które mogą być istotne dla produkcji, struktury, interpretacji oraz funkcji tekstu i wypowiedzi”<sup>20</sup>. W szerszym ujęciu zbiór sytuacji społecznych (zdarzeń komunikacyjnych), w których uczestniczą jednostki wyznaczać będzie wspólnotę komunikacyjną tworzoną w określonych społecznych ramach (np. instytucji, projektu, kręgu przyjaciół itp.). To właśnie przynależność do określonych komunikacyjnych wspólnot decyduje o kształcie naszej rzeczywistości społecznej, związkach, które nawiązujemy i kontynuujemy w trakcie naszego życia, o grupach, do których należymy, i o rzeczach, które decydujemy się robić.

Świat, w którym żyjemy, który nas bezpośrednio otacza (a zatem właśnie nasz świat lokalny, nasza skala mikro), jest bowiem wytwarzany i przetwarzany przede wszystkim w ramach komunikacji, którą prowadzimy zarówno z innymi ludźmi, jak i z samymi sobą. To właśnie w ramach dokonywanej nieustannie interpretacji i nazywania rzeczywistości „wytwarzamy” przedmioty i relacje, które w efekcie stają się dla nas jej znaczącymi elementami. Elementy te uzgadniamy z innymi – ustalamy znaczenia tak, abyśmy mogli w przyszłości czuć, że żyjemy w świecie wspólnym, uznanym i zaakceptowanym przez ważnych dla nas ludzi.

Nasze doświadczenia oraz – dzięki relacjom z innymi – poznane doświadczenia innych, wpływają na rozumienie świata, stanowiące podstawę społecznego funkcjonowania w różnego rodzaju grupach. Rozumienie to i porozumienie z innymi w ramach określonej wspólnoty komunikacyjnej możliwe jest dzięki modelom kontekstowym – swego rodzaju umysłowym reprezentacjom określonych zdarzeń komunikacyjnych – które z biegiem czasu współwypracowujemy. Modele kontekstowe, jako część osobistego doświadczenia jednostki, składają się na jej doświadczenie społeczne. Same w sobie „mają charakter osobisty, subiektywny i potencjalnie stronniczy (uprzedzeniowy), a jako takie odzwierciedlają osobowościowo zmienne interpretacje i opinie na temat zdarzeń komunikacyjnych”<sup>21</sup>. Modele kontekstowe, będące przedmiotem analiz w postępowaniu badawczym, obejmujące wspólnoty komunikacyjne (a więc pośrednio zbiór sytuacji społecznych), odzwierciedlają, w jaki sposób uczestnicy określonego zdarzenia komunikacyjnego postrzegają, interpretują i mentalnie odwzorowują te właściwości sytuacji społecznej, które są dla nich istotne.

<sup>20</sup> van Dijk T., *Ideology. A multidisciplinary approach*, Sage 1998, s. 211–212.

<sup>21</sup> *Ibidem*, s. 82.



Granice lokalności (czyli poziomu mikro) wyznaczone będą zatem poprzez określone wspólnoty komunikacyjne, w których pozostają jednostki. Wspólnota komunikacyjna jest oczywiście z założenia pojęciem szerszym niż bezpośrednia interakcja – uwzględnia bowiem jej historię, strukturę społeczną oraz modele kontekstowe, które wspólnota wykorzystuje.

Metoda badania tak rozumianej rzeczywistości społecznej w oczywisty sposób powinna odpowiadać charakterowi badanych zjawisk, czyli powinna umożliwiać uchwycenie procesu, w trakcie którego dochodzi do interpretacji rzeczywistości. Jak bowiem stwierdził Andrzej Piotrowski: „związek między teorią a metodą ma charakter implikacyjny: założenia o tym, czym jest rzeczywistość społeczna, wyznaczają wybór metody jej badania w tym sensie, że skłaniają do eliminacji procedur nie chwytających zjawisk istotnych na gruncie określonej teorii. I odwrotnie, wybór metody, owocujący określonym zbiorem danych i sposobów ich opracowania, przesądza w dużej mierze o zakresie zjawisk, o których badacz będzie mógł orzekać”<sup>22</sup>.

Przyjmując założenia o komunikacyjnej naturze lokalnej rzeczywistości społecznej, powinniśmy zatem wybierać takie metody, które umożliwiają nam odtworzenie procesów komunikacyjnych zachodzących na poziomie mikro. Zdaniem Moscovicio (który zaproponował teorię społecznych reprezentacji – swego rodzaju społecznie uzgodnionych i podzielanych teorii rzeczywistości społecznej<sup>23</sup>) powinniśmy: (a) dobierać materiał do badania tak, by reprezentował on próbę rozmów prowadzonych w społeczeństwie, (b) traktować reprezentacje społeczne jako środki odtworzenia rzeczywistości, (c) mieć na względzie, iż charakter reprezentacji ujawniany jest szczególnie wyraziście w czasie różnego rodzaju kryzysów i wstrząsów społecznych i (d) pamiętać, iż ludzi, którzy wypracowują reprezentacje, należy traktować jak naukowców-amatorów, a grupy, które tworzą, jako współczesne odpowiedniki stowarzyszeń naukowych, jakie istniały prawie wiek temu<sup>24</sup>. Program metodologiczny odpowiadający tym postulatom powinien umożliwiać badanie zjawisk społecznych, które powstają dopiero w trakcie i w wyniku współdziałania ludzi. „Jeśli chcemy badać nieprzerwany dialog, który wytwarza, modyfikuje i zmienia społeczne reprezentacje, to powinniśmy albo używać

---

<sup>22</sup> Piotrowski A., „Teoria a badanie empiryczne. Parę uwag o ich związku w orientacjach współczesnej socjologii” w: A. Giza-Poleszczuk, E. Mokrzycki (red.), *Teoria i praktyka socjologii empirycznej*, Warszawa: IFiS PAN 1990, s. 21.

<sup>23</sup> Więcej o teorii społecznych reprezentacji: Moscovici S., *Social representations. Explorations in social psychology*, Polity Press, Cambridge 2000; Trutkowski C., „Teoria społecznych reprezentacji i jej zastosowania” w: M. Marody (red.), *Wymiary życia społecznego. Polska na przełomie XX i XXI wieku*, Wydawnictwo Naukowe Scholar, Warszawa 2007.

<sup>24</sup> Moscovici S., „The phenomenon of social representations” w: R. Farr, S. Moscovici (red.), *Social Representations*, Cambridge University Press 1984, s. 52–59; patrz także: Flick U., „Social representations and the social construction of everyday knowledge: theoretical and methodological queries”, *Social Science Information* 1994, nr 33 (2).

metod przystosowanych do interpretacji naturalnie toczących się konwersacji albo też, jeśli to nie jest możliwe, powinniśmy próbować wytworzyć i podtrzymywać przejawy tego dialogu w wywiadach, w których nie kładzie się nacisku na standaryzację i duże próby ale pozostawia się badanym miejsce na rozwinięcie ich punktów widzenia i na [uchwycenie wykorzystywanych przez nich] sposobów odtwarzania (*re-create*) rzeczywistości lub tych jej części, które stanowią przedmiot badania”<sup>25</sup>.

Konieczne wydaje się tu podkreślenie, że zasadniczym celem prowadzonych badań jakościowych jest przede wszystkim zrozumienie świata osób badanych, a nie ustalenie zakresu rozpowszechnienia pewnych zjawisk czy konieczne określenie zewnętrznej ważności uzyskanych wyników. To właśnie poprzez zrozumienie sposobu rekonstrukcji zjawisk jesteśmy w stanie dojść do wniosków na temat postrzegania badanej rzeczywistości społecznej i nadawanych jej znaczeń, a tym samym określić możliwe i prawdopodobne sposoby jednostkowych i grupowych reakcji na wydarzenia w niej zachodzące. Rozumienie, o którym tu mowa, jest diltheyowskim rozumieniem hermeneutycznym, innymi słowy jest poszukiwaniem **wskazówek do interpretacji** rzeczywistości w wypowiedziach osób badanych, w przekazach płynących z mass mediów, czy w rozmowach prowadzonych w rodzinie i wśród znajomych<sup>26</sup>.

W świetle powyższych rozważań zapewne nie zaskoczy Czytelnika stwierdzenie, że jednym najważniejszych czynników odróżniających badania jakościowe od badań ilościowych (choć oczywiście nie jedynym) jest, w moim przekonaniu, kwestia różnicy perspektyw, z których próbują one ogarnąć rzeczywistość społeczną oraz wynikających z nich konsekwencji. Metody ilościowe pozwalają spojrzeć na społeczeństwo *sensu largo* i analizować procesy na poziomie makro, niekoniecznie sprowadzalne do poziomu jednostkowego, czy grupowego. Metody jakościowe zaś, jako te, które nie pozwalają na agregację, są interpretacyjne i bardziej elastyczne, nadają się do analiz prowadzonych na poziomie mikro – są przystosowane do interpretacji kontekstów, w których żyją grupy społeczne i jednostki. Czy ten podział można zatem przedstawić w postaci dychotomii metody kontekstowe – metody pozakontekstowe? Niestety, nie do końca. Problemem przy takiej próbie systematyzacji będą te techniki badawcze, które korzystają np. z danych ilościowych, ale są ściśle uzależnione od kontekstu, jak np. analiza treści. Bamberger uważa, że w związku z tym warto myśleć o dwóch kontinuumach: jednym, na którym można ulokować techniki badawcze w zależności od stopnia kwantyfikowalności

<sup>25</sup> Flick U., op. cit., s. 189.

<sup>26</sup> Szerzej na ten temat: Trutkowski C., *Społeczne reprezentacje polityki*, Wydawnictwo Naukowe Scholar, Warszawa 2000; Trutkowski C., „Zogniskowane wywiady grupowe – próba teoretycznego uzasadnienia” w: CD-ROM XI *Ogólnopolski Zjazd Socjologiczny. Materiały ze Zjazdu*; PTS i Wyższa Szkoła Społeczno-Gospodarcza w Tyczynie 2000.

wyników (kontinuum jakościowe-ilościowe) i drugim, które uwzględniałoby przystosowanie technik badawczych do analizy lokalnych kontekstów<sup>27</sup>. Wydaje mi się, że to drugie kontinuum jest bardziej czytelne, jako że uwzględnia nie tylko charakter gromadzonych danych, ale także bierze pod uwagę cele realizacji badań. „Metody kontekstowe próbują zrozumieć ludzkie działania w ramach społecznego, kulturowego, ekonomicznego i politycznego środowiska lokalności – zazwyczaj wioski, sąsiedztwa czy grupy społecznej, natomiast metody pozakontekstowe abstrahują od szczegółów lokalności po to, aby określić prawidłowości (trendy) generalne”<sup>28</sup>.

Pojawia się w tym miejscu zasadne pytanie – czy przy takim ujęciu różnic pomiędzy wskazanymi podejściami badawczymi w ogóle możliwa jest jakakolwiek ich integracja, czy też są to perspektywy wzajemnie nieprzystające, przeznaczone do badania odmiennych zagadnień? W literaturze znaleźć można wiele modeli łączenia metod jakościowych i ilościowych w ramach jednego projektu badawczego<sup>29</sup>. Proponowane podejścia to m.in. równoległe wykorzystanie różnych technik do badania różnego typu zjawisk (w zależności od kontekstualności przedmiotu badania), sekwencyjne stosowanie różnych metod w celu wypracowania lepszych narzędzi pomiaru, zawężenia pola obserwacji, czy też pogłębienia uzyskanych wniosków oraz różnorodnie rozumiana triangulacja, uzasadniana najczęściej jako sposób zwiększania trafności i rzetelności wyników. Przykładem integracji metod kontekstowych i pozakontekstowych może być sytuacja, w której wyniki badania jakościowego służą zidentyfikowaniu zmiennych numerycznych, które następnie mogą być włączone do sondażu jako część badań realizowanych w perspektywie makro, a więc pozwalających na agregację i generalizację wyników.

Niestety zbyt często pole prób integracji metod jakościowych i ilościowych okazuje się polem minowym. Część badaczy zapomina bowiem, że triangulacja nie służy weryfikacji wniosków z badań realizowanych w ramach odmiennych perspektyw (mikro/makro), że interpretacji jakościowej nie należy stosować we wnioskowaniu z danych sondażowych (i odwrotnie), że uczestników badań jakościowych (np. wielu grup fokusowych) nie można traktować jak respondentów wyłonionych w próbie losowej, czy wreszcie, że rozbieżności wyników uzyskanych różnymi metodami powinny być interpretowane w odniesieniu do założeń, na których te metody się wspierają (patrz wyżej). Jak zauważa Bamberger „w większości projektów badawczych [o integracyjnym podejściu metodologicznym] rozbieżności pomiędzy triangulowanymi wnioskami są albo ignorowane (co jest najczęstszą praktyką), albo wyjaśniane

<sup>27</sup> Bamberger M., op.cit.

<sup>28</sup> Ibidem, s. 9.

<sup>29</sup> Warto zapoznać się z tekstem Mikołaja Jasińskiego i Michała Kowalskiego „Falszywa sprzeczność...”, op. cit.

w sposób nieprzekonujący; bardzo nietypowe jest znalezienie systematycznej analizy rozbieżności w uzyskanych wynikach”<sup>30</sup>.

Triangulacja w założeniu powinna wzmacniać trafność i rzetelność wyników badań dzięki wykorzystywaniu różnych źródeł danych, przede wszystkim w ramach tej samej perspektywy. Postulat ten w odniesieniu do metod jakościowych oznacza przede wszystkim korzystanie z różnych technik badawczych, konfrontowanie opinii różnych badanych osób lub powracanie do respondentów po upływie jakiegoś czasu i ponowne stawianie podobnych pytań. W ramach perspektywy ilościowej oznacza porównywanie wyników różnych pomiarów, weryfikację danych sondażowych w odniesieniu do danych zbieranych innymi metodami (np. z danymi ze spisu powszechnego), czy testowanie hipotez przy wykorzystaniu różnych technik analitycznych.

Pozostaje jeszcze jedna kwestia związana ze stosowaniem metod jakościowych – mianowicie ich stosunkowo niewielka akulturacja. Wyniki badań ilościowych, głównie dzięki szerokiemu rozpowszechnieniu badań opinii publicznej<sup>31</sup>, zyskały legitymację i znacznie większe uznanie wśród decydentów (np. polityków, przedstawicieli administracji, czy biznesu) niż metody jakościowe. Te ostatnie często są wybierane nie ze względu na ich dostosowanie do problemu badawczego, ale ze względu na niższe koszty realizacji! W efekcie wielu zamawiających badania, czy decydujących o kształcie zamawianego projektu, traktuje metody jakościowe jako tańszy zamiennik badań „prawdziwych”, wybierany raczej z braku odpowiednich możliwości, a nie z uzasadnionej potrzeby. Cytowany już przeze mnie Bamberger zauważa (w świetle doświadczeń wyniesionych z pracy dla Banku Światowego), że nawet wtedy, gdy „decydenci sami zgadzają się z wynikami badań jakościowych, to często chcą, aby wyniki jakościowe były prezentowane z wykorzystaniem konwencjonalnych zasad statystycznych, po to, aby uczynić je bardziej akceptowalnymi dla innych agencji. Zaspokojenie potrzeby zwiększenia rzetelności wyników badań jakościowych poprzez użycie odpowiednich technik statystycznych jest często ważne dla zwiększenia akceptowalności i wykorzystania tychże wyników”<sup>32</sup>.

W praktyce badań społecznych realizowanych poza środowiskiem akademickim (np. w przypadku wielu badań ewaluacyjnych) często zdarza się, że elementem legitymizującym wykorzystanie słabo akulturowanych metod jakościowych jest osoba badacza (eksperta, konsultanta, ewaluatora itp). Swoim statusem, doświadczeniem i opanowaniem warsztatu metodologicznego często uprawomocnia on realizację badań, których istota i sens nie do końca są uświadomione. Dodać przy tym należy, że uwaga ta nie dotyczy raczej

<sup>30</sup> Bamberger M., op.cit., s. 22.

<sup>31</sup> Zob. Sułek A., *Sondaż polski*, Wydawnictwo IFiS PAN, Warszawa 2001.

<sup>32</sup> Bamberger M., op.cit., s. 7.

sfery badań marketingowych – tam, jak słusznie zauważają Górniak, Worek i Krupnik „przydatność tego typu badań doceniono już dawno (...) ich popularność systematycznie wzrasta, a wachlarz wykorzystywanych technik staje się coraz szerszy”<sup>33</sup>.

## Badania jakościowe a specyfika ewaluacji

Wielu autorów podkreśla, że podstawowym celem stosowania metod jakościowych w ewaluacji jest zrozumienie procesu działania interwencji i opis sposobów wdrażania podejmowanych działań w określonym otoczeniu społecznym, instytucjonalnym oraz politycznym. W odróżnieniu, metody ilościowe, zdaniem Bambergera, nie służą najlepiej „zrozumieniu środowiska socjoekonomicznego, procesów instytucjonalnych i politycznych, czy tego, jak różne gospodarstwa domowe w różnych kontekstach kulturowych odpowiedzą na określony [skierowany do nich] projekt”<sup>34</sup>. Należy bowiem pamiętać, że jakościowe spojrzenie na projekt jest zawsze spojrzeniem na silnie skontekstualizowaną rzeczywistość lokalną. Tak jak metody ilościowe są często niezbędne dla określenia szerszego oddziaływania interwencji, tak metody jakościowe służą z jednej strony funkcjom formatywnym w trakcie trwania projektu oraz zrozumieniu czynników wpływających na jego końcowe efekty. W szerszym ujęciu badania jakościowe często umożliwiają zatem zdobycie wiedzy i wypracowanie rekomendacji dotyczących sposobów wdrażania określonych typów interwencji w określonych warunkach.

Generalnie, podstawowym celem badań jakościowych jest dogłębne zrozumienie analizowanego zjawiska. W badaniach akademickich służy to poznaniu (odtworzeniu) społecznie wytworzonych i podzielanych przekonań na temat określonej rzeczywistości w celu zrozumienia zachodzących zjawisk społecznych, w badaniach marketingowych natomiast pozwala odtworzyć sposób postrzegania jakiegoś produktu przez konsumentów, rozpoznać ich preferencje i wreszcie dostosować określoną ofertę do ich wyobrażeń, potrzeb i oczekiwań. Badania ewaluacyjne mają natomiast zasadniczo inny cel – powinny one przyczynić się do ustalenia rzeczywistego wyniku interwencji oraz dokonać jego wartościującej oceny<sup>35</sup>. To prawda, że owa ocena jest efektem końcowym często wielowątkowego procesu badawczego, w którym metody jakościowe mogą stanowić jedynie część zastosowanego instrumentarium badawczego. Prawdą jest również, że procedury związane z wykorzystaniem metod jakościowych powinny być zasadniczo takie same w ewaluacji, jak

<sup>33</sup> Górniak J., Worek B., Krupnik S., op. cit., s. 117.

<sup>34</sup> Ibidem, s. 21.

<sup>35</sup> Por. Haber A., Kryszczuk M.D., „Wybrane problemy doboru technik badawczych w ewaluacji *ex-post*” w: A. Haber (red.), *Ewaluacja ex-post...*, op. cit., s. 140.

w badaniach akademickich, czy rynkowych. Ale specyfiką ewaluacji w przeciwieństwie do akademickich badań społecznych jest jednak wartościowanie. Realizacja badań ewaluacyjnych nie jest zatem czystym odtwarzaniem i poznawaniem rzeczywistości w celu zrozumienia procesów w niej zachodzących. Jest to wszak zawsze realizacja z tezą, z odniesieniem do wcześniej ustalonych (przyjętych lub narzuconych) kryteriów oceny.

I tu nieuchronnie rodzi się sprzeczność pomiędzy uniwersalnymi założeniami leżącymi u podstaw metodologii jakościowej a wymogami wynikającymi z celów ewaluacji. Otóż dogłębne, holistyczne zrozumienie badanego zjawiska musi zakładać próbę spojrzenia na rzeczywistość lokalną (skala mikro) z perspektywy jej uczestników i przy użyciu podzielanych przez nich, istotnych wymiarów postrzegania tej rzeczywistości. A wymiary te niekonięcznie przecież pokrywają się z przyjętymi przez badacza (lub narzuconymi mu) kryteriami oceny jakiejś interwencji. Maciej Szałaj bardzo trafnie wydobywa ten paradoks, pisząc: „oddziaływanie doboru kryteriów widać chociażby w metodach badawczych i sposobie gromadzenia dowodów, gdyż określa on rodzaje poszukiwanych danych (...). Określana przez kryteria oceny perspektywa może stanowić źródło legitymizacji dla badań, których **poprawność metodologiczna nie poddaje się prostej ocenie z wykorzystaniem zwyczajowych standardów naukowości** (podkreślenie C.T.). Obecność kryteriów oceny w kontekście badawczym jest, ogólnie rzecz biorąc, najbardziej chyba wyraźną ilustracją zaangażowanych w projekt wartości. Pokazuje to dobrze różnice pomiędzy ewaluacją a dziedzinami akademickimi, dostarczającymi ewaluacji inspiracji teoretycznych, narzędzi badawczych i wreszcie kadr”<sup>36</sup>.

W przeważającej większości badań ewaluacyjnych nie stosujemy zatem holistycznego ujęcia lokalnej rzeczywistości, ale raczej próbujemy ją przeanalizować ze względu na przyjęte wcześniej, predefiniowane kryteria oceny<sup>37</sup>. Dokonujemy tym samym sztucznego cięcia, abstrahując od sposobów

<sup>36</sup> Szałaj M. „Mechanizmy doboru kryteriów oceny w ewaluacji *ex-post*” w: A. Haber (red.), *ewaluacja ex-post...*, op. cit., s. 62.

<sup>37</sup> Istnieje jednak grupa badań ewaluacyjnych realizowana w ramach orientacji „ewaluacji nie opartej na celach” (ang. *goal free evaluation* – *GFE*). Przyjmuje się w niej odmienne podejście od prezentowanego w tym tekście (jako najbardziej rozpowszechnionego). Mianowicie zakłada się, że ewaluator przed rozpoczęciem ewaluacji nie odtwarza zakładanych celów interwencji i tym samym nie przyjmuje kryteriów jej oceny, lecz dopiero w trakcie ewaluacji ustala, jakie są jej skutki. Słowem dopiero w trakcie badań ewaluator dowiadyuje się, co wdrażana interwencja powoduje w określonej rzeczywistości zamiast sprawdzania, czy powoduje to, co wcześniej zamierzano. W ramach *GFE* wartość ustalana jest poprzez odniesienie efektów programu do rzeczywistych potrzeb populacji, do której jest on adresowany a nie poprzez jej *a priori* określenie. Więcej na ten temat Czytelnik znajdzie m.in. w tekstach: Scriven M., *Evaluation Thesaurus*, Sage 1991; Scriven M., „The pros and cons about goalfree evaluation”, *American Journal of Evaluation*, Vol. 12, No. 1, s. 55–62 (1991); Worthen B., „Program evaluation”, w: H. Walberg, G. Haertel (red.), *The international encyclopedia of educational evaluation*, Pergammon Press, Toronto, ON 1990.

percepcji reprezentowanych przez osoby badane. Poszukujemy wskazówek do wartościującej interpretacji zjawisk, postępując przy tym niejako w zgodzie z hermeneutycznym postulatem Schleiermachera, aby próbować zrozumieć autora równie dobrze albo lepiej, niż on sam siebie rozumie<sup>38</sup>. „Lepiej” oczywiście w odniesieniu do kryteriów oceny, które my przyjęliśmy. Sytuacja ta przypomina trochę problem demona Laplace’a, który dysponując kompletną wiedzą o położeniu wszystkich cząstek elementarnych wszechświata oraz wszystkich siłach na nie działających, byłby w stanie wskazać optymalną drogę do dowolnego wybranego celu, ale... nie mógłby wybrać tego celu, gdyż z wyborem takim wiążą się jakieś wartości. Badania jakościowe lokalnej rzeczywistości, np. przebiegu określonej interwencji, są w stanie dostarczyć odpowiedzi na pytanie, czy wdrażanie projektu było realizowane skutecznie i efektywnie, słowem, czy wybrano optymalną drogę do osiągnięcia zakładanych celów. Ale same w sobie nie są w stanie uzasadnić wyboru tych a nie innych kryteriów oceny realizacji danego projektu. Wyobraźmy sobie sytuację, w której jakiś program wdrażany był z dużymi problemami i w efekcie osiągnięte rezultaty nie są optymalne. Czy to oznacza, że jego ewaluacja powinna zakończyć się wynikiem negatywnym? Oczywiście odpowiedź brzmi: to zależy od przyjętych kryteriów. Na przykład konkluzją mogłoby być negatywne stwierdzenie, że program był realizowany nieskutecznie i nieefektywnie albo pozytywna konstatacja, że dzięki zrealizowaniu celów (nawet nieoptymalnemu) udało się uniknąć konieczności zwrotu środków europejskich bądź udało się przygotować grunt pod kontynuację reformatorskich działań w jakimś obszarze.

Owa, prezentowana tu, specyfika badań ewaluacyjnych wymaga od badacza-ewaluatora szczególnej uwagi. W procesie ewaluacji mamy bowiem do czynienia z dokonywaniem ocen na podstawie wybranych wartości. Proces badawczy powinien w związku z tym być szczególnie przejrzysty i czytelny dla wszystkich zaangażowanych w projekt stron. Odnosi się to w najwyższym stopniu do badań jakościowych, z natury swojej, jak już wspomnieliśmy, subiektywnych, mniej ustrukturyzowanych i gorzej akulturowanych niż badania ilościowe. Dobrym sposobem zapewnienia przejrzystości procesu badawczego i prezentacji podstaw wypracowanych wniosków jest zaangażowanie realizatorów projektu w dyskusję nad uzyskanymi wynikami.

Wreszcie ostatnim zagadnieniem, które chciałbym poruszyć przy okazji prezentowanych tu rozważań, jest kwestia współczesnej praktyki metodologicznej w projektach ewaluacyjnych realizowanych w Polsce. Jeśli bowiem, jak stwierdziliśmy, uniwersalne kanony dotyczące rozplanowania badań jakościowych nie zostały dotąd wypracowane, to skąd się biorą określone metody i schematy badawcze w ofertach przetargowych? Jakie są generalnie źró-

<sup>38</sup> Schleiermacher F., *Hermeneutik*, Heidelberg 1959; zob. także Ricoeur P., *Język, tekst, interpretacja*, PIW, Warszawa 1989, s. 196.

dła decyzji o wyborze określonych modeli realizacji badań (np. dotyczących zakresu wykorzystania metod jakościowych i ilościowych oraz rodzaju stosowanych technik badawczych)? Wynikiem obserwacji badań ewaluacyjnych w praktyce musi być nieuchronna konstatacja, że nie zawsze kanony metodologiczne (np. opisywana w tym tekście specyfika metod i związanych z nimi założeń dotyczących badanej rzeczywistości) są tu czynnikiem decydującym. Wydaje się, że przynajmniej kilka kwestii wpływa na ostateczny kształt realizowanych badań.

Po pierwsze, badania jakościowe są po prostu tańsze w realizacji od badań ilościowych. Przeciętny projekt jakościowy przeważnie będzie wymagał niższego budżetu niż sondaż realizowany na próbie reprezentatywnej. Jeśli dodamy do tego, wspomniany już, brak sztywnych kanonów realizacji badań jakościowych, to jeszcze wyraźniej zobaczymy potencjalne źródło oszczędności dla zleceniobiorców ewaluacji. Odmienne bowiem, niż w badaniach ilościowych, przy realizacji serii wywiadów pogłębionych nie jesteśmy przecież w stanie zmierzyć wielkości błędu (uzależnionego wszak m.in. od wielkości próby). A realizacja sześciu wywiadów zamiast dziesięciu da się jakoś uzasadnić (np. często stosowanym pojęciem – wytrychem „nasylenia teoretycznego”<sup>39</sup>).

Po drugie, jak słusznie zauważają Haber i Kryszczuk, „instytucje mają skłonność do instytucjonalizacji pewnych praktyk badawczych, które są z biegiem czasu doskonalone i nabierają charakteru odrębnych kanonów metodologicznych”<sup>40</sup>. Samo w sobie nie musi to oczywiście oznaczać zmniejszenia jakości realizowanych ewaluacji. Pojawia się tu jednak niewątpliwie groźba zaniku refleksji teoretycznej nad wyborem oferowanych metod, co w następstwie rodzić może niebezpieczeństwo braku świadomości konsekwencji owego wyboru dla wypracowanych wniosków. Dodajmy jeszcze, że odrębne kanony metodologiczne mogą być charakterystyczne tak dla instytucji, jak i dla poszczególnych konsultantów–ewaluatorów.

Po trzecie wreszcie, zdarza się że sama instytucja zamawiająca ewaluację określa rodzaj metod badawczych, które mają być w jej ramach zrealizowane. I tu również możemy mieć do czynienia zarówno z względami budżetowymi, jak i zinstytucjonalizowaniem określonego kanonu badawczego. Zdarzały się także kuriozalne przypadki, gdy o wyborze wykonawcy ewaluacji decydowała różnorodność i ilość zaproponowanych technik badawczych (im więcej zaproponowanych technik tym lepsza ocena oferty), a nie konieczność uмотywowania ich zastosowania.

<sup>39</sup> Por. Helling K., op. cit.

<sup>40</sup> Ibidem, s. 139.



## Podsumowanie

Przywołana przeze mnie debata o wyższości metod ilościowych nad jakościowymi (lub odwrotnie) w historii rozwoju badań ewaluacyjnych stanowiła jedno z głównych pól sporu ewaluatorów. Z biegiem czasu szczęśliwie przerodziła się ona w poszukiwania możliwości integracji obu podejść i jako taka nie straciła ani na znaczeniu, ani na intensywności. Integracja metod ilościowych i jakościowych nie jest bowiem rzeczą prostą – z pewnością nie da się jej osiągnąć pod atrakcyjnymi, hasłami prostej triangulacji czyli postulatu badania tego samego zjawiska różnymi metodami i porównywania wyników w celu minimalizacji błędu. Należy bowiem pamiętać, że metody jakościowe i metody ilościowe służą do badania różnych zjawisk – wykorzystanie jednych bądź drugich oznacza przyjęcie odmiennych perspektyw oglądu, prowadzenie analiz w odmiennej skali, odmienny stosunek do zmiennych kontekstowych i wreszcie odmienny poziom szczegółowości opisu.

Warto podkreślić, że stosowanie metod jakościowych i ilościowych w ewaluacji oznacza szereg odmiennych konsekwencji wynikających z dokonywanego wyboru.

Celem stosowania metod jakościowych jest osiągnięcie dogłębnego zrozumienia badanego zjawiska. Metody te oznaczają przyjęcie perspektywy mikro i realizację badań na poziomie lokalnym (instytucji, grupy, wspólnoty), co w efekcie prowadzi do holistycznego ujęcia przedmiotu badania (np. określonej interwencji, wybranego procesu społecznego, wydarzenia itp.). Możliwości ekstrapolacji ustaleń badań jakościowych są jednak mocno ograniczone – zazwyczaj wyniki można odnosić wyłącznie do bezpośrednio badanej rzeczywistości lokalnej. Istnieją co prawda podejścia teoretyczne (jak wspomniana tu już wcześniej teoria społecznych reprezentacji), uzasadniające możliwość ograniczonej generalizacji wniosków – mają one jednak zawsze charakter teoretyczny i raczej nieweryfikowalny. Realizując badania jakościowe, a następnie prezentując ich wyniki, należy zatem zawsze pamiętać o konieczności rygorystycznego zachowania systematyczności procesu badawczego – wyczerpującego uzasadnienia wyboru technik, doboru respondentów, czasu i miejsca badania, sposobu analizy materiału i źródeł zgłaszanych rekomendacji. Wszystko po to, aby uniknąć jakże częstych pod adresem badań jakościowych zarzutów uprawiania publicystyki i przedstawiania subiektywnych sądów.

Z drugiej strony, metodom ilościowym przynależna jest perspektywa makro, poszukiwanie ogólnych prawidłowości i możliwość generalizacji wyników. Kontekst lokalny jest tu nieistotny, a szczegółowość analizy pojedynczych przypadków zasadniczo niewskazana. Badania ilościowe, o ile zrealizowane zgodnie z zasadami sztuki, umożliwiają uzyskanie danych weryfikowalnych, niezależnych od subiektywizmu badacza. Są to zatem dane „twarde” i porównywalne.

Badacz ilościowy, w przypadku niewłaściwego zdefiniowania przedmiotu i/lub obiektów badania, może się po pierwsze narazić na zarzut powierzchowności, a po drugie na oskarżenie o tworzenie artefaktów. Niebezpieczeństwo powierzchowności związane jest z faktem, że nie zawsze prawidłowości generalne, odnoszące się do skali makro, mogą być sprowadzone do poziomu jednostkowego (skali mikro, rzeczywistości lokalnej). Artefakty zaś rodzą się z zadawania niewłaściwych pytań lub pytania niewłaściwych ludzi. Wielu metodologów podkreślało, że jeśli zadajemy respondentom pytanie, to musimy liczyć się z faktem, że uzyskamy na to pytanie odpowiedź. Odpowiedź zaś może paść nawet wtedy, gdy respondent pytania nie rozumie, nie ma wiedzy, aby jej udzielić, albo nie chce nam jej udzielić. Ankieter powinien taką odpowiedź zapisać i nie ma możliwości dopytać (jak np. w trakcie wywiadu pogłębionego), co respondent miał na myśli, wybierając określoną odpowiedź spośród kategorii przewidzianych przez badacza.

Integracja obu podejść – jakościowego i ilościowego – w ramach badania społecznego, czy ewaluacyjnego jest możliwa pod warunkiem zachowania świadomości powyżej zarysowanych różnic i konsekwencji z nich wynikających. Przede wszystkim należy pamiętać o nieprzekładalności perspektyw, właściwych obu orientacjom metodologicznym – zasadniczo trudno znaleźć uzasadnienie dla wykorzystania metod jakościowych i ilościowych do badania tego samego zjawiska (choć można je wykorzystywać do badania różnych zjawisk w ramach określonej lokalności).

Na koniec chciałbym jeszcze raz podkreślić, że jakościowe badania społeczne (akademickie) i jakościowe badania ewaluacyjne to w gruncie rzeczy dwa dość odrębne podejścia. Celem pierwszych jest wszak poznanie prawdy o rzeczywistości (nawet jeśli jest to jedynie prawda lokalna) i przez to zrozumienie procesów w niej zachodzących. Celem drugich natomiast jest ocena jakiejś rzeczywistości lokalnej (np. realizowanego programu) ze względu na przyjęte *a priori* kryteria. I choć z pewnością możemy mówić o jednakowo intensywnych w obu podejściach wysiłkach na rzecz obiektywizacji ustaleń, to w ewaluacji będzie to zawsze próba obiektywizacji w odniesieniu do określonych wartości. Wartościowanie jest zatem tym czynnikiem, który przede wszystkim odróżnia jakościowe badania ewaluacyjne od takich samych badań akademickich.

---

**Cezary Trutkowski** – doktor nauk społecznych, adiunkt w Zakładzie Psychologii Społecznej Instytutu Socjologii UW. Jest konsultantem i trenerem, autorem publikacji naukowych, nauczycielem akademickim i doświadczonym badaczem społecznym. Przez wiele lat zarządzał projektami budowania potencjału instytucjonalnego samorządów te-

rytorialnych. Jako ekspert Rady Europy, wspomagał rozwój samorządów w Albanii, Armenii, Chorwacji, Czarnogórze, Macedonii, Mołdawii, Rosji, Serbii i Słowenii. W Polsce nadzorował komponent szkoleniowy w ramach Programu Aktywizacji Obszarów Wiejskich (gdzie odpowiadał m.in. za wewnętrzny monitoring i ewaluację projektu), przez wiele lat uczestniczył jako trener w Programie Przeciw Korupcji Fundacji im. Stefana Batorego i pracował jako ekspert w programie „Przejrzysta Polska”. Ma polskie i międzynarodowe doświadczenia w projektach oceny potrzeb szkoleniowych oraz ewaluacji programów rozwojowych i edukacyjnych (pracował dla Rady Europy, Banku Światowego, UNDP oraz polskich instytucji rządowych i pozarządowych). Przez osiem lat prowadził badania rynku i opinii publicznej w firmie, której był współwłaścicielem. Obecnie odpowiada za monitoring i ewaluację Programu Integracji Społecznej w ramach Poakcesyjnego Programu Wsparcia Obszarów Wiejskich (realizowanego przez MPiPS w ramach pożyczki Banku Światowego).

## Bibliografia

- Antoszkiewicz J., *Metody heurystyczne. Twórcze rozwiązywanie problemów*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1990.
- Babbie E., *Badania społeczne w praktyce*, Wydawnictwo Naukowe PWN, Warszawa 2003.
- Bamberger M., „Opportunities and challenges for integrating quantitative and qualitative research”, w: M. Bamberger (red.), *Integrating quantitative and qualitative research in development projects*, World Bank, Washington, DC 2000.
- Chung K., „Issues and approaches in the use of integrated methods”, w: M. Bamberger (red.), *Integrating quantitative and qualitative research in development projects*, World Bank, Washington, DC 2000.
- van Dijk T., *Ideology. A multidisciplinary approach*, Sage 1998.
- Flick U., „Social representations and the social construction of everyday knowledge: theoretical and methodological queries”, *Social Science Information* 1994, nr 33 (2).
- Górnjak J., Worek B., Krupnik S., „Zastosowanie podejścia badań jakościowych w ewaluacji *ex-post*”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.
- Gouldner A., „Teoretyczny kontekst socjologii”, w: E. Mokrzycki, *Kryzys i schizma*, PIW, Warszawa 1984.
- Haber A., Kryszczuk M.D., „Wybrane problemy doboru technik badawczych w ewaluacji *ex-post*”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.
- Helling I.K., „Metoda badań biograficznych”, w: *Metoda biograficzna w socjologii*, J. Włodarek, M. Ziółkowski, PWN, Warszawa 1990.
- Jasiński M., Kowalski M., „Falszywa sprzeczność: metodologia jakościowa czy ilościowa?”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.
- Maison D., *Zogniskowane wywiady grupowe. Jakościowa metoda badań marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 2001.
- Miller G., Dingwall R. *Analyzing Social Settings*, Wadsworth Publishing, Belmont, CA 1997.
- Morgan D.L., „Focus groups”, *Annual Review of Sociology* 1996, vol. 22.
- Morgan D.L., *Successful focus groups. Advancing the state of the art*, Sage 1993.
- Moscovici S., *Social representations. Explorations in social psychology*, Polity Press, Cambridge 2000.

- Moscovici S., „The phenomenon of social representations”, w: Farr R., Moscovici S. (red.), *Social Representations*, Cambridge University Press 1984.
- Piotrowski A., „Teoria a badanie empiryczne. Parę uwag o ich związku w orientacjach współczesnej socjologii”, w: A. Giza-Poleszczuk, E. Mokrzycki (red.), *Teoria i praktyka socjologii empirycznej*, IFiS PAN, Warszawa 1990.
- Ricoeur P., *Język, tekst, interpretacja*, PIW, Warszawa 1989.
- Rubacha K. *Metodologia badań nad edukacją*, Wydawnictwa Akademickie i Profesjonalne 2008.
- Rubin H.J., Rubin, I.S., *Qualitative interviewing: The art of hearing data*, Sage Publications, Thousand Oaks, CA 2004.
- Sadish W.R., „Evaluation theory is who we are”, *American Journal of Evaluation* 1998, vol. 19, nr 1.
- Schleiermacher F., *Hermeneutik*, Heidelberg 1959.
- Schwandt T.A., *Qualitative inquiry*, Sage, Thousand Oaks, CA 1997.
- Scriven M., *Evaluation Thesaurus*, Sage 1991.
- Scriven M., „The pros and cons about goal-free evaluation”, *American Journal of Evaluation* 1991, Vol. 12, No. 1.
- Silverman D., *Interpretacja danych jakościowych*, PWN 2008.
- Straś-Romanowska M. (red.), *Metody jakościowe w psychologii współczesnej*, Wydawnictwo Uniwersytetu Wrocławskiego 2000.
- Sulek A., *Sondaż polski*, Warszawa: Wydawnictwo IFiS PAN 2001.
- Szałaj M. „Mechanizmy doboru kryteriów oceny w ewaluacji *ex-post*” w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.
- Trutkowski C., „Teoria społecznych reprezentacji i jej zastosowania” w: M. Marody (red.), *Wymiary życia społecznego. Polska na przełomie XX i XXI wieku*, Wydawnictwo Naukowe Scholar, Warszawa 2007.
- Trutkowski C., „Zogniskowane wywiady grupowe – próba teoretycznego uzasadnienia” w: CD-ROM XI Ogólnopolski Zjazd Socjologiczny. *Materiały ze Zjazdu*; PTS i Wyższa Szkoła Społeczno-Gospodarcza w Tyczynie 2000.
- Trutkowski C., *Społeczne reprezentacje polityki*, Wydawnictwo Naukowe Scholar, Warszawa 2000.
- Worthen B., „Program evaluation”, w: H. Walberg, G. Haertel (red.), *The international encyclopedia of educational evaluation*, Pergammon Press, Toronto, ON 1990.



*Tomasz Kaczor*

## **Modele ekonomiczne w ewaluacji, czyli o czym jej odbiorca wiedzieć powinien**

Ewaluacja, szczególnie w polskich warunkach, wydaje się nie doceniać roli i możliwości modeli ekonomicznych. Oceniając przeprowadzone w Polsce ewaluacje<sup>1</sup>, zobaczymy, że bardzo niewiele z nich wykorzystuje je jako narzędzie analizy. Można nawet odnieść wrażenie, że pojawiają się wyłącznie tam, gdzie trudno sobie wyobrazić użycie innych technik, są więc swego rodzaju narzędziem ostatecznej szansy. Jest ku temu kilka powodów. Modele na ogół nie są intuicyjnie zrozumiałe. Ich stosowanie oraz interpretacja wyników wymaga zwykle specjalistycznej wiedzy, często dotyczącej nawet specyficznej klasy modeli. Drugą, istotną przyczyną pomijania modeli jest fakt, że często są prezentowane w sposób, w którym sam model staje się ważniejszy od wyniku i jego interpretacji, co stoi w sprzeczności z samą ideą ewaluacji. Trzecią wreszcie przyczyną jest niewielka, w porównaniu z innymi technikami ewaluacyjnymi, znajomość potencjalnych możliwości tego narzędzia pośród odbiorców ewaluacji. To wszystko sprawia, że potencjał narzędzi, jakimi mogą być modele nie jest w pełni wykorzystywany.

Poniżej chcemy powiedzieć nieco o możliwych zastosowaniach modeli – czemu poświęcono pierwszą część rozdziału – oraz o znaczeniu różnych elementów badań modelowych dla właściwego odbioru wyników, co znajduje się w części drugiej. W kolejnej opisano niektóre z ograniczeń modelowania, a na samym końcu Czytelnik znajdzie tendencyjny „niezbędnik” tego, co powinna zawierać ewaluacja wykorzystująca model ekonomiczny.

### **Modele, ich rodzaje i zastosowania**

Ewaluacja – w tym znaczeniu, w jakim będziemy ją stosować w tym tekście – jest szczególną formą tradycyjnych badań społeczno-ekonomicznych. Wy różnia ją przede wszystkim specyficzny cel, w jakim jest prowadzona, którym

---

<sup>1</sup> Np. zawarte w bazie danych ewaluacyjnych MRR <http://www.fundusze-strukturalne.gov.pl/EWALUACJA+funduszy+w+Polsce/Baza+badan+ewaluacyjnych/>

jest wsparcie procesu decyzyjnego. Nie dziwi zatem, że ewaluacja często korzysta z narzędzi stworzonych na potrzeby innego rodzaju prac badawczych i tak właśnie stało się z modelami ekonomicznymi, których specyficzne cechy czynią je naturalnym orężem ewaluatora. W zasadzie dla precyzji należałoby użyć terminu „ilościowy model ekonomiczny”, bowiem przez model rozumieć będziemy formalne narzędzie oparte na zapisie ekonomicznych lub społecznych zależności w postaci równań matematycznych, których elementy stanowią ilościowe wyrażenie analizowanych zjawisk. O ich przydatności w ewaluacji decydują dwie cechy – mierzalny charakter dostarczanych wyników oraz relatywnie niewielka doza arbitralności. Pierwsza jest niemal oczywista (podobnie jak korzyści z uzyskania wyników, które można na przykład przedstawić na wykresie) i wynika wprost z zaproponowanej definicji modelu. Druga jest pochodną sztywnego stosowania reguł: zależności i założeń zapisane w równaniach działają bezwzględnie (lecz trzeba zwrócić uwagę, że są one tak samo bezkrytyczne jak bezwzględne). Pamiętajmy jednak, że sam dobór założeń jest procesem arbitralnym i z tego względu powinien być przedmiotem czujnej uwagi, o czym więcej w dalszej części tekstu. Warto jeszcze wspomnieć o dodatkowej zalecie badań wykorzystujących modelowanie ekonomiczne. W wielu przypadkach możliwe jest ustalenie, oprócz samych wartości interesujących odbiorcę ewaluacji, także precyzji uzyskanego oszacowania.

Przed wnikiem w zastosowania tego typu narzędzi do celów ewaluacyjnych warto poświęcić nieco miejsca ich najważniejszym typom. Modele ekonomiczne są grupą daleką od jednorodności i niełatwo poddającą się klasyfikacji, jednak wyraźnie rysują się dwa podziały. Kryterium pierwszego jest obszar objęty modelowaniem i z tego punktu widzenia każdy model będzie albo całościowym, albo częściowym. Model całościowy opisuje komplet zależności czy procesów dotyczących analizowanego organizmu, np. gospodarki kraju, regionu lub przedsięwzięcia. Druga grupa skupia się na wybranych spośród nich obszarach, najistotniejszych z punktu widzenia celów analizy.

Podkreślmy od razu, że pominięcie mniej istotnych zależności nie powoduje, że modele częściowe są w jakiś sposób ułomne. To, którego z nich warto użyć, definiuje charakter przedmiotu ewaluacji (czyli tego, w jaki sposób ma działać ewaluowany projekt lub program) oraz zakres efektów, które mają być obserwowane. Tam, gdzie ewaluacja dotyczy szeregu działań dotyczących różnych obszarów, a wykorzystywane instrumenty mogą wzajemnie wpływać na swoje funkcjonowanie i potrzebny jest opis skutków dla różnych charakterystyk badanego organizmu, naturalnym wyborem będzie narzędzie całościowe. Natomiast tam, gdzie mechanizm działania jest relatywnie prosty, a interakcje z otoczeniem ograniczone, kuszącym wyborem jest model częściowy, relatywnie prosty, szybszy w tworzeniu i mający mniejsze wymagania. Za cenę nieszkodliwego uproszczenia opisu oddziaływań otrzy-

mujemy znacznie większą detaliczność wyników. Podkreślimy też naturalną komplementarność modeli całościowych i cząstkowych w niektórych sytuacjach. Te ostatnie mogą służyć jako narzędzia uzupełniające, pozwalające nadać wybranym wynikom uzyskanym dzięki modelowi całościowemu większą szczegółowość.

Drugą, równie istotną linią demarkacyjną, rozróżniającą modele, jest stopień, w jakim są oparte na jawnie określonej teorii ekonomicznej. Z jednej strony stoją modele, w których badacz specyfikuje zaledwie podstawowy charakter relacji między zmiennymi (np. liczbą miejsc pracy a nakładami na różne polityki rynku pracy). W ich przypadku precyzyjna natura zależności ustalana jest przy zastosowaniu technik statystycznych i ekonometrycznych na podstawie danych historycznych. Z drugiej strony występują narzędzia, w których ewaluator określa szczegółową postać relacji, wykorzystując teorię ekonomiczną czy wyniki innych badań. Zalety pierwszego typu to doskonałe dopasowanie do danych i możliwość zastosowania tam, gdzie teoria nie daje jasnych przesłanek co do charakteru zależności. Wadą jest konieczność dostępu do dużej ilości danych historycznych i ograniczone możliwości interpretacji wyników<sup>2</sup>.

Tabela 1. Podział modeli ze względu na...

... kompletność opisu rzeczywistości	
Modele całościowe – w sposób kompletny opisują relacje ewaluowanego organizmu gospodarczego	np.: modele makroekonometryczne, modele równowagi ogólnej
Modele cząstkowe – opisują wybrane mechanizmy istotne z punktu widzenia działania przedmiotu ewaluacji	np: jedno- i wielorównaniowe modele ekonometryczne, modele równowagi cząstkowej
... rolę danych i teorii	
Modele oparte na teorii ekonomicznej definiującej dokładny kształt zależności między zmiennymi	np: modele równowagi ogólnej, niektóre modele ekonometryczne
Modele, w których teoria mówi o podstawowym charakterze zależności, a jej szczególny charakter ujawniają dane	np: modele VAR, ARIMA, niektóre modele ekonometryczne

Źródło: Opracowanie własne.

Tych ograniczeń nie ma drugi typ modeli – wyniki są zawsze łatwe do wyjaśnienia w oparciu o bazową teorię i mogą być stosowane w odniesieniu do nowych zjawisk i procesów, o ile założenia teoretyczne do nich pasują.

<sup>2</sup> Mniejsze możliwości interpretacji wynikają z braku dokładnej informacji o tym, jaki jest szczegółowy charakter zależności powodujących, że przedmiot ewaluacji działa w taki a nie inny sposób i czy te relacje będą identyczne w odmiennych warunkach.



Tabela 2. Zastosowania modeli oraz uzyskiwane wyniki

Zastosowania modeli	Przykładowe wyniki i zastosowania
Opis stanu faktycznego i wpływu rzeczywistych polityk, programów i projektów: jakie zmiany zaszyły na skutek ewaluowanego programu i dlaczego	wartości wybranych mierników (dynamik, struktur, poziomów) rzeczywistych oraz hipotetycznych w sytuacji niezrealizowania programu (np. PKB per capita, stopa inwestycji) <sup>3</sup>
Analiza porównawcza różnych działań i mierzenie ich efektywności – szczególnie cenne tam, gdzie środki z jednego źródła zasilają wiele działań o bezpośrednio nieporównywalnych efektach	porównanie „społecznej stopy zwrotu” z nakładów finansowanych z funduszy UE w zależności od celu, na jaki są wydatkowane <sup>4</sup>
Analiza złożonych programów, których liczne działania mogą mieć wzajemny wpływ na skuteczność i skalę efektów	analiza wpływu funduszy strukturalnych na zatrudnienie <sup>5</sup>
Prognoza wpływu przedmiotu ewaluacji na dany organizm, rynek czy zjawisko	analiza wpływu infrastruktury transportowej (obwodnicy metropolii) na gospodarkę <sup>6</sup>
Optymalizacja polityki, programu lub projektu czyli dobór jego parametrów tak, by maksymalizować rezultaty lub minimalizować koszty	optymalizacja poziomu dopłat do cen paliw <sup>7</sup> , optymalizacja poziomu dopłat do opieki nad dziećmi w celu zwiększenia aktywności zawodowej kobiet <sup>8</sup>
Badanie efektu brutto i netto <sup>9</sup>	wartość inwestycji finansowanych bezpośrednio z projektu <i>versus</i> całkowita zmiana wartości inwestycji

Źródło: Opracowanie własne.

<sup>3</sup> Np. w „Badanie wpływu Narodowego Planu Rozwoju 2004–2006 i Narodowych Strategicznych Ram Odniesienia 2007–2013 na wybrane wskaźniki dokumentów strategicznych”, T. Kaczor, R. Socha, IBnGR, 2008, lub „Wpływ realizacji inwestycji finansowanych z funduszy unijnych na kształtowanie się głównych wskaźników dokumentów strategicznych – NPR i NSS oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modelu krajowego i modeli regionalnych Hermin”, WARR, 2008.

<sup>4</sup> A. de la Fuente, *The effect of Structural Fund Spending on the Spanish regions: an assessment of the 1994-99 Objective 1 CSF*, Instituto de Analisis Economico (CSIC), 2002.

<sup>5</sup> Raport końcowy z badania nt.: „Wpływ interwencji z funduszy strukturalnych UE na zatrudnienie”, PAG Uniconsult sp z o.o. /PENTOR Research International, 2007.

<sup>6</sup> Asensio J., Roca O., *Evaluation of transport project infrastructures beyond cost-benefit analysis. An application to Barcelona's 4th ring road*, Universitat Autònoma de Barcelona.

<sup>7</sup> Vedenov D., Wetzstein M., *Toward an Optimal U.S. Ethanol Fuel Subsidy*, University of Georgia 2006.

<sup>8</sup> Domeija D., Klein P., *Should daycare be subsidized?*, 2008.

<sup>9</sup> Rozróżnienie efektu brutto i netto dobrze ilustruje następujący przykład. Prowadząc ewaluację projektu nakierowanego na poprawę szans długotrwale bezrobotnych na znalezienie pracy za pomocą szkoleń, zwykle badany jest efekt brutto – np. jaka część szkolonej grupy podejmuje zatrudnienie w przeciągu pół roku od zakończenia szkoleń. Jakkolwiek daje to dobre pojęcie o tym, ile projekt oferuje uczestnikom, nie pozwala jednak ocenić, jaki jest cały skutek dla społeczeństwa (bo nie uwzględnia np. faktu, że inne osoby nie uzyskały pracy – będąc gorzej wykształcone od uczestników – którą zdobyłyby, gdyby szkoleń nie było) – co ma szczególnie duże znaczenie w przypadku większych programów. Modele w wielu przypadkach pozwalają uwzględnić tego typu zjawiska, szacując skutek całkowity.

Takie podziały, jak przytoczone powyżej, pozornie proste, nie są jednoznaczne i nie można *a priori* określić, jakiego typu model może być skutecznie wykorzystany w danej ewaluacji. W większości sytuacji skutek można osiągnąć, używając kilku, czasem bardzo różnych rodzajów, i nie można zdecydować, że ewaluator powinien wykorzystywać jeden, konkretny typ. Opisane wyżej podziały są bardzo płynne i wiele modeli trudno jest jednoznacznie zakwalifikować do którejkolwiek z grup.

Potencjalny zakres zastosowań modeli ekonomicznych jest bardzo szeroki, ograniczony wyłącznie wyobraźnią ewaluatora. Wszędzie, gdzie przedmiot ewaluacji opisany jest mierzalnymi charakterystykami – co w obecnej chwili jest w zasadzie standardem – otwiera się potencjalne pole wykorzystania jednego z takich narzędzi. Model sprawdzi się we wszystkich trzech podstawowych typach ewaluacji: *ex ante*, bieżącej (*on-going*) i *ex post*. Szczególnie w przypadku tej pierwszej jego użyteczność będzie nieoceniona, w wielu przypadkach może być jedynym źródłem danych liczbowych, gdy zastosowanie innych narzędzi ilościowych jest z definicji wykluczone. Tabela 2 przedstawia najczęstsze zastosowania modeli oraz kształt uzyskiwanych wyników. Z pewnością nie jest to lista zamknięta, wyczerpująca wszystkie możliwe zastosowania, jednak zawiera te najbardziej powszechnie wykorzystywane. Niektóre z przykładów, jakkolwiek nie są w ścisłym słowa znaczeniu ewaluacjami, stanowią dobrą ilustrację zagadnień, jakie mogą być analizowane za pomocą modelu.

### **Analiza przy użyciu modelu to nie tylko cyfry, czyli o znaczeniu założeń i interpretacji**

Mówiąc o stosunkowo niewielkiej arbitralności analiz przy użyciu modeli, wspomniano o roli założeń. Chociaż dziwnie to może zabrzmieć, założenia stojące u podstaw modelu powinny stanowić część jego wyników. Bez ich jasnego i precyzyjnego określenia i opisanie model stanowi zaledwie „czarną skrzynkę”, z której „coś” wypada. Dopiero znajomość założeń umożliwia interpretację uzyskanych danych liczbowych. Najistotniejsze dla wyników (rozumianych zarówno jako uzyskane z modelu dane liczbowe, jak i ich interpretacja) są niewątpliwie założenia dotyczące trzech elementów: ekonomicznej teorii, na której opiera się konstrukcja modelu, mechaniki wpływu przedmiotu ewaluacji oraz danych i parametryzacji modelu<sup>10</sup>.

<sup>10</sup> Przez mechanikę wpływu przedmiotu ewaluacji rozumiemy to, jakie zależności ekonomiczne są wykorzystywane w analizie (przykładowo: wyższa dotacja → większe nakłady inwestycyjne → większa konkurencyjność → wzrost eksportu).

Żaden model nie jest idealnym odwzorowaniem rzeczywistości. Znajomość podstaw teoretycznych pozwala ocenić, jakie procesy doprowadziły do uzyskanego rezultatu oraz jakie zjawiska mogły zostać pominięte. Podobny cel ma sprawdzenie, w jaki sposób ewaluowany proces oddziałuje na badany organizm, na jakie zmienne w modelu wpływa. Konieczna jest jasność, jak przedmiot ewaluacji, zapisany np. słowami projektu, przekłada się na matematyczny język modelu i jakie skutki rodzi akurat taka forma przekładu.

Wreszcie, dla każdego modelu istnieje wiele zestawów danych i parametrów. Od tego, które wybrał ewaluator, zależy będzie skala, a czasami nawet kierunek oddziaływania, dlatego też badanie nie może zostawiać wątpliwości, jakie techniki są używane. Dopiero wiedza o tych trzech grupach założeń wsparta wiedzą o wykorzystanej technice modelowania pozwala dobrze rozumieć wyniki, ocenić ewaluację, porównać ją z innymi lub wykorzystać ponownie w przyszłości<sup>11</sup>.

Założenia stanowią ten element ewaluacji przy zastosowaniu modelu, w którym wyraźnie pojawia się czynnik ekspercki. Drugim takim miejscem jest interpretacja wyników. Z punktu widzenia odbiorcy ewaluacji niebagatelne znaczenie ma wyraźne rozgraniczenie pomiędzy tym, co jest bezpośrednim wynikiem obliczeń przy użyciu modelu, a co stanowi ich rozszerzenie w związku z wiedzą ewaluatora, zewnętrzną względem modelu. Druga z tych części nie ma bowiem prawa korzystać z „autorytetu” modelu oraz stojącej za nim teorii i wymaga oddzielnego uzasadnienia. W przypadku jego braku powinna być interpretowana wyłącznie jako opinia ewaluatora, która, jakkolwiek wartościowa, zawsze będzie skażona relatywnie dużą dozą arbitralności.

Tymczasem w wielu badaniach, nie tylko ewaluacyjnych, pojawiają się w omówieniu wyników modelu kwestie, których model nie porusza i tylko przypadkowo pasują do interpretacji wyników. Na usprawiedliwienie badaczy podkreślimy, że ze względu na złożoność narzędzia, jakim jest model, czasem samemu ewaluatorowi nie jest łatwo ocenić, gdzie się kończy model, a zaczyna znacznie szersze, własne rozumienie świata.

---

<sup>11</sup> Dobrą ilustracją roli, jaką odgrywają założenia teoretyczne i parametryzacja modeli, jest zestawienie analiz wpływu NPR i NSRO na realizację polityk spójności („Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004–2007”, MRR, 2008). Badanie wykonane przy użyciu dwóch narzędzi o odmiennych podstawach teoretycznych (modele HERMIN i MaMoR2), przy takich samych danych wejściowych pokazuje znaczące różnice w skali i rozkładzie czasowym skutków. Dopiero znajomość podstaw obu modeli pozwala zrozumieć, dlaczego uzyskane wyniki różnią się w takim stopniu i co implikują te różnice.

## Ograniczenia zastosowań modeli

Modele ekonomiczne są bardzo elastycznymi narzędziami nadającymi się do analizy różnorodnych problemów ewaluacyjnych. Jednocześnie, każdy model jest instrumentem mniej lub bardziej dedykowanym zagadnieniu, jakiemu go poświęcono. Wybór założeń, mechaniki modelu czy nawet danych jest silnie uzależniony od bardzo specyficznego charakteru problemu ewaluacyjnego i zastosowany do pozornie zbliżonego zagadnienia może dać błędne wyniki. W szczególności trudne jest uzyskanie przy użyciu jednego narzędzia wyników o silnie zróżnicowanym stopniu detaliczności. Model całej gospodarki narodowej, zaprojektowany do oceny wpływu polityki na podstawowe agregaty makroekonomiczne będzie miał problemy z jednoczesnym wyznaczeniem jej oddziaływania, np. na bardzo wąską grupę społeczną. Ewaluatorowi postawionemu przed tego typu problemem jako rozwiązanie proponuje się więc stworzenie dwóch modeli: ogólnego oraz cząstkowego, który w jakimś stopniu bazowałby na wynikach tego pierwszego, oferując nieosiągalny dlań wgląd w detale. Jednak takie rozwiązanie zawsze powoduje dodatkowe koszty.

Podobnym ograniczeniem z tej grupy jest kwestia skali ilościowej analizowanych procesów. Analiza zjawisk zachodzących masowo oparta jest na innych zasadach niż badanie pojedynczych przypadków. Czasem rodzi się pokusa by użyć narzędzia zaprojektowanego dla pierwszej grupy dla przeanalizowania przykładowych przypadków, np. wybranych jednostek z grupy podlegającej badaniu. To, co pozornie łatwe i tanie („przecież jest już model, to wystarczy przeliczyć dla tych kilku gmin”), często kryje w sobie błąd metodologiczny – model jest zaprojektowany z myślą o wykorzystaniu np. statystycznych charakterystyk populacji, przez co działa dla grupy podmiotów, ale nie dla każdego z nich osobno.

Kolejnym, istotnym w naszej ocenie, ograniczeniem zastosowań modelowania jest kwestia ich wymagań dotyczących danych. Zaczniemy od stwierdzenia, że model jest narzędziem ilościowym, co między innymi oznacza, że dane wejściowe powinny mieć charakter liczbowy. W konsekwencji także ewaluowany program czy projekt musi przedstawiać działania, których skutki mają być badane w formie ilościowej, najczęściej dotyczyć to będzie nakładów. Na twórcy podmiotu ewaluacji spoczywa ciężar nadania działaniom takiego właśnie kształtu, w innym przypadku ewaluacja będzie bezwartościowa.

Chociaż modele bardzo różnią się między sobą wymaganiami dotyczącymi danych, to jednak dla poprawnego działania zawsze potrzebują ich systematyczności i kompletności. W przypadku większości z nich niezbędne dane pochodzić będą z więcej niż jednego źródła, a nawet te pochodzące z jednego różnią się nieraz terminem publikacji często o kilka kwartałów. Tymczasem wiele typów modeli jest w stanie działać wyłącznie, gdy dostęp-

ne są wszystkie dane, np. z danego okresu lub regionu. Oznacza to, że brak nawet jednej danej wyklucza użycie pozostałych, choć oczywiście istnieją techniki pozwalające na uzupełnienie pewnych mniej znaczących braków. Dlatego zdarza się że w ewaluacji zostanie użyty zestaw danych pozornie bardzo odległych w czasie<sup>12</sup>. Fakt ten nie odbiera jednak aktualności wynikom ewaluacji, gdyż zmiany w gospodarce zachodzą na tyle wolno, że zwykle można zaakceptować liczne opóźnienia bez znaczącej straty dla jakości obliczeń.

Model zawsze jest uproszczonym odwzorowaniem rzeczywistości. Wykorzystane przy jego konstrukcji techniki pozwalają dobrze ją opisywać mimo pominięcia pewnych zależności. Umożliwiają także analizę sytuacji hipotetycznych w niewielkim stopniu od niej odbiegających. To rodzi kolejne ograniczenie. Modelowanie sytuacji drastycznie różniących się charakterem od historycznych doświadczeń – na których w mniejszym lub większym stopniu oparty jest każdy model – rodzi ryzyko uzyskania przypadkowych wyników. Dotyczy to również finansowej skali przedsięwzięć. O ile można przyjąć, że zwiększenie wydatków na jakiś cel o jeden procent da zbliżony skutek jak każdy procent z dotychczasowych, to przyjęcie podobnej logiki przy zwiększeniu trzykrotnym zwykle będzie już błędem.

Kluczowym elementem większości ewaluacji jest porównanie dwóch stanów. Jeden z nich, często nazywany referencyjnym lub bazowym, obrazuje rzeczywistość bez przedmiotu ewaluacji, podczas gdy drugi odpowiada sytuacji, w której zrealizowano ewaluowany program lub działanie. To rozróżnienie, nie zawsze wyraźnie zasygnalizowane, w przypadku badań z użyciem modeli będzie na ogół sercem analiz, których wiele sprowadza się po prostu do porównania tych stanów. Ten stan referencyjny, który został przyjęty do badania, ma znaczący wpływ na ostateczny wynik ewaluacji<sup>13</sup>. O ile w przypadku badania *ex-post* stan jest zwykle definiowany przez dostępne dane (co oznacza, że jest stanem obiektywnym), o tyle w pozostałych dwóch typach ewaluacji jest on na ogół ustalany przez ewaluatora. Z tej przyczyny nieodłącznym elementem takiego badania i jego wyników powinno być wyraźnie przedstawienie stanu referencyjnego z podkreśleniem tych jego elementów, które mogą mieć istotne znaczenie dla uzyskanych rezultatów. Dodajmy jeszcze, że porównanie dwóch ewaluacji tego samego przedmiotu powinno dotyczyć nie tylko ich bezpośrednich wyników, ale także możliwego wpływu stanu referencyjnego.

<sup>12</sup> Dla przykładu: najbardziej aktualne oficjalnie opublikowane przez GUS tabele *input-output*, kluczowe dla niektórych typów modeli, pochodzą z roku 2000, a więc mają 8 lat.

<sup>13</sup> Przykładowo, dodatkowy efekt środków przeznaczonych na wspieranie innowacyjnych inwestycji przedsiębiorstw w sytuacji, gdy ich kondycja finansowa jest zła, może być bardzo duży. Jednak w sytuacji, gdy firmy dysponują nadmiarem zgromadzonej gotówki, lecz nie inwestują (co miało np. miejsce przed rozpoczęciem aktualnego boomu inwestycyjnego), efekt tych samych środków może być niewielki.

Wreszcie, tylko w niewielu badaniach ewaluacyjnych model będzie narzędziem wystarczającym do pełnej oceny przedmiotu ewaluacji. Analiza modelowa potrafi zwykle określić skalę oddziaływania projektu, jednak nie mówi nic o tym, czy wpływ w ogóle występuje<sup>14</sup>. Używając jednego z wcześniejszych przykładów, model może wyliczyć skutki nakładów na szkolenia dla lokalnego rynku pracy, natomiast nie jest w stanie stwierdzić, czy szkolenia trafią do właściwych osób. Ilościowa analiza przy użyciu modelu jest zawsze uzupełnieniem analizy jakościowej, bowiem to ta druga właśnie pozwala powiedzieć, jaki rodzaj oddziaływania uwzględnić w modelu. Oczywiście nie zawsze taka analiza musi być częścią ewaluacji, może bazować na wcześniejszych doświadczeniach i badaniach, jednak zawsze powinna istnieć w świadomości ewaluatora. Odbiorca ewaluacji musi mieć pewność, że nie modelujemy nieistniejących zjawisk.

Spróbujmy stworzyć małe podsumowanie tego, co powinna zawierać ewaluacja przy użyciu modelu, tak by jej odbiorca mógł mieć pewność co do jakości uzyskanych rezultatów. Poniższa lista (tabela 3) z pewnością nie wyczerpuje zagadnienia, ale może ułatwić przygotowanie i odbiór takich badań.

Tabela 3. Wybiórczy niezbędnik ewaluacji przy użyciu modelu

Co powinna zawierać ewaluacja?	Czemu to jest ważne?
Opis teorii i założeń stojących u podstaw badania...	... aby być pewnym, że uwzględniono wszystkie ważne czynniki wpływu przedmiotu ewaluacji na otoczenie
Charakterystyka modelu...	... aby rozumieć jego ograniczenia
Opis danych...	... bo bez tego trudno interpretować wyniki
Na czym polegał eksperyment, opis stanu referencyjnego...	... aby wiedzieć, czy i jak wpłynęły na wyniki
Interpretacja odróżniająca produkty modelu od innych elementów...	... aby ocenić obiektywność wyników

Źródło: Opracowanie własne.

Podsumowując, modele ekonomiczne stanowią mogą wartościowe narzędzie ewaluacyjne, weryfikujące działanie lub dostarczające informacji nieosiągalnych przy użyciu innych technik. Jednak, by ich rezultaty miały wartość, badania ewaluacyjne nie mogą ograniczać się do cyfr, lecz muszą zawierać szereg informacji pozwalających je zrozumieć i zinterpretować, także niewy-

<sup>14</sup> Pomijamy tu niektóre specyficzne modele służące do analiz *ex-post* mających na celu właśnie stwierdzenie wystąpienia rzeczywistego oddziaływania.

specjalizowanemu odbiorcy. Stosunkowo niewielka świadomość ich możliwości i potencjalnych zastosowań wydaje się być przeszkodą w ich szerszym zastosowaniu.

---

**Tomasz Kaczor** – doktor nauk ekonomicznych, absolwent Wydziału Nauk Ekonomicznych UW, specjalizuje się w wykorzystaniu metod ilościowych w badaniu gospodarki. W latach 1997–2001 pracownik naukowy Instytutu Badań nad Gospodarką Rynkową, od 2001 roku do chwili obecnej współpracownik. Przez sześć lat pracował jako wicedyrektor departamentu finansów w centrali ZUS i był odpowiedzialny m.in. za budżet FUS i FRD oraz prowadzenie analiz makroekonomicznych i finansowych. Obecnie pełni funkcję Głównego Ekonomisty Banku Gospodarstwa Krajowego. Prowadzi prace badawcze i analizy dotyczące makroekonomii, finansów publicznych i zagadnień europejskich. Ma wieloletnie doświadczenie w tworzeniu modeli ekonomicznych i ich wykorzystaniu do prowadzenia analiz ekonomicznych i sporządzania prognoz. Autor i współautor szeregu publikacji i ewaluacji, w szczególności licznych prac dotyczących badania wpływu funduszy strukturalnych na różne aspekty gospodarki. Twórca modelu MaMoR2 wykorzystywanego do analiz makroekonomicznego wpływu polityki spójności oraz szeregu innych modeli.

## Bibliografia

- Asensio J., Roca O., *Evaluation of transport project infrastructures beyond cost-benefit analysis. An application to Barcelona's 4th ring road*, Universitat Autònoma de Barcelona.
- „Badanie wpływu Narodowego Planu Rozwoju 2004–2006 i Narodowych Strategicznych Ram Odniesienia 2007–2013 na wybrane wskaźniki dokumentów strategicznych”, T. Kaczor, R. Socha, IBnGR, 2008.
- Domeija D., Klein P., *Should daycare be subsidized?*, 2008.
- de la Fuente A., *The effect of Structural Fund Spending on the Spanish regions: an assessment of the 1994-99 Objective 1 CSF*, Instituto de Analisis Economico (CSIC), 2002.
- „Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004–2007”, MRR, 2008.
- Raport końcowy z badania nt.: „Wpływ interwencji z funduszy strukturalnych UE na zatrudnienie”, PAG Uniconsult sp z o.o. /PENTOR Research International, 2007.
- Vedenov D., Wetzstein M., *Toward an Optimal U.S. Ethanol Fuel Subsidy*, University of Georgia 2006.
- „Wpływ realizacji inwestycji finansowanych z funduszy unijnych na kształtowanie się głównych wskaźników dokumentów strategicznych – NPR i NSS oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modelu krajowego i modeli regionalnych Hermin”, WARR, 2008.
- <http://www.fundusze-strukturalne.gov.pl/EWALUACJA+funduszy+w+Polsce/Baza+badan+ewaluacyjnych/>

Stanisław Bienias

## Wykorzystanie wyników modelowania ekonomicznego w procesie ewaluacji polityki spójności

Czego spodziewa się odbiorca wyników ewaluacji, sięgając po raport końcowy z badania? Odbiorca chce zobaczyć konkrety, a coś jest bardziej konkretnego niż liczby pokazujące efekt badanej interwencji? Jeżeli liczby te dotyczą jeszcze oddziaływań, a nie wyłącznie informacji odnoszących się do poziomu produktów i rezultatów (w końcu tych dostarcza nam monitoring<sup>1</sup>), to większość z nas jest prawie w pełni ukontentowana. Prawie, ponieważ osoby zajmujące się dłużej badaniami chciałyby zobaczyć liczby dotyczące rzeczywistych wniosków (tzw. efektów netto), a nie tylko wyników brutto interwencji będącej przedmiotem ewaluacji. Co więcej, czasami chcemy oszacować skalę tych efektów (patrz ewaluacja *ex ante*), zanim jeszcze nasz program/projekt zostanie uruchomiony.

Jak widać, usatysfakcjonowanie odbiorcy ewaluacji nie jest łatwe – istnieją jednak narzędzia, techniki badawcze, dzięki którym możemy osiągnąć pożądany efekt. Narzędziem takim są właśnie modele ekonomiczne, które znakomicie się do tego celu nadają. Dzięki modelowaniu ekonomicznemu uzyskujemy dokładną kwantyfikację efektów naszej interwencji, a nie wyłącznie ogólne stwierdzenie (jakże często występujące w raportach ewaluacyjnych), że „wydaje się, iż projekt/program wywołał znaczące pozytywne efekty zarówno w sferze społecznej, jak i gospodarczej, lecz z racji złożoności interwencji skala tych oddziaływań jest trudna do oszacowania”. Stosując modele, jesteśmy w stanie pokazać efekty netto, a nie, jak w większości tradycyjnych metod badawczych, efekty brutto. Możemy szacować

---

<sup>1</sup> W Polsce w okresie programowania 2004–2006, przy bardzo dobrze działającym monitoringu postępu finansowego (premier 20 dni po zakończeniu każdego miesiąca dostaje na biurko informację – kto, w ramach jakiego działania, nie wydatkuje pieniędzy), istnieje bardzo ograniczona informacja na temat postępu rzeczowego. Wskaźników jest zbyt dużo, są one niespójne i w większości przypadków nie pozwalają na agregację informacji na wyższe poziomy wdrażania. Problemy dodatkowo potęguje niedziałający system informatyczny SIMIK 2004–2006.



efekty naszej interwencji, które wystąpią w przyszłości, a nawet określić przewidywaną trwałość tych efektów – już po zakończeniu programu/projektu. Jesteśmy w stanie uzyskać informacje dotyczące wpływu interwencji na zmienne makroekonomiczne, takie jak PKB, stopa zatrudnienia czy wydajność pracy – przy odpowiednio skonstruowanym narzędziu ograniczenia w tym zakresie są bardzo niewielkie. Ponadto modele pozwalają nam pokazać efekty, jakie nasza interwencja wywołuje w całej gospodarce (krajowej/regionalnej/lokalnej) i odnieść je do całości zachodzących przemian – widzimy efekty interwencji w dynamicznym kontekście. Modele mają jeszcze jedną niewątpliwą zaletę – mimo dużego skomplikowania narzędzia i jednocześnie znacznych nakładów pracy związanych z jego opracowaniem, a następnie utrzymaniem – jest ono relatywnie tanie<sup>2</sup>.

Czy oznacza to, że wszystkie badania ewaluacyjne powinniśmy oprzeć wyłącznie na modelowaniu ekonomicznym? Oczywiście, nie. Jak każde narzędzie, tak i modele ekonomiczne, mają pewne ograniczenia. Nie będę tutaj poruszał zagadnień dotyczących konstrukcji narzędzia, lecz wyłącznie te związane z wykorzystaniem samych wyników modelowania.

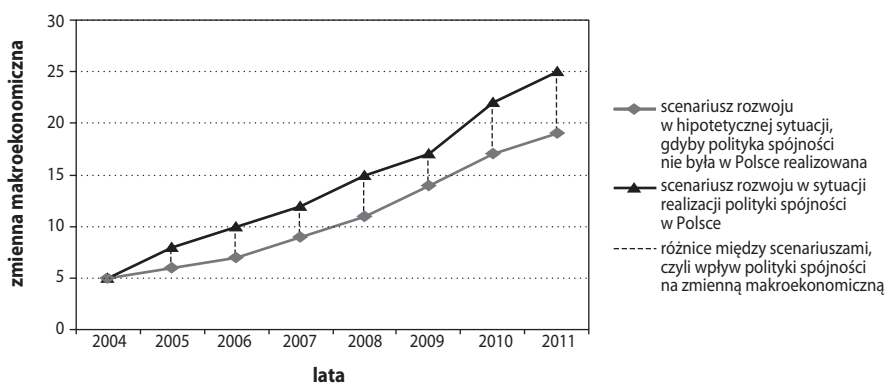
Z uwagi na specyficzną terminologię, i konieczność posiadania przez odbiorcę chociaż minimalnego zasobu wiedzy z dziedziny makroekonomii, pierwszą trudność napotykamy już w momencie prezentacji wyników modelowania. Dla zdecydowanej większości osób badania różnice pomiędzy „punktem procentowym” a „procentem” są niewielkie. Jeszcze trudniejsze jest rozróżnienie np. między „poziomem Produktu Krajowego Brutto” a „dynamiką Produktu Krajowego Brutto”. Bardzo często (a celują w tym zwłaszcza media) terminy te w odniesieniu do konkretnej liczby są stosowane zamiennie! Możemy więc sobie wyobrazić, jak duże błędy w ten sposób powstają! Dodatkową trudność stanowi fakt, że w przypadku badań ewaluacyjnych na wykresach, w tabelach prezentowany jest wyłącznie wpływ interwencji na daną zmienną makroekonomiczną (czyli różnica między dwoma możliwymi scenariuszami rozwoju – z uwzględnieniem realizacji interwencji oraz w hipotetycznej sytuacji – gdyby dana interwencja nie była realizowana). Prowadzi to w wielu przypadkach do dodatkowego zwiększenia błędu w zrozumieniu wyników. Często dane prezentujące efekt skumulowany badanej interwencji (łączy efekt narastający w ciągu kilku kolejnych lat) są do siebie dodawane, co daje wielokrotne przeszacowanie wyniku. Dalej przedstawiony został sposób, w jaki należy rozumieć prezentowane wyniki oddziaływania makroekonomicznego polityki spójności.

---

<sup>2</sup> Realizacja badania mającego na celu ocenę efektu netto za pomocą ankiet na dużych grupach kontrolnych będzie w większości przypadków droższa niż zastosowanie modelu ekonomicznego.

## Jak należy interpretować wyniki modelowania ekonomicznego?

Na wykresie 1 przedstawione zostały dwa scenariusze rozwoju. Scenariusz rozwoju uwzględniający realizację polityki spójności w Polsce – linia szara oraz hipotetyczny scenariusz rozwoju w sytuacji, gdyby Polska nie była beneficjentem tej polityki. Linie przerywane (różnica między dwoma scenariuszami) obrazują wpływ realizacji polityki spójności na daną zmienną. Jest to efekt netto (mamy bowiem do czynienia z odniesieniem do scenariusza rozwoju w sytuacji kontrfaktycznej). W celu większej przejrzystości prezentacji wyników na ogół na wykresach pokazuje się wyłącznie jedną linię przedstawiającą wpływ interwencji na określoną zmienną ekonomiczną. Spójrzmy na konkretny przykład prezentacji efektów polityki spójności.



Wykres 1. Wpływ polityki spójności na zmienną makroekonomiczną

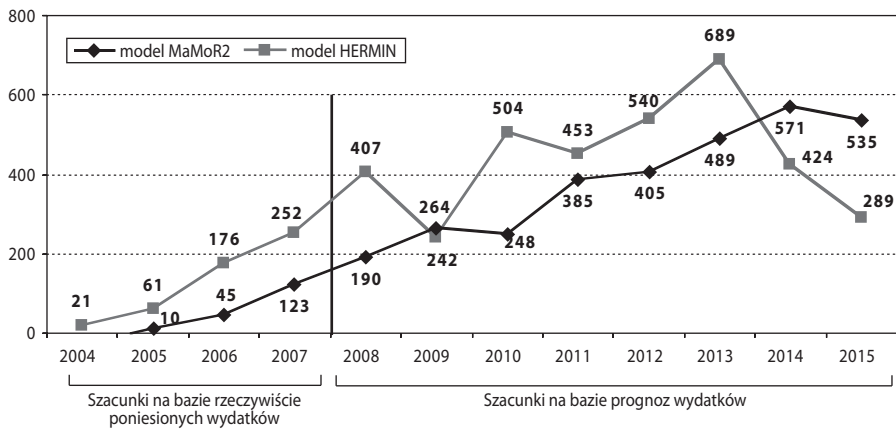
Źródło: Opracowanie na podstawie materiałów roboczych Krajowej Jednostki Oceny, MRR.

Na wykresie 2 zaprezentowany został wpływ polityki spójności na liczbę pracujących w wieku od 15 do 64 lat, oszacowany za pomocą dwóch modeli makroekonomicznych (model HERMIN – linia szara i model MaMoR2 – linia czarna). Według szacunków przeprowadzonych za pomocą modelu HERMIN w 2007 r. ilość osób pracujących<sup>3</sup> była wyższa o 252 tys. osób niż w przypadku hipotetycznej sytuacji, gdyby Polska nie była beneficjentem polityki spójności i nie korzystała ze wsparcia finansowego z funduszy unij-

<sup>3</sup> Liczonej na podstawie wyników Badań Aktywności Ekonomicznej Ludności (BAEL). Według metodologii BAEL do pracujących zalicza się osoby, które w badanym tygodniu:

- wykonywały, przez co najmniej 1 godzinę, pracę przynoszącą zarobek lub dochód albo pomagały (bez wynagrodzenia) w prowadzeniu rodzinnego gospodarstwa w rolnictwie lub rodzinnej działalności gospodarczej poza rolnictwem,
- nie wykonywały pracy (np. z powodu choroby, urlopu, przerwy w działalności zakładu, trudnych warunków atmosferycznych), ale formalnie były zatrudnione (wartości średnioroczne).

nych (zwróćmy uwagę, że odległość między linią szarą a osią poziomą na wykresie 2 odpowiada długości linii przerywanej – obrazującej wpływ – na wykresie 1). Należy mieć na uwadze, że 252 tys. osób w 2007 r. pokazuje efekt skumulowany. Błędna byłaby interpretacja, że niniejsze miejsca pracy powstały tylko w roku 2007. Znaczna część osób – z prezentowanych 252 tys. – znalazła bowiem zatrudnienie (w wyniku realizacji polityki spójności) już w latach poprzednich<sup>4</sup>! Idąc dalej tym tropem, zauważymy, że między rokiem 2008 a 2009 efekt realizacji polityki spójności będzie miał negatywny wpływ na rynek pracy<sup>5</sup> – chociaż efekt skumulowany pozostanie nadal pozytywny – 242 tys. osób. Analogicznie powinniśmy interpretować szacunki opracowane przy użyciu drugiego modelu – MaMoR2 (linia czarna na wykresie).



Wykres 2. Skumulowany wpływ polityki spójności na liczbę pracujących w wieku od 15 do 64 lat (w tys. osób)

Źródło: *Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004–2007*, Departament Koordynacji Polityki Strukturalnej, Ministerstwo Rozwoju Regionalnego, lipiec 2008 r.

Musimy mieć na uwadze, że każda instytucja, każdy ekspert w nieco inny sposób prezentuje otrzymane wyniki modelowania ekonomicznego, co dodatkowo utrudnia ich bezproblemowe zrozumienie. Sugeruje się raczej ograniczanie ilości informacji, dotyczących różnych zmiennych makroekonomicz-

<sup>4</sup> Jeszcze większym błędem byłoby twierdzenie, że w latach 2004–2007 powstało 510 tys. miejsc pracy (21 + 61 + 176 + 252). Takie błędne informacje były rozpowszechniane w uznanych gazetach o ogólnopolskim zasięgu.

<sup>5</sup> Spadek z 407 do 242 tys. osób. Główną przyczyną takiego stanu będzie zakładane, prognozowane zmniejszenie transferów do gospodarki, związane z zakończeniem realizacji perspektywy 2004–2006, a jeszcze niepełnym „rozpędzeniem” wydatkowania środków z perspektywy 2007–2013.

nych prezentowanych w raporcie z ewaluacji, na rzecz poświęcenia większej ilości miejsca szczegółowemu opisaniu i zinterpretowaniu kilku kluczowych wartości, najważniejszych z punktu widzenia realizacji badania. Niezwykle istotne jest ponadto, by sam raport pisany był językiem niefachowym, zrozumiałym dla szerszego grona odbiorców, a nie tylko dla wąskiej grupy specjalistów. Kwestie te są szczególnie istotne, gdy możemy się spodziewać, że wyniki naszego raportu będą interesujące również dla opinii publicznej i mediów. Dzięki temu uniknąć możemy wielu nieporozumień i pisaniasprostowań do artykułów prasowych<sup>6</sup>.

Zdefiniowaliśmy więc pierwszy problem związany z wykorzystaniem wyników modelowania – są one stosunkowo mało zrozumiałe dla szerszego grona odbiorców. Kolejny problem pojawia się w momencie wyjaśnienia, dlaczego otrzymaliśmy taki, a nie inny wynik. Dla większości osób (praktycznie wszystkich niezajmujących się bezpośrednio modelowaniem) sam model stanowi pewnego rodzaju „pudełko-zagadkę”. Po wrzuceniu do modelu kilku wartości – w przypadku polityki spójności są to płatności rzeczywiste (na bazie danych historycznych) lub/i planowane (prognozy płatności w kolejnych latach) na rzecz beneficjentów pomocy w rozbiciu na kategorie wydatków<sup>7</sup> – otrzymujemy wyniki, które nas interesują – wpływ na PKB, liczbę osób pracujących. Mechanizm, jaki kryje się za tym przeistoczeniem, dla większości odbiorców ewaluacji na zawsze pozostanie tajemnicą. Problemy pojawiają się najczęściej w momencie, gdy efekty są inne, niż się spodziewaliśmy – np. mamy do czynienia z nagłym zwrotem trendu oddziaływania, dużą fluktuacją prognoz lub innymi nieoczekiwanymi wynikami.

## Kontrowersyjne wyniki modelowania makroekonomicznego

Ciekawym przykładem może być tutaj analiza opracowana przez Urząd Marszałkowski Województwa Dolnośląskiego<sup>8</sup> na potrzeby optymalizacji alokacji dla programu regionalnego. W ramach analizy opracowano kilka możliwych scenariuszy alokacji środków (różny udział środków służących interwencjom, które można określić jako sprzyjające osiągnięciu celów Strategii Lizbońskiej) i ich spodziewanego wpływu na rozwój gospodarki regionalnej.

---

<sup>6</sup> Przy prezentacji wyników na konferencjach prasowych należy ponadto pamiętać o przygotowaniu notatki dla dziennikarzy, która powinna być napisana łatwym i zrozumiałym językiem, oraz jej umieszczeniu na odpowiedniej stronie internetowej.

<sup>7</sup> Modele wykorzystywane w Polsce uwzględniają trzy główne kategorie interwencji – czyli kierunek wykorzystania środków: inwestycje w sektor produkcyjny, inwestycje w infrastrukturę oraz wydatki na rozwój zasobów ludzkich.

<sup>8</sup> Bradley J., Zaleski J., Tomaszewski P., Zembaty M., *Ocena efektu makroekonomicznego Regionalnego Programu Operacyjnego na lata 2007–2013 na gospodarkę Województwa Dolnośląskiego za pomocą modelu HERMIN*, Wrocław 2006.

Okazało się – ku zaskoczeniu wszystkich – że scenariusz alokacji mniej „prolizboński” zapewniał lepsze efekty rozwojowe! Czy to oznacza, że nie powinniśmy realizować Strategii Lizbońskiej z wykorzystaniem środków w ramach polityki spójności? Oczywiście, nie!

Pamiętajmy, że model jest tylko instrumentem, a za otrzymanymi wynikami stoją pewne założenia i ograniczenia dla samego narzędzia. Po pierwsze, różnice między poszczególnymi scenariuszami były relatywnie niewielkie i mieściły się w granicy błędu modelowania. Po drugie, pamiętać należy, że model widział różne scenariusze wydatków w znacznie bardziej uproszczonej formie, niż my je rozumiemy, posługując się pojęciem „wydatki prolizbońskie”. Dane dostarczane do modelu pogrupowane były w ramach trzech kategorii interwencji (inwestycje w sektor produkcyjny, inwestycje w infrastrukturę oraz wydatki na rozwój zasobów ludzkich). Dla przykładu w kategorii infrastruktura znajdowały się zarówno wydatki „prolizbońskie”, na przykład budowa autostrad, jak i wydatki, które ciężko podejrzewać o generowanie wysokich impulsów rozwojowych – np. budowa basenów. Dla modelu ten sam efekt na wzrost PKB generuje jedno euro przeznaczone na budowę autostrady, jak i wydane na budowę infrastruktury społecznej. W tym przypadku instytucja, która realizowała niniejszą analizę, nie pozostawiła wyników bez wyjaśnienia i jasno wskazała, że uzyskane szacunki nie pozwalają na wysnuć wiążących wniosków w zakresie efektywności środków „prolizbońskich”. Gdyby niniejsze wyniki pozostawione zostały bez odpowiedniego komentarza ze strony twórców modelu, mogłyby służyć niewłaściwym decyzjom podejmowanym na szczeblu politycznym.

Przyczyny otrzymania nietypowych/niespodziewanych wyników mogą być bardzo różne – poczynając od założeń przyjętych w modelu, po teorie ekonomiczne leżące u podstaw modelowania<sup>9</sup>. Jeżeli interpretację kontrowersyjnych wyników pozostawimy wyłącznie odbiorcom ewaluacji, możemy spodziewać się sformułowania wielu nieprawidłowych lub niezasadnych wytłumaczeń. Najbardziej właściwymi osobami, które każdorazowo powinny przedstawiać stosowną interpretację, są twórcy samego modelu, ponieważ to oni znają model najlepiej i wiedzą, jakie mechanizmy przyczyniły się do osiągnięcia takich, a nie innych wyników. Zlecający ewaluację powinni natomiast dopilnować, by odpowiednie wyjaśnienia znalazły się w raporcie z badania<sup>10</sup>.

Ponieważ, jak zostało to powiedziane wcześniej, dla większości osób model jest swoistym „pudełkiem zagadką”, bardzo duże znaczenie przy realizacji

<sup>9</sup> W tym miejscu spójrzmy, jak różnie zachowują się modele np. przy nagłym zaprzestaniu finansowania, bądź znacznym ograniczeniu jego skali. Jedne modele prognozują w takiej sytuacji bardzo gwałtowne – negatywne „tąpnięcie” w sferze realnej gospodarki (HERMIN), podczas gdy inne reagują stosunkowo spokojnie (MaMoR2).

<sup>10</sup> Praktyka wskazuje, że czasami sami twórcy nie są w stanie przedstawić jasnej i klarownej interpretacji „wątpliwych” wyników i sami nie wiedzą, dlaczego model pokazuje takie, a nie inne prognozy.

badania z wykorzystaniem modelowania ma autorytet twórców samego narzędzia. Jest to istotne, gdyż większość modeli stanowi własność konkretnych instytucji i nie ma pełnej dostępnej dokumentacji prezentującej wszystkie elementy modelu<sup>11</sup>. W większości przypadków publikowane są jedynie ogólne opisy modelu oraz pakiet założeń przyjętych przez autorów. Wiarygodność wyników jest zatem niejako „firmowana” autorytetem i doświadczeniem twórców narzędzia, a wiele informacji przyjmuje się, bazując wyłącznie na opinii i informacji uzyskanych bezpośrednio od osób realizujących badanie. Istotne jest więc, by mieć pewność, że za wynikami prac stoi wiedza i doświadczenie twórców modelu, gdyż między innymi od tego zależeć będzie wiarygodność wyników samego modelowania.

Kolejnym problemem, który pojawia się przy wykorzystaniu wyników modelowania ekonomicznego, jest ograniczona możliwość porównania danych. Trudność polega na tym, że wyniki prognoz przeprowadzone przy wykorzystaniu poszczególnych modeli znacznie się różnią. Różnice te dotyczą zarówno wielkości oddziaływania, momentu, w którym to oddziaływanie jest największe, jak i reakcji gospodarki na ograniczenie skali dopływu środków europejskich itp.

## Różnice w wynikach prognoz

Przykładem mogą tu być wyniki analiz przeprowadzonych na zlecenie Ministerstwa Rozwoju Regionalnego przy zastosowaniu dwóch modeli szacujących wpływ na poziomie regionalnym<sup>12</sup>. Oceniając całościowy wpływ polityki spójności, przy zastosowaniu modelu MaMoR2, otrzymujemy wyniki prognozowania, które wskazują, że polityka ta będzie przyczyniać się do przyspieszenia konwergencji międzyregionalnej w Polsce. Wszystkie najbied-

---

<sup>11</sup> Wyjątek stanowi tutaj dokumentacja dotycząca polskiej implementacji modelu HERMIN. W związku z finansowaniem większości prac ze środków publicznych, zarówno Ministerstwo Rozwoju Regionalnego, jak i Wrocławska Agencja Rozwoju Regionalnego (współtwórca polskiej wersji modelu HERMIN) prowadzi bardzo otwartą politykę informacyjną i upublicznia wszystkie informacje dotyczące konstrukcji modelu. Działanie takie, niewątpliwie właściwe z uwagi na finansowanie przedsięwzięcia ze środków publicznych, jednocześnie narażają sam model na dość intensywną krytykę środowisk naukowo-ekspertkich. Ograniczona ilość dostępnych informacji nie pozwala bowiem na tak pogłębioną krytykę innych modeli.

<sup>12</sup> Kaczor T., Socha R., „Badanie wpływu Narodowego Planu Rozwoju 2004–2006 i Narodowych Strategicznych Ram Odniesienia 2007–2013 na wybrane wskaźniki dokumentów strategicznych”, Instytut Badań nad Gospodarką Rynkową, Prevision, kwiecień 2008 r.

„Wpływ realizacji inwestycji finansowanych z funduszy unijnych na kształtowanie się głównych wskaźników dokumentów strategicznych – NPR i NSS oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modelu krajowego i modeli regionalnych Hermin”, Zespół Wrocławskiej Agencji Rozwoju Regionalnego pod kierownictwem prof. J. Zaleskiego we współpracy z dr J. Bradleyem, maj 2008 r.

niejsze województwa rozwijać się będą zdecydowanie szybciej niż średnia dla kraju. Natomiast wyniki prognozowania, przy zastosowaniu modelu HERMIN, dają wręcz odwrotne rezultaty – realizacja polityki spójności przyczyni się do pogłębienia różnicowań międzyregionalnych w naszym kraju (cztery z pięciu najbiedniejszych województw będą się rozwijały wolniej niż średnia dla kraju).

Wpływ funduszy unijnych na poziom PKB na mieszkańca w relacji do średniej dla kraju dla 5 najbiedniejszych i 5 najbogatszych polskich województw przedstawia tabela 1.

Tabela 1

Województwo	Wpływ funduszy unijnych na poziom PKB na mieszkańca w relacji do średniej dla kraju dla 5 najbiedniejszych i 5 najbogatszych polskich województw		
		MaMoR2	HERMIN
		2015	
Średnia zmiana dla najbiedniejszych województw:		+	-
lubelskie	Najbiedniejsze województwa	+1,8	-2,57
podkarpackie		+2,3	+0,35
podlaskie		+2,8	-2,59
świętokrzyskie		+2,4	-2,7
warmińsko-mazurskie		+3,7	-0,13
średnia zmiana dla najbiedniejszych województw			-
śląskie	Najbogatsze województwa	+1,4	+4,72
mazowieckie		-6,7	-3,73
pomorskie		+1,7	+2,94
śląskie		-0,1	+2,46
wielkopolskie		-1,2	+2,99
Średnia zmiana dla najbogatszych województw		-	+

Źródło: Opracowanie własne na podstawie *Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004–2007*, Departament Koordynacji Polityki Strukturalnej, Ministerstwo Rozwoju Regionalnego, lipiec 2008 r.

Różnice w wynikach modelowania przy zastosowaniu dwóch różnych narzędzi, chociaż zastanawiające, z naukowego punktu widzenia wydają się zrozumiałe. Modele przygotowują dwa różne zespoły badawcze, u podstaw każdego z narzędzi stoją różniące się teorie ekonomiczne, a do przeprowadzenia badań przyjmuje się nieco inne założenia. Wykorzystując modele ekonomiczne, należy zawsze mieć na uwadze, że są one zaawansowanym i skomplikowanym, ale jednak tylko narzędziem, a otrzymane wyniki nie zawsze muszą odpowiadać rzeczywistości. Podobnie jak ma to miejsce w przypadku realizacji ankiet – na identycznie zadane pytanie możemy uzyskać różne odpowiedzi, zależnie od tego, jakim typem ankiety się posłużymy. Z podobną sytuacją (nieco bardziej złożoną) mamy do czynienia, stosując modele ekonomiczne. Specyfika modelu i przyjęte założenia bardzo silnie wpływają na otrzymywane wyniki. Należy zatem dokładnie zapoznać się z założeniami procesu badawczego, a dla wzmocnienia wiarygodności otrzymanych wyników warto rozważyć wykorzystanie więcej niż jednego modelu ekonomicznego (triangulacja). Daje to możliwość oparcia procesu decyzyjnego o lepiej zweryfikowane informacje.

W konsternację mogą natomiast wprowadzić bardzo duże różnice wyników przygotowane za pomocą tych samych modeli tylko w różnym czasie. W niektórych przypadkach wyniki modelowania różnią się nawet o 50% między badaniami przeprowadzanymi rok po roku.

W takich sytuacjach konieczne jest, by twórcy modelu przedstawiali stosowne wyjaśnienia, które tłumaczą takie, a nie inne wyniki prognozowania. Najczęstsza przyczyna tego, że prognozy różnią się w czasie, wynika z rozszerzenia szeregów czasowych o kolejne okresy (uaktualnienie modelu). Należy również pamiętać, że „modelarze” nieustannie się doskonalą, porównują otrzymane wyniki i w rezultacie wprowadzają pewne modyfikacje do swoich narzędzi, co następnie skutkuje odmiennymi wynikami<sup>13</sup>.

## Podsumowanie

Polska na tle innych państw członkowskich UE jest krajem, który stosunkowo dużą wagę przywiązuje do znaczenia badań oddziaływania makroekonomicznego w realizacji polityki spójności z wykorzystaniem modeli ekonomicznych. Modele są w naszym kraju wykorzystywane nie tylko do przygotowywania samych prognoz, które to następnie są brane pod uwagę przez ewaluatorów (tak było w przypadku ewaluacji *ex ante* dokumentów programowych na lata 2007–2013). Często modele są integralną częścią ewaluacji

---

<sup>13</sup> Ciekawym przykładem takiego postępowania, było upublicznienie przez Komisję Europejską IV Raportu Kohezyjnego. Przedstawione w nim wyniki prognozowania m.in. dla Polski stanowiły swoisty benchmark dla wyników prognoz przygotowywanych przez polskich ekonomistów.



*on-going*<sup>14</sup>, gdzie wykorzystywane są jako jedna z metod analizy, a nie jako jedyne źródło informacji.

Pozostaje jeszcze odpowiedź na pytanie, czy biorąc pod uwagę wszystkie trudności związane z wykorzystaniem wyników modelowania, powinno się podejmować wysiłek na rzecz dalszego zwiększania ich roli w procesie ewaluacji polityki spójności? Zdecydowanie tak! Korzyści z wykorzystania wyników modelowania są nieporównywalnie większe niż problemy, z którymi możemy się spotkać. Trzeba natomiast starać się ograniczać skalę ewentualnych trudności:

- wyniki modelowania powinny być prezentowane w sposób łatwy w odbiorze, a sam raport powinien być pisany językiem, który nie wymaga posiadania specjalistycznej wiedzy;
- w raporcie więcej miejsca powinno być poświęcone interpretacji kluczowych wyników, a nie „zalewaniu” odbiorcy badania olbrzymią ilością różnych danych, które są mało zrozumiałe;
- kontrowersyjne, zaskakujące wyniki nigdy nie powinny pozostawać bez odpowiedniego komentarza autorów badania, którzy z racji znajomości narzędzia są najwłaściwszymi osobami mogącymi poprawnie zinterpretować otrzymane wyniki;
- przed przystąpieniem do interpretacji wyników należy zawsze zapoznać się z założeniami procesu badawczego;
- za wynikami badania realizowanego przy wykorzystaniu modeli ekonomicznych powinni stać twórcy narzędzia, których wiedza i doświadczenie podnosić będzie wiarygodność otrzymanych szacunków i prognoz;
- z uwagi na trudności z porównywalnością wyników modelowania, tam, gdzie jest to możliwe, wskazana jest realizacja badania cyklicznie przy wykorzystaniu tego samego narzędzia;
- zwiększeniu dokładności i wiarygodności wyników badań służy zastosowanie równoległe więcej niż jednego modelu ekonomicznego<sup>15</sup>;
- w celu podniesienia precyzji otrzymywanych wyników, ważne jest, by modele ekonomiczne, które prezentują wyniki w skali makro, bazowały w swoich założeniach na badaniach ewaluacyjnych w skali mikroekonomicznej;

---

<sup>14</sup> Np. badanie dotyczące *Wpływu interwencji finansowanych z funduszy strukturalnych UE na zatrudnienie*, zrealizowane przez PAG Uniconsult wraz z Pentor Research International na zlecenie MRR. Ewaluacja bazowała zarówno na wynikach modelowania, jak i na rozległych badaniach ankietowych. Wyniki uzyskane różnymi technikami badawczymi były następnie ze sobą zestawione, co pozwalało na uzyskanie bardzo ciekawych, a jednocześnie wiarygodnych informacji. Innym badaniem, w ramach którego model był jednym z narzędzi badawczych, było badanie *Wpływu programów operacyjnych i NPR na realizację celów Strategii Lizbońskiej* zrealizowane przez firmę Ecorys na zlecenie MRR.

<sup>15</sup> W IV Raportie Kohezyjnym KE wykorzystano trzy różne modele ekonomiczne. Realizacja cyklicznych badań oddziaływania makroekonomicznego, przy wykorzystaniu trzech różnych modeli ekonomicznych, jest również planowana przez Ministerstwo Rozwoju Regionalnego.

- w badaniach ewaluacyjnych warto jest wykorzystywać model jako jedno z narzędzi badawczych, a nie jedyne źródło danych – łączenie różnych metod daje możliwość triangulacji i tym samym podnosi to wiarygodność otrzymanych wyników.

Na koniec warto jest jeszcze przyrzeć się kontekstowi, w jakim w najbliższych latach przyjdzie nam w Polsce wykorzystywać wyniki modelowania makroekonomicznego. Polska, począwszy od 2007 r., stała się największym beneficjentem funduszy strukturalnych UE. Z uwagi na skalę zaangażowanych środków staliśmy się swoistym laboratorium polityki spójności w skali całej Europy<sup>16</sup>. Od tego, czy będziemy w stanie trafnie, skutecznie i efektywnie wykorzystać tak duże fundusze, uzależniony będzie nie tylko przyszły interes Polski, lecz również przyszłość całej polityki spójności. Przed ewaluacją w naszym kraju stoi kluczowe zadanie w dostarczeniu argumentów w dyskusji nad kształtem i skalą polityki spójności po roku 2013. Wyniki modelowania makroekonomicznego, które pozwalają na pokazanie całościowych efektów tej polityki, będą jednym z najważniejszych argumentów w rozpoczynającej się debacie.

---

**Stanisław Bienias** – absolwent Szkoły Głównej Handlowej w Warszawie oraz podyplomowych studiów z zakresu ewaluacji projektów na Uniwersytecie Warszawskim. Od czterech lat pracuje w Ministerstwie Rozwoju Regionalnego, od 2006 roku na stanowisku naczelnika Wydziału Ewaluacji (Krajowa Jednostka Oceny). Odpowiedzialny za realizację procesu ewaluacji na poziomie Narodowego Planu Rozwoju na lata 2004–2006 i Narodowych Strategicznych Ram Odniesienia na lata 2007–2013 oraz za koordynację procesu ewaluacji na poziomie programów operacyjnych realizowanych w latach 2004–2006 i 2007–2013. Brał udział w realizacji ponad 50 badań ewaluacyjnych, był organizatorem i prelegentem licznych konferencji i seminariów dotyczących budowy kultury ewaluacyjnej w naszym kraju. Reprezentuje Polskę w spotkaniach „Evaluation Network” w ramach Dyrekcji Generalnej do Spraw Regionalnych w Komisji Europejskiej, współpracuje z OECD i innymi instytucjami krajowymi i zagranicznymi. Jeden z ekspertów obecnie realizowanego Twinningu Niemiecko-Polsko-Rumuńskiego. Prowadzi wykłady na studiach podyplomowych z zakresu ewaluacji, monitoringu i wskaźników. Współautor kilku publikacji z zakresu ewaluacji i funduszy strukturalnych.

---

<sup>16</sup> Więcej na ten temat w *Polityka spójności po 2013 r. Pożądane kierunki reformy*, MRR, styczeń 2008 r.

## Bibliografia

- Bradley J., Zaleski J., Tomaszewski P., Zembaty M., *Ocena efektu makroekonomicznego Regionalnego Programu Operacyjnego na lata 2007–2013 na gospodarkę Województwa Dolnośląskiego za pomocą modelu HERMIN*, Wrocław 2006.
- Kaczor T., Socha R., „Badanie wpływu Narodowego Planu Rozwoju 2004–2006 i Narodowych Strategicznych Ram Odniesienia 2007–2013 na wybrane wskaźniki dokumentów strategicznych”, Instytut Badań nad Gospodarką Rynkową, Prevision, kwiecień 2008 r.
- Polityka spójności po 2013 r. Pożądane kierunki reformy*, MRR, styczeń 2008 r.
- „Wpływ realizacji inwestycji finansowanych z funduszy unijnych na kształtowanie się głównych wskaźników dokumentów strategicznych – NPR i NSS oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modelu krajowego i modeli regionalnych Hermin”, Zespół Wrocławskiej Agencji Rozwoju Regionalnego pod kierownictwem prof. J. Zaleskiego we współpracy z dr J. Bradley’em, maj 2008 r.

*Dominik Batorski*

## **Metody analizy sieci i ich zastosowanie w ewaluacji**

Badania ewaluacyjne korzystają z wielu technik zbierania danych. Jednak w praktyce to, jakie dane są zbierane i wykorzystywane, zależy w dużej mierze od możliwości ich analizy. Celem niniejszego rozdziału jest przedstawienie metod analizy sieci społecznych (ang. *social network analysis*), nazywanych czasem także analizą sieciową. W odróżnieniu od klasycznych metod analiz prowadzonych w badaniach ilościowych, które koncentrują się na badaniu atrybutów jednostek (zmiennych), ich właściwości i zależności między nimi, analiza sieciowa wymaga badania relacji pomiędzy samymi jednostkami. Jest to o tyle istotne, że każdy z nas, a także firmy i organizacje, funkcjonuje w sieciach relacji z innymi osobami i organizacjami. Sieci te mogą mieć kluczowe znaczenie dla osiągnięć jednostek, sukcesu instytucji i konkretnych projektów. Badania pokazują, że charakter sieci, w które zaangażowane są jednostki, ma wpływ praktycznie na wszystko od zdrowia<sup>1</sup>, przez sukces zawodowy<sup>2</sup>, aż do tożsamości. Podobnie sytuacja organizacji w dużym stopniu zależy od otoczenia, w którym funkcjonują.

Wykorzystując dane o relacjach, analiza sieciowa pozwala na badanie struktury powiązań pomiędzy jednostkami, a także zależności struktury od atrybutów jednostek i wpływu na procesy, które zachodzą poprzez relacje (transakcje, przepływ informacji, kooperacja). Metody te wykorzystywane są nie tylko w socjologii, ale też w wielu innych dziedzinach. W badaniach organizacyjnych analiza struktury komunikacji w organizacji może pozwolić zidentyfikować nieformalne relacje, znaczenie poszczególnych jednostek dla przepływów informacji i wiedzy, a także zidentyfikować nieformalnych liderów<sup>3</sup>. W kryminologii metody sieciowe są wykorzystywane do ustalania relacji pomiędzy obiektami i zwalczania zorganizowanej przestępczości. W sektorze finansowym (bankach, firmach ubezpieczeniowych) analiza da-

---

<sup>1</sup> Osoby, które mają więcej bliskich relacji z innymi ludźmi, żyją dłużej. Ci, którzy mają bardziej zróżnicowane kontakty społeczne, są bardziej odporni i rzadziej chorują.

<sup>2</sup> Relacje i ich struktura wpływają na szansę znalezienia pracy, a także częstotliwość uzyskiwania awansów.

<sup>3</sup> Por. Kilduff M., Tsai W., *Social Networks and Organizations*, Sage, Londyn 2003.

nych relacyjnych pozwala na wykrywanie nadużyć. W epidemiologii i farmakologii służy badaniu rozprzestrzeniania się chorób i nałogów, oraz planowaniu skutecznego im przeciwdziałania. To jednak tylko przykłady – obszarów zastosowań analiz sieci jest znacznie więcej, bo wszędzie tam, gdzie relacje mają istotne znaczenie, badanie powiązań i ich struktury ma dużą wartość.

Niniejszy rozdział stanowi wprowadzenie do metod analizy sieciowej. Pierwsza część zawiera podstawowe definicje i pokazuje specyfikę podejścia sieciowego w porównaniu z tradycyjnymi badaniami. Druga część rozdziału poświęcona jest prowadzeniu badań sieciowych, metodom zbierania danych i ich typom. Trzecia część zawiera omówienie podstawowych metod analitycznych wykorzystywanych w analizie sieci. W ostatniej części przedstawiona została problematyka zastosowania podejścia sieciowego w badaniach ewaluacyjnych.

## **Sieci i podejście sieciowe**

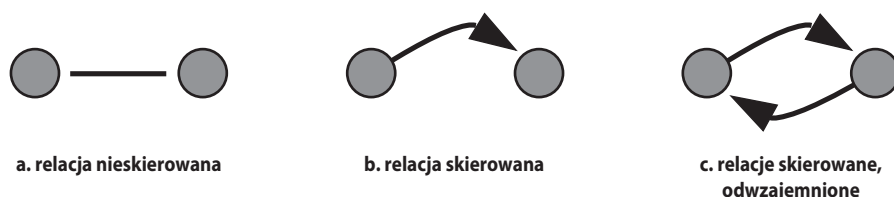
### **Podstawowe definicje**

W najprostszym ujęciu, sieć to każdy zbiór obiektów powiązanych pewnymi relacjami. Sieć społeczna to najczęściej zbiór osób połączonych społecznie znaczącymi relacjami. Jednak w praktyce analiza sieciowa nie ogranicza się do badania wyłącznie sieci społecznych i zarówno obiekty, jak i relacje w sieci mogą być bardzo różne. Jednostkami, czyli węzłami sieci (grafu), mogą być osoby, idee, publikacje, wydarzenia, projekty, organizacje, a także państwa. Podobnie jak w innych typach badań zależy to przede wszystkim od badanego problemu. Przykładowo możemy mówić o sieci w przypadku komunikacji między osobami, a także handlu i współpracy pomiędzy firmami, czy migracji między państwami. Wybór badanej populacji, a więc określenie granic badanej sieci, może stanowić poważny problem. Zanim jednak przejdziemy do omówienia tego problemu, należy nieco więcej miejsca poświęcić relacjom.

Jako że sieć składa się z jednostek i relacji między nimi, to koniecznym jest określenie nie tylko tego, jakie jednostki, ale również jakie relacje bierzemy pod uwagę. W gruncie rzeczy połączeniami między węzłami sieci mogą być wszelkiego rodzaju relacje między nimi. Często badanymi relacjami w tradycyjnej analizie sieci społecznych są na przykład: znajomości, komunikacja, współpraca, udzielanie porad. Danymi relacyjnymi są również członkostwo w tej samej organizacji, udział w tych samych spotkaniach, a pomiędzy organizacjami transakcje, przepływy finansowe, współpraca, a także wiele innych.

Z punktu widzenia analizy sieci społecznych istotnych jest kilka właściwości formalnych relacji, ponieważ wpływają one znacząco na to, jakie metody

analiz można zastosować. Zasadniczą sprawą jest oczywiście rodzaj relacji. A więc, o jaką relację chodzi, czego ona dotyczy oraz w jaki sposób możliwe jest ustalenie, czy pomiędzy dwoma jednostkami istnieje taka relacja, czy nie. Nie wystarczy na przykład powiedzieć – relacja przyjaźni, bez podania kryteriów zakwalifikowania relacji jako takiej<sup>4</sup>. Kolejną sprawą jest możliwość określenia kierunku relacji. Relacje mogą być skierowane lub nieskierowane (symetryczne). W pierwszym przypadku jest dokładnie określone, kto jest inicjatorem relacji, a kto jej odbiorcą. Na przykład X jest rodzicem Y, lub X jest zwierzchnikiem Z. Relacją symetryczną byłoby bycie rodziną, X i Y są rodziną, a także praca, przy tym samym projekcie X pracuje z Z.



Rysunek 1. Przykładowe typy relacji

Ważne są też topologiczne i metryczne własności relacji. Przede wszystkim możemy wyróżnić relacje, o których można powiedzieć, że mają charakter zero-jedynkowy – połączenie istnieje lub nie istnieje oraz takie, dla których możliwe jest określenie siły połączenia (na przykład częstotliwość odwiedzin, komunikacji, długość znajomości, wartość transakcji itp.). Czasami może to być znak relacji, na przykład pozytywna–negatywna, relacja lubienia–niełubienia.

W tym miejscu warto zaznaczyć, że na ogół węzłami w sieci są jednostki tego samego rodzaju – na przykład tylko osoby, albo tylko organizacje. Taką sieć nazywamy jednomodalną. Stosunkowo często badane są również sieci, w których węzły należą do dwóch bądź kilku typów (sieć dwu- wielomodalna). Sieci dwumodalne są najczęściej nazywane sieciami afiliacji, ponieważ relacje łączące węzły są zazwyczaj relacjami przynależności. Przykładem są relacje przynależności osób do organizacji, uczestnictwa w projektach, udziału w wydarzeniach.

Zdarza się również, że dla badanej zbiorowości możliwe jest ustalenie większej liczby różnych relacji. Mówimy wtedy o sieciach wielokrotnych. Są to sieci, w których te same jednostki są połączone różnymi relacjami – na przykład współpracą w tych samych projektach, wspólnym spędzaniem czasu

<sup>4</sup> Może to być na przykład obustronne stwierdzenie i opisanie drugiej osoby jako przyjaciela, a może być kategoria spełniająca określone warunki, takie jak wspólne spędzanie czasu wolnego, pomaganie sobie itp.

wolnego, dyskutowaniem spraw zawodowych itd. W przypadku organizacji mogą to być transakcje, współpraca, powiązania personalne (gdy te same osoby należą do dwóch lub więcej organizacji), współwłasność. Możliwe byłoby rozpatrywanie każdej z tych sieci z osobna, ale istnieją metody pozwalające analizować takie wielokrotne relacje łącznie. Wielokrotne połączenia są na ogół mocniejsze i trwalsze.

Od opisanych tu formalnych własności sieci (jednostek i relacji między nimi) w dużym stopniu zależą również metody, jakimi można i warto daną sieć analizować. Inne miary mogą być wykorzystane, gdy relacje mają określony kierunek i wartość, a inne, gdy powiązania nie mają takiego charakteru. Wszystko to, jak również konieczność używania osobnych programów komputerowych dedykowanych do analizy sieci, sprawia, że metody te mogą wydać się początkującym badaczom bardzo skomplikowane.

## Badanie i znaczenie struktur

Metody analiz sieci społecznych są szeroko stosowane w wielu dyscyplinach, przy badaniu bardzo różnych rodzajów zjawisk: poznanie kształtowania się postaw, ruchliwości społecznej i ekonomicznej, dyfuzji innowacji i informacji, komunikacji, struktury wspólnot, firm, organizacji, oraz stosunków międzyludzkich i zachowań politycznych. Jednak najczęściej służą po prostu do badania samej struktury. Zgodnie ze słownikową definicją, struktura to „zbiór, układ wzajemnie powiązanych ze sobą elementów i/lub układ relacji występujących pomiędzy częściami czy elementami jakiegoś systemu, pewnej całości”. Nieco więcej mówi inna definicja struktury, według której są to regularności we wzorach relacji między jednostkami<sup>5</sup>. Analiza sieci społecznych dąży przede wszystkim do odkrycia wzoru interakcji między ludźmi, organizacjami, państwami itp. Służy wskazaniu wzorców struktur społecznych i wpływu takich struktur na inne zmienne oraz zmian samej struktury.

Analiza sieciowa oparta jest na przekonaniu, że wzory relacji między ludźmi, a także między organizacjami, nie są przypadkowe. Oraz że to, jak ludzie się zachowują, w dużym stopniu zależy od relacji, w jakich są z innymi ludźmi (i jakimi ludźmi). Często też sukces organizacji, wspólnot, a nawet całych społeczeństw zależy od wzoru ich wewnętrznej struktury. Dobrym przykładem może być zależność wydajności pracy i efektywności zespołów zadaniowych od ich struktury<sup>6</sup>. Także działanie mechanizmów społecznych – na przykład wpływu społecznego, przepływu informacji, rozprzestrzeniania się innowacji, związane jest z ilością i jakością kontaktów społecznych i strukturą komunikacji.

<sup>5</sup> Knoke D., Kuklinski J.H., *Network analysis*, Sage 1982.

<sup>6</sup> Cummings J.N., Cross R., „Structural properties of work groups and their consequences for performance”, *Social Networks* 2003, nr 25(3), s. 197–210.

Dzięki metodzie analizy sieci społecznych możliwe jest badanie nie tylko całych grup lub tylko izolowanych relacji, ale różnych pośrednich zależności. Możemy badać gęstość sieci, wyodrębniać silniej powiązane podgrupy, analizować, czy sieć jest scentralizowana, czy urozmaicona w swojej wielkości i spójności, czy relacje są wąsko wyspecjalizowane, czy raczej wielotematyczne, oraz to, jak niebezpośrednie powiązania wpływają na zachowania jednostek. Analiza sieci społecznych pozwala uwzględnić (lepiej niż spojrzenie na całą grupę) kontekst, w którym znajduje się badana relacja i poszczególne jednostki.

## Podejście sieciowe

Analiza sieci wywodzi się z socjometrii stworzonej przez Moreno<sup>7</sup>. Pierwsza znacząca faza rozwoju metod analizy sieci nastąpiła w latach 70.<sup>8</sup> Następny etap rozwoju tej dziedziny miał miejsce w ciągu ostatnich kilkunastu lat. Po pierwsze, pojawiły się metody umożliwiające nie tylko opis sieci<sup>9</sup>, ale również wnioskowanie o zależnościach występujących w tych sieciach<sup>10</sup> i procesach w sieciach zachodzących<sup>11</sup>. Po drugie zaś, sieciami zainteresowali się matematycy i fizycy, wprowadzając nową perspektywę do badań społecznych<sup>12</sup>.

Jednak analiza sieci to coś więcej niż zestaw miar i narzędzi. To paradygmat i sposób podejścia do rzeczywistości społecznej i badania zjawisk społecznych, opierający się przede wszystkim na analizie relacji pomiędzy obiektami<sup>13</sup>. Analiza sieci jest w gruncie rzeczy osobną perspektywą badawczą

---

<sup>7</sup> Moreno J.L., „Foundations of sociometry: An introduction”, *Sociometry* 1941, nr 4(1), s. 15–35.

<sup>8</sup> Holland P.W., Leinhardt S. (red.), *Perspectives on social network research*, Academic Press 1979.

<sup>9</sup> Wasserman S., Faust K., *Social Networks Analysis: Methods and applications. Structural analysis in the social sciences*, Cambridge University Press 1994; Scott J., *Social Network Analysis: a handbook*, Sage Publications, Londyn 2000.

<sup>10</sup> Snijders T.A.B., Pattison P.E., Robins G.L., Handcock M.S., „New specifications for exponential random graph models”, *Sociological Methodology* 2006, s. 99–153.

<sup>11</sup> Robins G., Pattison P., „Random graph models for temporal processes in social networks”, *Journal of Mathematical Sociology* 2000, nr 25, s. 5–41; Robins G., Pattison P., Elliott P., „Network models for social influence processes”, *Psychometrika* 2001, nr 66, s. 161–190; Snijders T.A.B., „Models for Longitudinal Network Data”, w: P. Carrington, J. Scott, S. Wasserman (red.), *Models and methods in social network analysis*, Cambridge University Press, New York 2005, Chapter 11.

<sup>12</sup> Watts D.J., „Networks, dynamics, and the small-world phenomenon”, *American Journal of Sociology* 1999, nr 105(2), s. 493–527; Watts D.J., *Six degrees: the science of a connected age*, W.W. Norton 2003; Barabási A.-L., *Linked: how everything is connected to everything else and what it means for business, science and everyday life*, A Plume Book 2003; Newman M.E.J., „The structure and function of complex networks”, *SIAM Review* 2003, nr 45, s. 167–256.

<sup>13</sup> Knoke, D., Kuklinski J.H., op. cit.; Wellman B., „Structural analysis: from method and metaphor to theory and substance”, w: B. Wellman, S.D. Berkowitz (red.), *Social structures: a network approach*, Cambridge University Press 1988, tom 15 serii *Contemporary studies in sociology*, rozdział 2,



w obrębie nauk społecznych. W znacznie większym stopniu koncentruje się bowiem na relacjach pomiędzy wchodzącymi ze sobą w interakcje jednostkami. Istnieje też sporo teorii socjologicznych i psychologicznych nawiązujących do koncepcji sieciowych i wyjaśniających zjawiska społeczne z odwołaniem się do czynników relacyjnych i strukturalnych<sup>14</sup>.

Badając zjawiska społeczne, można obserwować zarówno cechy jednostek, jak i relacje między nimi. Rozważmy to na przykładzie. O grupie społecznej mówimy w kategoriach zależności pomiędzy cechami jednostek albo też w kategoriach relacji między jednostkami. W pierwszym przypadku, na przykład w szkole, za grupę możemy uznać uczniów chodzących do tej samej klasy. W drugim, za grupę uznamy uczniów raczej ze względu na relacje istniejące między nimi, a nie ze względu na sam fakt przynależności do jednej klasy. W wielu sytuacjach same atrybuty, które określają przynależność do grupy, nie wystarczą. Korzenie takiego podejścia znajdziemy w myśli Geорга Simm-la. Badając zjawiska społeczne, nie możemy analizować wyłącznie „takich czy innych właściwości jednostek jako jednostek w oderwaniu od ich stosunków wzajemnych”<sup>15</sup>.

Badania społeczne, w tym również znaczna część badań ewaluacyjnych, wymagają uwzględnienia obydwu możliwych podejść – zarówno atrybutów jednostek, jak i relacji między tymi jednostkami. Według Barry Wellmana: „Struktury relacji społecznych są dużo potężniejszym źródłem wyjaśnień socjologicznych niż atrybuty osób należących do danej zbiorowości”<sup>16</sup>. Podejście sieciowe wyróżnia szereg własności odmiennych od tradycyjnego podejścia:

- Zachowanie jest interpretowane w kategoriach strukturalnych ograniczeń nakładanych na działanie, a nie woluntarystycznie – w terminach wewnętrznych sił skłaniających jednostkę do działania i osiągnięcia jakiegoś celu.
- Analizy koncentrują się na relacjach pomiędzy jednostkami, zamiast dzielić jednostki na kategorie wyróżniane poprzez ich atrybuty.
- Głównym zagadnieniem jest to, jak wzorce relacji pomiędzy innymi jednostkami wpływają na zachowania aktorów będących w sieci. Tym samym też, relacje nie są traktowane w izolacji (od innych relacji). Podejście sieciowe jest istotne, ponieważ przebieg wielu procesów społecznych zależy od relacji pomiędzy jednostkami i strukturami tworzonymi przez relacje.
- Struktura jest postrzegana jako sieć sieci, która może, ale nie musi być dzielona na rozłączne grupy. Nie trzeba zakładać, że zwarte i dobrze wy-

---

s. 19–61; Wasserman S., Faust K., op. cit.; Hanneman R.A., Riddle M., *Introduction to social network methods*, University of California, Riverside 2005. Opublikowane elektronicznie na: <http://faculty.ucr.edu/~hanneman/>.

<sup>14</sup> Monge P.R., Contractor N.S., *Theories of communication networks*, Oxford University Press, New York 2003.

<sup>15</sup> Szacki J., *Historia myśli socjologicznej*, Warszawa 1981, s. 512.

<sup>16</sup> Wellman B., op. cit., 1988, s. 31.

odrębnione grupy są elementami składowymi struktury, bo nie wszystkie sieci muszą mieć grupową strukturę.

- Metody analizy bezpośrednio zajmują się wzorcami i relacyjną naturą struktury społecznej. Czasem zastępują, a czasem uzupełniają mainstreamowe metody statystyczne, które wymagają niezależnych jednostek analizy.

## Organizacja badań sieciowych

### Wybór badanej zbiorowości

Jednym z pierwszych etapów badania sieci społecznych jest określenie granic badanego układu, a więc tego, jakie jednostki składają się na sieć i jakie relacje pomiędzy nimi mają być analizowane. Nie jest to żadnym problemem w sytuacji, gdy badana społeczność, grupa lub organizacja jest mała i dobrze zdefiniowana. Również wtedy, gdy węzłami sieci są projekty w ramach większego programu, lub firmy bądź organizacje, wyodrębnienie badanych jednostek bywa znacznie prostsze. Dużo trudniejsza jest sytuacja, gdy badana populacja jest bardzo liczna. Nie chodzi tylko o to, że trudne jest dotarcie do każdej osoby, ale bardziej o problem zidentyfikowania wszystkich istniejących relacji. Zadawanie nawet jednej osobie pytań o jej relacje w stosunku do każdej innej osoby z badanej populacji byłoby bardzo czasochłonne. Znacznie częściej też dla dużych sieci łatwiej jest ustalić sieci afiliacji.

### Dane sieciowe i ich zbieranie

Oprócz wspomnianego wcześniej problemu wyodrębnienia granic badanego układu zdarza się wiele innych problemów ze zbieraniem danych sieciowych. Szczególne trudności związane są z badaniem dużych zbiorowości. W badaniach sieci społecznych, w sytuacji, kiedy jest to możliwe, bada się całą sieć. Od wszystkich osób zbierane są informacje o ich relacjach ze wszystkimi innymi osobami, co pozwala na dokładne określenie pozycji każdej osoby w sieci. Ponieważ badanie takie możliwe jest tylko wtedy, kiedy sieć jest stosunkowo nieduża i ma jasno określone granice, taka metoda zbierania danych jest najbardziej adekwatna do badania niewielkich populacji. Pewnym wyjątkiem są sytuacje, gdy możliwe jest wykorzystanie danych zastanych o wszystkich jednostkach należących do zbiorowości, którą chcemy badać i relacjach między nimi.

Częściowym rozwiązaniem problemu badania dużych populacji jest badanie tylko niektórych – najważniejszych relacji. W przypadku dużej zbiorowości osób można zamiast pytań o relacje ze wszystkimi osobami (liczba pytań

jest tu proporcjonalna do liczebności badanej grupy), prosić o wymienienie wszystkich przyjaciół, albo też zapytać jedynie o troje najważniejszych. Podobnie, w badaniach prowadzonych wewnątrz dużych organizacji, zadaje się pytania nie o każdą inną osobę, a jedynie prosi się o wybranie osób, z którymi dana jednostka najczęściej współpracuje, rozmawia o sprawach zawodowych, których się radzi, i innych w zależności od problematyki badania. Jednak ten sposób badania ma również swoje wady, prowadzi bowiem do uzyskania danych jedynie o najważniejszych, a więc i najmocniejszych relacjach. Uzyskanie informacji o słabych więziach jest wówczas właściwie niemożliwe. Jest to problemem szczególnie w takich badaniach, w których słabe relacje mają duże znaczenie – np. w badaniach kapitału społecznego, gdzie słabe relacje są kluczowe dla przepływów informacji<sup>17</sup>.

Inną możliwością uniknięcia trudności związanych z wyczerpującym badaniem dużej populacji jest przeprowadzenie go na próbie osób<sup>18</sup>. Trudno dowiedzieć się o relacje między wszystkimi osobami, gdy jest ich dużo, dlatego też badania są prowadzone na próbach. Takie rozwiązanie jest również lepsze w sieciach o gorzej zdefiniowanych granicach. W przypadku badań realizowanych na próbach badane są tak zwane sieci egocentryczne. Tego typu badania są w największym stopniu zbliżone do tradycyjnych badań. Zbierane są dane o poszczególnych respondentach (*ego*) oraz o osobach, z którymi respondenci są połączeni (*alter*). Często nie jest to tylko określenie relacji dla badanych jednostek, ale również estymacja relacji pomiędzy kontaktami danej osoby, robiona na podstawie pytań zadawanych respondentowi albo na podstawie informacji uzyskanych bezpośrednio od jego kontaktów. Takie podejście jest najczęściej stosowane w badaniach bardzo dużych sieci, gdy nie dysponujemy danymi zastanymi, a badanie całej populacji i relacji pomiędzy wszystkimi osobami nie jest możliwe. W przypadku badań, które obejmują sieci egocentryczne oraz pewien stopień dotarcia do szerszych kręgów społecznych (na przykład znajomych znajomych), mówimy o badaniach sieci cząstkowych (*partial network*). Jednocześnie jest to mniej pracochłonne niż ustalenie powiązań pomiędzy wszystkimi parami aktorów w badanej zbiorowości.

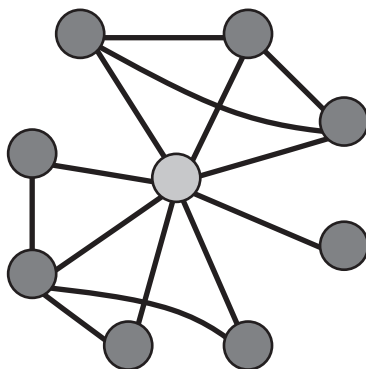
Dobrym przykładem badań sieciowych prowadzonych na próbach jest wiele badań poświęconych kapitałowi społecznemu<sup>19</sup>. Zazwyczaj używana jest technika generowania imion (*name generator*). Jednostka wymienia osoby, z którymi jest w relacji i opisuje tę relację. Inną metodą jest wykorzystywa-

<sup>17</sup> Granovetter M., „The strength of weak ties”, *American Journal of Sociology* 1973, nr 73, s. 1360–1380.

<sup>18</sup> Erickson B.H., Nosanchuk T.A., „Applied network sampling”, *Social Networks* 1983, nr 5, s. 367–382.

<sup>19</sup> Np.: Lin N., Cook K., Burt R.S. (red.), *Social capital: theory and research*, 2001; Burt R.S., *Brokerage and Closure: An introduction to Social Capital*, Oxford University Press 2005.

na w badaniach kapitału społecznego technika generowania pozycji (*position generator*). Respondent jest proszony o podanie, czy zna jakieś osoby zajmujące pozycje dające dostęp do zasobów cenionych w zbiorowości<sup>20</sup>.



Rysunek 2. Przykładowa sieć egocentryczna

Rodzaj zbieranych i analizowanych danych (cała sieć vs. sieci egocentryczne) w zasadniczy sposób związany jest z wyborem badanej populacji i typem badanej sieci. Z drugiej strony, możliwe do przeprowadzenia analizy w dużym stopniu zależą od rodzaju danych, którymi dysponujemy. Wielu analiz, które są możliwe dla danych dotyczących całej sieci, nie da się zrobić, mając jedynie dane o wybranych sieciach egocentrycznych.

Ze względu na wspomniane trudności z badaniem dużych populacji, stosunkowo często zbiera się w nich dane o sieci afiliacji. Często tego typu dane są też znacznie łatwiej dostępne. Co więcej, sieć afiliacji (a więc sieć dwumodalna) może zostać zredukowana do dwóch sieci jednomodalnych. Przykładowo sieć przynależności osób do projektów może zostać zastąpiona siecią relacji między projektami, poprzez posiadanie tych samych osób, oraz siecią relacji między osobami, gdzie relacją jest przynależność do tych samych projektów.

Innym, niezwykle istotnym problemem badań sieciowych jest zbiór danych wrażliwych, a informacje o relacjach często do takich należą. Dotyczy to zarówno wielu relacji interpersonalnych, jak również relacji pomiędzy firmami (na przykład zawieranych transakcji i ich wartości). Skłonność respondentów do podawania tego typu danych jest niewielka, a dodatkowo przed rozpoczęciem badań należy rozważyć kwestie ich etyczności. W przypadku badań sieciowych, szczególnie takich, gdzie badana jest nieduża zbiorowość, znacznie trudniej jest zagwarantować anonimowość badanych.

<sup>20</sup> Lin N., Fu Y., Hsung R.-M., „The position generator: measurement techniques for investigations of social capital”, w: N. Lin, K. Cook, R.S. Burt (red.), op. cit., s. 57–81.

Przede wszystkim, gdy prezentowane wyniki dotyczą poszczególnych osób i gdy są w postaci graficznej, to osoby należące do badanej populacji bardzo często mogą się zorientować się, kto jest kim. Dlatego też badacz powinien zachować szczególną ostrożność, aby ani procedura przeprowadzania badania, ani prezentowane wyniki nie zaszkodziły badanym jednostkom i ich relacjom.

Należy równocześnie podkreślić, że braki danych w badaniach sieci są bardzo niebezpieczne. W porównaniu ze zwykłymi badaniami ilościowymi w badaniach sieci każdy brak danych ma dużo większe znaczenie. Powoduje bowiem utratę danych o relacjach, czego konsekwencją jest jednocześnie wypaczenie informacji o znacznej części osób, które wzięły udział w badaniu i które mogły być w relacjach z jednostkami, do których nie udało się dotrzeć. Braki danych uniemożliwiają na przykład dokładne stwierdzenie stopnia odwzajemnienia relacji oraz wielu innych bardziej złożonych analiz. Szerzej o doborze próby w badaniach sieci społecznych pisali Erickson i Nosanchuk<sup>21</sup> i Frank<sup>22</sup>.

Relacje społeczne istniejące w różnych zbiorowościach są często postrzegane w sposób odbiegający od rzeczywistości. Wpływ na to, jak jednostki postrzegają relacje, ma ich osobowość<sup>23</sup>, a także to, na ile blisko danej relacji są i jak dobrze ją znają<sup>24</sup>. Różnice pomiędzy percepcją a rzeczywistością komunikacją mają swoje konsekwencje dla jakości zbieranych danych, powinny więc mieć także wpływ na sposoby ich zbierania<sup>25</sup>. Jest to kolejny argument, obok niedogodności związanych z doбором próby i zagrożeniami związanymi z brakiem danych, ograniczający prowadzenie badań za pomocą wywiadów.

Metody zbierania danych w analizach sieci mogą być bardzo różne i nie muszą się ograniczać do przeprowadzania wywiadów lub ankiet. Wykorzystywane są również: obserwacje, dzienniczki i oraz dane zastane z różnego rodzaju archiwów, spisów i rejestrów. W praktyce każdy sposób zbierania danych może być wykorzystywany również w badaniach sieciowych. Często też łączy się wiele różnych metod.

---

<sup>21</sup> Erickson B.H., Nosanchuk T.A., op. cit.;

<sup>22</sup> Frank O., „Network sampling and model fitting”, w: P.J. Carrington, J. Scott, S. Wasserman (red.), op. cit..

<sup>23</sup> Casciaro T., „Seeing things clearly: social structure, personality, and accuracy in social network perception”, *Social Networks* 1998, 20(4), s. 331–351.

<sup>24</sup> Krackhardt D., Kilduff M., „Whether close or far: Social distance effects on perceived balance in friendship networks”, *Journal of Personality and Social Psychology* 1999, nr 76, s. 770–782.

<sup>25</sup> Marsden P.V., „Network data and measurement”, *Annual Review of Sociology* 1990, nr 16, s. 435–463.

## Metody analiz sieci

Analizy sieci społecznych można prowadzić w różny sposób. Przede wszystkim możliwe są różne poziomy analizy sieci. Poziom jednostek, czyli sieci egocentrycznych, koncentruje się na relacjach posiadanych przez jednostkę i powiązanych z nią osób. Kolejnym poziomem jest poziom podgrup wyodrębnionych w sieci, a ostatnim poziom całej sieci. W analizie sieci społecznych istnieje niezmiernie dużo metod i możliwości analizowania własności relacji oraz ich struktur. Istotnym elementem badań jest wizualizacja danych sieciowych. Sieci są przedstawiane za pomocą grafów, składających się z węzłów i łączących je krawędzi. Obecnie grafy coraz bardziej są metodą prezentacji i wstępnej eksploracji danych niż analiz<sup>26</sup>.

W tym miejscu scharakteryzowane zostaną wyłącznie najbardziej podstawowe zagadnienia analizowane przy użyciu danych sieciowych. Dla lepszego wyjaśnienia opisywanych metod odwołam się do prostych przykładów sieci. Kilka przedstawionych dalej zastosowań tych metod odnosić się będzie do sieci współpracy pomiędzy organizacjami, których działania dotyczą podobnych spraw. Możliwe byłoby ich zastosowanie w ewaluacji chociażby projektów szkoleniowych, mających na celu promocję *clusteringu* jako formy kooperacji pomiędzy przedsiębiorstwami oraz partnerami publicznymi i trzecim sektorem<sup>27</sup>. Celem tego rodzaju projektów jest: powstanie trwałych form współpracy ukierunkowanych na tworzenie się lokalnych „wspólnot” wypracowujących wspólną markę oraz utrzymujących wielopłaszczyznowe, intensywne kontakty (np. Dolina Lotnicza, Podkarpacie).

## Sieciowe charakterystyki jednostek

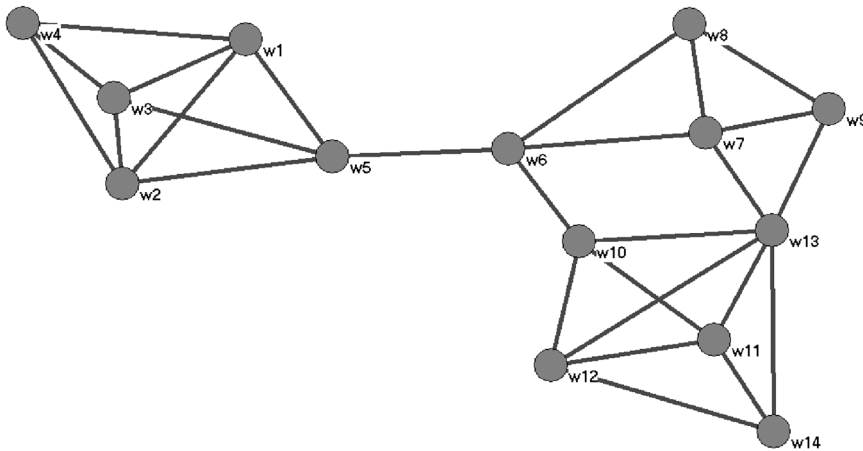
Jednym z podstawowych pytań analizy sieci społecznych, jest pytanie o pozycję jednostki. Bardzo często formalna struktura osób w organizacji bądź organizacji w większych projektach nie odpowiada rzeczywistości, a nieformalne relacje i hierarchie mają olbrzymie znaczenie dla funkcjonowania poszczególnych jednostek i całych organizacji. Analiza sieci umożliwi scharakteryzowanie pozycji jednostki w strukturze sieci na podstawie jej relacji z innymi jednostkami. Pozwala odpowiedzieć na pytania, które jednostki są bardziej centralne, a które położone na peryferiach – które mają

---

<sup>26</sup> de Nooy W., Batagelj V., Mrvar A., *Exploratory Social Network Analysis with Pajek*, Cambridge University Press 2005.

<sup>27</sup> Przykładowym projektem tego typu jest „Turystyka wspólna sprawa”, por. [www.tws.org.pl](http://www.tws.org.pl), <http://www.parp.gov.pl/index/more/24>. Jednak możemy wyobrazić sobie dowolne inne przykłady organizacji mających inne cele – np. działających na rzecz promocji zdrowego trybu życia, albo też instytucje, których działania mają znaczenie dla informatyzacji kraju.

większe, a które mniejsze znaczenie? Jakie jest znaczenie położenia w sieci dla dostępu do informacji i kontroli przepływów informacji (pozycja brokera)? Jakie jest znaczenie pozycji w sieci dla dostępności szans i możliwości, którym jednostkom ich położenie w strukturze daje przewagę konkurencyjną, a wobec których jest ograniczające<sup>28</sup>? Które jednostki dysponują większą władzą niż inne? Wszystkie te pytania dotyczą pozycji jednostek w stosunku do innych osób.



Rysunek 3. Przykładowa sieć składająca się z 14 węzłów o zróżnicowanej pozycji

Najprostszą charakterystyką pozycji jest liczba relacji posiadanych przez daną jednostkę (*degree*). Na rysunku 3 węzły mają od 3 do 6 relacji (węzeł 13). Jednostki posiadające więcej relacji mają na ogół większe znaczenie i możliwości, choć nie jest to regułą. W przypadku, gdy relacje w analizowanej sieci mają określony kierunek, możliwe jest również dokonanie rozróżnienia ze względu na wybory dokonywane przez jednostkę i jej partnerów. Możemy wyróżnić liczbę relacji inicjowanych przez jednostkę (*outdegree*) oraz liczbę relacji przychodzących (*indegree*), ta druga jest dobrą miarą popularności bądź prestiżu. W tym przypadku, badając liczbę wskazań, jakie dana osoba otrzymuje (jak często jest wybierana przez inne), możemy mówić o poziomie prestiżu lub popularności. Możliwe jest również uwzględnienie tego, jak bardzo popularne są osoby wybierające. W takim przypadku fakt wybrania przez osobę popularną może być dużo istotniejszy niż bycie wybranym przez kilka osób, których nikt nie wybrał.

W przypadku relacji skierowanych możliwe są cztery typy bezpośrednich połączeń między dwoma węzłami (często nazywanymi diadą). Węzły mogą

<sup>28</sup> Por. Burt R.S., *Structural holes: the social structure of competition*, Harvard University Press 1992; Burt R.S., *Brokerage and Closure...*, op. cit.

nie być powiązane, relacja może być nieodwzajemniona (wychodząca lub przychodząca), albo może zachodzić w obie strony (odwzajemniona). W wielu sieciach stopień odwzajemnienia relacji jest ważną charakterystyką sieci. Jednak posiadając dane o całej sieci, mamy możliwość analizowania nie tylko diad, które są powiązane bezpośrednimi relacjami, ale także takich, które są połączone poprzez inne jednostki. O ile sieć jest połączona (o czym szerzej w kolejnej części), to możliwe jest określenie dystansu dla każdej pary węzłów, poprzez ścieżki relacji i inne węzły. W mierzeniu dystansu wykorzystuje się na ogół najkrótsze ścieżki (tzw. *geodesic distance*). Najdłuższą spośród najkrótszych ścieżek w sieci nazywa się średnicą sieci (*diameter*). W przypadku grafu zaprezentowanego na rysunku 3 najbardziej oddalone od siebie są węzły 4 i 14, a średnica sieci wynosi 6.

Pozycje w sieci związane z władzą, większym dostępem do informacji i większymi możliwościami wywierania wpływu charakteryzuje się za pomocą pojęcia centralności. Miar centralności jest dużo<sup>29</sup>, ale w praktyce wszystkie one opierają się na różnych koncepcjach zaangażowania jednostki w relacje w sieci. Im większe zaangażowanie we wszystkie relacje sieci, tym większa centralność danego aktora. Same relacje mogą być jednak różnie definiowane. Wspomniana powyżej liczba relacji (*degree*) jest najprostszą miarą centralności jednostki w sieci, dlatego też bywa również określana jako *actor centrality*. Typowym przykładem jednostki centralnej jest gwiazda socjometryczna. Jednak ta miara uwzględnia jedynie bezpośrednie zaangażowanie w relacje. Uwzględnienie powiązań niebezpośrednich w określaniu pozycji węzła w sieci prowadzi do wyodrębnienia bardziej złożonych, ale też mających większą wartość informacyjną, indeksów centralności pozycji węzła. Warto podkreślić, że większość z tych miar wymaga uwzględnienia całej sieci, a nie tylko relacji jednostki.

Przykładowo *betweenness centrality* to częstość, z jaką dana jednostka pojawia się w najkrótszych ścieżkach pomiędzy różnymi parami węzłów. Potencjalnie są to osoby posiadające większą kontrolę przepływu informacji i wpływ na inne osoby. Na rysunku 3 jednostką o najwyższej wartości tego indeksu jest węzeł 6, który kontroluje komunikację pomiędzy wieloma fragmentami sieci. Niewiele mniej ważny jest węzeł 5, a dalej kolejno 10, 13 i 7. Warto zwrócić uwagę na różnice pomiędzy węzłami 6 i 13, ten pierwszy ma tylko cztery bezpośrednie połączenia, drugi aż sześć, jednak to węzeł 6 ma znacznie istotniejszą pozycję.

Inną miarą centralności jest *closeness centrality*, określana poprzez odległość jednostki od wszystkich innych w sieci. Freeman uznaje to za miarę niezależności od innych osób i odporności na wpływ<sup>30</sup>. Po tym względem

<sup>29</sup> Freeman L.C., „Centrality in social networks: I. conceptual clarification”, *Social Networks* 1979, nr 1, s. 215–239; Wasserman S., Faust K., op. cit.

<sup>30</sup> Freeman L.C., op. cit.



również najlepsza jest pozycja węzła 6, jednak niewiele gorsza jest pozycja węzłów 7 i 10, których średnia odległość od innych węzłów w sieci jest mniejsza niż dla węzła 5 i pozostałych jednostek.

Warto też zaznaczyć, że niektórzy autorzy, między innymi Scott<sup>31</sup> rozróżniają lokalną i globalną centralność. Rozróżnienie to ma szczególnie sens dla dużych sieci społecznych, w których liczba węzłów jest znacznie większa niż średnia liczba relacji. Miary lokalne pokazują pozycję węzła w jego otoczeniu (sąsiedztwie), a drugie w całej sieci.

Dodatkowo dla każdej miary centralności możliwe jest także policzenie odpowiadającej jej miary centralizacji całej sieci. Miary centralizacji przyjmują wartości od zera do jeden i określają, na ile dana sieć jest zcentralizowana w stosunku do maksymalnie zcentralizowanej sieci. Dzięki temu możliwe jest porównywanie różnych sieci pomiędzy sobą.

Znaczenie pozycji węzła w sieci może opierać się również na innych charakterystykach niż opisane tu miary centralności. Oba prezentowane podejścia mogą dawać podobne rezultaty, jako że opisane tu własności przekładają się na wyższą centralność węzłów. Pewne osoby w sieci mogą mieć kluczowe znaczenie dla przepływów informacji, gdy są łącznikami pomiędzy różnymi fragmentami sieci. Pozycję takiego brokera zajmują węzły, których usunięcie z sieci powodowałoby rozłączenie większych fragmentów sieci (w teorii grafów nazywa się je punktami odcięcia). W sieci zaprezentowanej na rysunku 3 takimi węzłami są węzły 5 i 6. Łącznikiem pomiędzy dwoma grupami może być również pojedyncza relacja pomiędzy dwoma jednostkami (relacja między węzłami 5 i 6 na rysunku 3). Takie relacje, których usunięcie z sieci powoduje jej rozłączenie, nazywane są mostami. Podsumowując, można stwierdzić, że znaczenie poszczególnych węzłów i relacji dla komunikacji, stabilności i efektywności procesów zachodzących w sieci oraz innych procesów w zależności od rodzaju sieci, może zostać ustalone poprzez analizowanie efektów usunięcia danego węzła/relacji z istniejącej struktury.

Poszczególne jednostki mogą być opisywane również za pomocą miar charakteryzujących ich otoczenie w sieci. Przykładem takiej miary jest gęstość lokalna (*ego-network density*)<sup>32</sup>, czyli proporcja połączeń między węzłami powiązanych z jednostką. Przykładowo na rysunku 3, prawie wszystkie jednostki, będące w relacji z węzłem 11, mają relacje między sobą (lokalna gęstość wynosi 5/6), inaczej jest na przykład w przypadku węzłów 7 i 6.

Analizy charakterystyk położenia jednostki w sieci moglibyśmy z powodzeniem zastosować we wspomnianym wcześniej przykładzie analizowania sieci współpracy pomiędzy organizacjami. Pozwoliłoby to odpowiedzieć na

<sup>31</sup> Scott J., *Social Network Analysis: a handbook*, Sage Publications, Londyn 2000, s. 82.

<sup>32</sup> Gęstość lokalna bywa również nazywana inaczej. W literaturze stosowanych jest kilka nazw, między innymi jednostkowa spistość, a także *clustering*.

pytania, które spośród organizacji pracujących przy realizacji większego programu mają większe znaczenie dla jego realizacji. Które z nich współpracują z większą liczbą partnerów, i których pozycja jest ważniejsza dla efektywnego współdziałania wielu podmiotów, a także bez których podmiotów nie sposób byłoby się obejść, realizując dany program? Długość ścieżek pomiędzy jednostkami pozwala stwierdzić, na ile możliwe są przepływy wiedzy i doświadczeń pomiędzy poszczególnymi organizacjami. Prowadzenie badań w czasie pozwoliłoby dodatkowo sprawdzić, jaka jest trwałość sieci współpracy w czasie.

### Struktury grupowe

Większość miar, które stosuje się do opisu pozycji pojedynczych osób, można po niewielkich modyfikacjach stosować do opisu właściwości podgrup w sieci. Podgrupy takie mogą być określone na podstawie atrybutów jednostek (np. płci w przypadku osób, branży w przypadku firm itp.). Możliwe jest jednak również wyodrębnienie podgrup i dokonanie podziału sieci na podstawie wzorców relacji. Z tej perspektywy grupa jest grupą ze względu na specyficzną strukturę relacji między jednostkami ją tworzącymi, a nie tylko ze względu na posiadanie przez jednostki jakiejś jednej cechy czy charakterystyki.

We wspomnianym przykładzie moglibyśmy wyodrębnić grupy organizacji zarówno na podstawie ich atrybutów (np. firmy, instytucje administracji, organizacje pozarządowe), jak również na podstawie tego, kto z kim współpracuje, wyodrębniając w ten sposób ściśle współpracujące grupy. To, na ile sieć współpracy organizacji ma grupowy charakter, może mieć kluczowe znaczenie dla powodzenia i efektywności całego realizowanego przez nie programu. Czasem też powstanie trwałych form współpracy pomiędzy organizacjami jest celem samym w sobie. W obu przypadkach wyodrębnienie grup współpracujących organizacji, możliwe dzięki metodom sieciowym, jest niezbędne. Ważne są również pytania: czy działania podejmowane są niezależnie przez grupy organizacji, z których każda tworzy zamkniętą klikę, niewspółpracującą i niewymieniającą doświadczeń z innymi, czy też współpraca obejmuje znacznie więcej podmiotów i nie jest możliwe wyodrębnienie grup organizacji?

Zanim jednak przejdziemy do problemu wyodrębniania grup w ramach sieci, warto omówić nieco bardziej podstawowe koncepcje związane z podziałem węzłów należących do sieci na fragmenty. Przede wszystkim należy sobie zadać pytanie, czy sieć, którą analizujemy, jest połączona, to znaczy, czy każda para węzłów jest powiązana przez ścieżkę relacji? Jeżeli nie, to znaczy, że sieć ma więcej niż jeden komponent. Komponentem sieci nazywa się taki maksymalny zbiór węzłów, dla którego możliwe jest powiązanie dowolnej pary węzłów w tym komponencie poprzez relacje i inne węzły należące do

tego samego komponentu. Wyróżnia się różne typy komponentów ze względu na wymagania względem typu relacji łączących węzły<sup>33</sup>.

W analizie sieci możliwe są dwa podstawowe podejścia i sposoby myślenia o zbiorowościach i grupach w obrębie sieci, jako strukturach wyróżnianych na podstawie danych empirycznych – połączeń między jednostkami. Pierwszym jest kohezja społeczna: grupy wyodrębniane są na podstawie połączeń między konkretnymi aktorami. Drugim ekwiwalenty strukturalne: gdy grupy identyfikuje się na podstawie podobieństwa kontaktów z innymi osobami. Oczywiście oba te podejścia nie wykluczają wyodrębnienia grup na podstawie innych – niesieciowych atrybutów. W tym miejscu zajmiemy się wyodrębnianiem spójnych podgrup na podstawie kohezji społecznej, natomiast ekwiwalenty strukturalne zostaną omówione w następnym części rozdziału.

Na poziomie lokalnym spójność sieci można rozważać poprzez analizowanie relacji w triadach – tzn. pomiędzy trzema jednostkami, bo dopiero w przypadku trzech jednostek można mówić o grupie. Teoretycznie możliwych jest 16 różnych kombinacji relacji w triadach. Interesujące jest, jakiego rodzaju triady występują jak często? Na podstawie takiej informacji (tak zwanego cenzusu triadowego) możliwe jest dokonanie analiz dotyczących przechodniości relacji, a także spójności sieci i jej fragmentów<sup>34</sup>. Większa przechodniość relacji, czyli większe prawdopodobieństwo, że dwie jednostki, które są w relacjach z trzecią są również w relacji ze sobą, występuje w strukturach, w których podziały na grupy są silniejsze.

Spójne podgrupy to zbiory aktorów, pomiędzy którymi są relatywnie mocne, bezpośrednie, intensywne, częste lub pozytywne relacje. Jednak ze względu na to, że pojęcie grupy społecznej może odnosić się do wielu różnych własności sieci społecznych, to samych definicji podgrup sieci społecznych jest również bardzo wiele. Podstawowe własności w konceptualizacjach pojęcia podgrup to:

- wzajemność (*mutuality*) relacji,
- bliskość (*closeness*) i dostępność (*reachability*) członków podgrupy,
- częstość relacji pomiędzy członkami grupy,
- względna częstość relacji pomiędzy członkami grupy w porównaniu do relacji z innymi jednostkami.

Z tego też wywodzą się najważniejsze podejścia do wyodrębniania spójnych podsieci:

- kompletna wzajemność,

<sup>33</sup> Mówi się o słabych komponentach, gdy kierunek relacji nie jest brany pod uwagę. O mocnych komponentach, gdy dla każdej pary węzłów A i B możliwe jest poprowadzenie ścieżki relacji skierowanych zarówno od A do B, jak i od B do A. Oraz o rekursywnych komponentach, gdy ścieżki prowadzą przez odwzajemnione relacje.

<sup>34</sup> Por. Heider F., „On balance and attribution”, w: P.W. Holland, S. Leinhardt (red.), *Perspectives on social network research*, Academic Press 1979, s. 11–23.

- dostępność i odległość,
- *degree* węzłów (bezpośrednie połączenie z odpowiednio wieloma członkami podgrupy)<sup>35</sup>.

Formalnych definicji podgrup jest bardzo dużo. Wybór konkretnej powinien być uzasadniony charakterem sieci i relacji łączących węzły. Jedną z podstawowych definicji grupy o dużej spójności w obrębie sieci jest klika. Jest to maksymalny<sup>36</sup> kompletny podgraf, a więc zbiór kompletnie połączonych węzłów. Innymi słowy każda jednostka musi mieć relacje z każdą inną należącą do tej samej kliki. Kliki w sieci (szczególnie większej) mogą się częściowo nakładać, podział na podgrupy nie jest więc podziałem rozłącznym i jeden węzeł może należeć do więcej niż jednej kliki. W sieci pokazanej na rysunku 3 największe kliki liczą po cztery węzły – np. węzły {1, 2, 3, 4} lub {1, 2, 3, 5}. Liczne są także kliki łączące trzy jednostki.

Ze względu na wymaganie kompletności definicja klik jest niezwykle restrykcyjna i brak jej odporności na błędy pomiaru. Brak jednej relacji powoduje, że podgrupa nie jest kliką, co więcej wielkość podgrup jest ograniczona przez *degree* aktorów. Nie ma też praktycznie żadnych (strukturalnych) różnic pomiędzy członkami kliki, nie jest więc możliwe badanie wewnętrznej struktury. Dlatego też różne inne definicje przyjmują, iż nie jest konieczne, aby relacja istniała między każdą parą członków grupy. Inne definicje uwzględniają również odległości pomiędzy osobami lub proporcje relacji w obrębie grupy (połączenia wewnątrz grupy są gęste, a na zewnątrz rzadkie).

W oparciu o koncepcję dostępności (*reachability*), moglibyśmy zdefiniować grupę jako zbiór aktorów blisko ze sobą związanych (na podstawie danej relacji). Wywodzące się z tej koncepcji pojęcie *n*-kliki to maksymalny podgraf, w którym do każdego elementu z kliki można dojść w *n* krokach. Gdy *n* = 1, mamy doczynienia ze zwykłymi klikami. Najczęściej stosowany jest dystans *n* = 2. Według tej definicji nie wszystkie ścieżki muszą przechodzić przez osoby z danej *n*-kliki i może się zdarzyć, że odległość *n* jest tylko poprzez węzły należące do innej podgrupy. Dlatego też średnica podgrupy może być większa od *n*. Inaczej jest w przypadku tak zwanych *n*-clans, czyli *n*-klik, w których wszystkie odległości przechodzą przez ścieżki w ramach podgrafu. Zgodnie z tymi definicjami w grafie z rysunku 3 można byłoby wyodrębnić między innymi takie grupy węzłów, jak {1, 2, 3, 4, 5} lub {6, 7, 9, 10, 11, 12, 13}.

Sama odległość nie jest też najlepszym wskaźnikiem spójności (dobrym przykładem jest tu druga z wymienionych przed chwilą podgrup). Dlatego czasem wprowadza się inne definicje. Stosunkowo często wykorzystywane są definicje *k*-plex i *k*-cores. W pierwszym przypadku każdy węzeł należący do podgrupy liczącej *g* węzłów musi być bezpośrednio powiązany przynajmniej z *g*-*k* innymi jej członkami. Tym samym 1-plex jest grupą, w której każdy

<sup>35</sup> Jest to szczególnie ważna definicja, gdy istotne są bezpośrednie kontakty.

<sup>36</sup> W znaczeniu największy możliwy.

ma relacje z każdym. Natomiast definicja *k*-cores wymaga, by każdy węzeł w podgrafie był połączony z przynajmniej *k* innymi węzłami z tego podgrafu. Zamiast więc nakładać wymagania na liczbę krawędzi, które mogą być nieobecne, wymaga się danej liczby istniejących relacji. 3-cores to na rysunku 3 wyłącznie grupy {1, 2, 3, 4, 5} oraz {10, 11, 12, 13, 14}.

Jeszcze inne podejście do wyodrębniania spójnych podsieci możliwe jest, gdy w analizowanej sieci węzły połączone są relacjami o zróżnicowanej sile. W takich przypadkach możliwe jest stosowanie definicji maksymalnie silnych komponentów (często nazywanych również *m*-slices). Jest to maksymalny podgraf, dla którego wszystkie węzły połączone są relacjami o sile *m* lub większej.

### Pozycje i role w ramach sieci

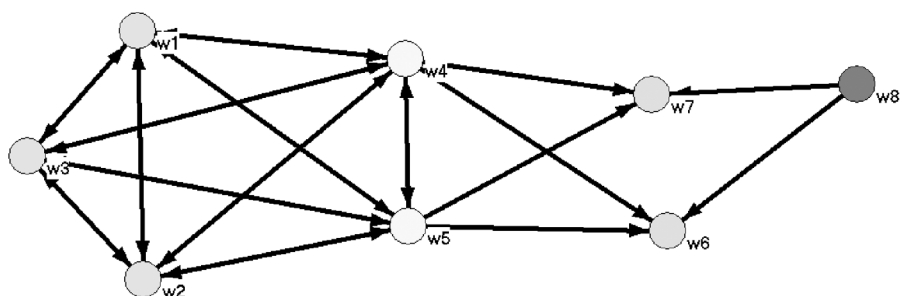
Innym, niż przed chwilą opisane, podejściem do wyodrębniania podgrup w ramach sieci jest wyróżnianie pozycji i ról społecznych. Jednostki zajmujące taką samą pozycję w sieci są w ten sam lub podobny sposób połączone z całym systemem społecznym, ale nie muszą mieć relacji między sobą. Ekwiwalenty strukturalne są to podgrupy w obrębie sieci, zdefiniowane przez wzór relacji łączących aktorów między sobą. Dwa węzły są strukturalnie ekwiwalentne, jeżeli mają dokładnie takie same relacje ze wszystkimi innymi węzłami w sieci. Jest to tożsame z tym, że sieci, z których usunęlibyśmy jeden z tych węzłów i jego relacje, byłyby takie same.

Rysunek 4 przedstawia cztery grupy strukturalnie ekwiwalentnych jednostek. Warto zwrócić uwagę, że choć węzły od 1 do 5 tworzą klikę, to jednak można tu wyróżnić dwie pozycje, jako że węzły 4 i 5 są dodatkowo powiązane z węzłami 6 i 7. Te dwa ostatnie również mają taką samą pozycję mimo tego, iż nie są powiązane ze sobą.

W praktyce, w większych sieciach takie sytuacje zdarzają się stosunkowo rzadko. Dlatego też definiuje się pewną miarę podobieństwa wzorców relacji dla każdej pary węzłów. Może to być na przykład liczba lub odsetek jednostek, z którymi dane dwa węzły mają różne relacje. Dysponując danymi o macierzy odległości dla wszystkich par węzłów, stosuje się metody hierarchicznego *clusteringu* i dzieli się sieć na podgrupy.

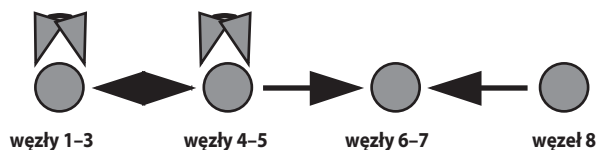
Nieco inną metodą, w ostatnich latach rzadziej stosowaną, jest metoda CONCOR<sup>37</sup>. Umożliwia ona podzielenie sieci na dwie podgrupy pozycji, z których każda może być dalej dzielona. Metoda polega na tworzeniu macierzy korelacji z macierzy relacji pomiędzy jednostkami. Wielokrotne powtarzanie tej operacji powoduje, że w macierzy zostają wyłącznie wartości +1 między węzłami, które są w tej samej grupie i wartości -1 dla relacji międzygrupowych.

<sup>37</sup> Knoke, D., Kuklinski J.H., op. cit.



Rysunek 4. Pozycje w sieci – cztery grupy strukturalnie ekwiwalentnych jednostek (relacje skierowane)

Wyodrębnienie podgrup w ramach sieci nie jest celem samym w sobie. Można zredukować sieć między jednostkami i badać zależności pomiędzy wyodrębnionymi w sposób rozłączny grupami. Przykład takiej redukcji pokazuje rysunek 5, na którym pokazane zostały relacje pomiędzy pozycjami z poprzedniego przykładu. Grupa pierwsza to węzły 1–3, które mają relacje między sobą oraz z węzłami 4 i 5. Węzły 6 i 7 nie inicjują żadnych relacji, ale otrzymują relacje od węzłów 4, 5 i 8. Oprócz redukcji całej sieci możliwe jest analizowanie kontekstu funkcjonowania poszczególnych podgrup poprzez redukcję pozostałych podgrup, ale bez redukowania danej grupy. Metody takie są stosowane szczególnie w analizach dużych sieci<sup>38</sup>.



Rysunek 5. Relacje pomiędzy pozycjami zidentyfikowanymi na rysunku 4

W przykładzie badań sieci współpracy pomiędzy organizacjami możliwe byłoby wyodrębnienie za pomocą tych metod grup organizacji, które pełnią takie same funkcje w projekcie i ich role są do pewnego stopnia wymienne. Możliwa byłaby identyfikacja grup jednostek bardziej i mniej zaangażowanych w realizację danego programu, co może być także pierwszym krokiem do odtwarzania nieformalnych hierarchii. Wspomniane tu metody mogłyby również posłużyć do próby wyodrębnienia ról/typów zaangażowania różnych organizacji – na przykład takich, które podejmują konkretne działania i tych, które finansują działania innych, nie angażując się bezpośrednio.

<sup>38</sup> de Nooy W., Batagelj V., Mrvar A., op. cit.

## Własności całej sieci

O wielu własnościach całych sieci wspomniałem już przy innych poziomach analizy. Jednym z podstawowych pytań jest to, czy sieć jest połączona, a jeśli nie, to z ilu i jakich komponentów się składa? Jeśli sieć składa się tylko z jednego komponentu, to ważne jest, jaka jest jej średnica, a także to, jakie są przeciętne odległości pomiędzy węzłami.

Kolejną, prostą, a zarazem istotną, charakterystyką sieci jest gęstość relacji. Gęstość sieci można obliczyć na podstawie liczby osób i liczby relacji. Stanowi ona proporcję istniejących relacji do wszystkich możliwych (liczbę możliwych relacji można obliczyć na podstawie ilości osób  $N$  i wynosi ona  $N*(N-1)$  dla sieci skierowanych). Gęstość sieci opisuje, na ile sieć jest spójna – jej wartość jest duża w sytuacji, gdy wszystkie węzły tworzą jedną grupę, w której węzły są mocno powiązane. Z drugiej strony wartość gęstości jest niewielka, gdy węzły są słabo ze sobą powiązane lub gdy sieć jest podzielona na wiele grup, które są mocno wewnątrz powiązane, ale słabo między sobą. Inną miarą, uzupełniającą informacje dostarczane przez wartość gęstości dla sieci, w których relacje mają określony kierunek, jest spójność sieci (*network cohesion*). Jest to proporcją relacji odwzajemnionych, czyli liczba wzajemnych wyborów, podzielona przez maksymalną liczbę takich wyborów.

Całą sieć można również charakteryzować przez jej centralizację, liczoną jako różnicę pomiędzy centralnością najbardziej centralnych jednostek i wszystkich pozostałych. Różnice te są następnie odnoszone do maksymalnej możliwej sumy różnic dla sieci o takiej samej liczbie węzłów. Miara ta pokazuje, na ile sieć jest zorganizowana wokół najbardziej centralnych węzłów, a na ile ma bardziej rozproszoną strukturę.

Istotną charakterystyką sieci może być także rozkład poszczególnych cech węzłów, w najprostszym przypadku rozkład liczby relacji. Jak się okazuje, w bardzo wielu sieciach, rozkład ten można opisać zależnością potęgową, to znaczy jest bardzo dużo jednostek, które mają bardzo mało relacji i bardzo niewiele takich, które relacji mają bardzo dużo<sup>39</sup>.

Poszczególne miary sieci mogą mieć znaczenie przy porównywaniu struktury różnych sieci między sobą. Jednak wiele charakterystyk sieci jest silnie zależnych od liczby jednostek wchodzących w jej skład, dlatego też porównywanie sieci bardzo różniących się wielkością nie jest wskazane.

## Zależność cech jednostek i relacji w sieci

Opisane powyżej metody analizy sieci koncentrowały się na samej strukturze sieci, wykorzystując prawie wyłącznie informacje o relacjach pomiędzy jednostkami. Jednak w wielu badaniach największy sens ma uwzględnienie

<sup>39</sup> Barabási A.-L., op. cit.

zarówno relacji, jak i atrybutów węzłów. Jedynym do tej pory wspomnianym wykorzystaniem informacji o charakterystykach węzłów była możliwość wyodrębniania podgrup w ramach sieci, będąca podstawą do redukcji większych sieci do sieci relacji pomiędzy grupami. W praktyce możliwości łączenia różnych typów danych jest znacznie więcej.

W dużym uproszczeniu możemy wyróżnić w tym przypadku dwa pytania. Po pierwsze, na ile relacje w sieci zależą od atrybutów jednostek? – jest to pytanie o zjawisko selekcji społecznej. Po drugie, na ile atrybuty jednostek zależą od relacji między nimi? Przykładowo, czy węzły, które są ze sobą powiązane w sieci, mają podobne właściwości? Jest to pytanie o wpływ społeczny. Stosunkowo często występujące w sieciach zjawisko homofilii – podobieństwa węzłów będących ze sobą w relacjach, może być efektem zarówno procesu selekcji, jak i wpływu społecznego. Tego rodzaju zagadnienia były badane między innymi na przykładzie grup młodzieżowych i zachowań, takich jak palenie papierosów, branie narkotyków, picie alkoholu. W przywoływanym już przykładzie warto byłoby weryfikować chociażby to, jak współpraca pomiędzy organizacjami zależy od ich typu – czy przedsiębiorstwa, jednostki administracji i organizacje pozarządowe częściej współpracują z takimi samymi jednostkami, albo na ile trwale grupy współpracujących podmiotów składają się z różnych typów organizacji?

Nieco innym przypadkiem wykorzystania danych o atrybutach jednostek są badania kapitału społecznego, gdzie jednym z istotnych elementów jest dostęp do zasobów poprzez relacje społeczne. Tego typu analizy koncentrują się na tym, do jakich zasobów jednostka może skutecznie dotrzeć, mobilizując własną sieć społeczną.

Badanie efektów sieciowych jest do pewnego stopnia możliwe poprzez użycie zmiennych sieciowych, na przykład opisujących pozycję jednostki, jako zmiennych zależnych w tradycyjnych analizach statystycznych. Jednak często analiza tego typu zależności jest bardzo trudna. Podstawowym problemem jest to, że tradycyjne metody statystyczne wymagają, by jednostki w analizowanej populacji były od siebie niezależne. W przypadku analizy sieci jest to możliwe wyłącznie w badaniach sieci egocentrycznych prowadzonych na próbie. W konsekwencji w ostatnich latach rozwiniętych zostało wiele modeli umożliwiających testowanie hipotez dla danych sieciowych – tak zwane exponential random graphs models ERGM<sup>40</sup>. A także metod analiz efektów wpływu społecznego i selekcji społecznej<sup>41</sup>.

---

<sup>40</sup> Snijders T.A.B., Pattison P.E., Robins G.L., Handcock M.S., op. cit.; Robins G.L., Snijders T.A.B., Wang P., Handcock M., Pattison P., „Recent developments in exponential random graph (p\*) models for social networks”, *Social Networks* 2007, nr 29, s. 192–215.

<sup>41</sup> Snijders T.A.B., op. cit.



## SNA w ewaluacji

Jak zauważają Durland i Fredericks, stosunkowo niedawno zaczęto używać metod analizy sieci w badaniach ewaluacyjnych i nie są one jeszcze w pełni wykorzystywane<sup>42</sup>. Może to wynikać z tego iż metod analizy sieci społecznych i różnego rodzaju miar sieciowych jest bardzo dużo, ich wykorzystanie wymaga umiejętności korzystania ze specjalnych programów do analizy sieci<sup>43</sup> i w konsekwencji prowadzenie badań sieciowych przez początkujących badaczy jest bardzo trudne<sup>44</sup>.

Obecnie metody analizy sieci społecznych są używane przez ewaluatorów w różnych krajach. Jednym z przykładów jest zastosowanie analiz sieciowych w programach zdrowotnych i epidemiologicznych, szczególnie mających na celu przeciwdziałanie rozpowszechnianiu się takich chorób, jak HIV/AIDS, SARS. Ale również w promocji zdrowia<sup>45</sup>, przeciwdziałaniu alkoholizmowi, narkomanii, zwalczaniu palenia tytoniu i otyłości. Podejście sieciowe jest w tych przypadkach używane jako podstawa do konstruowania konkretnych programów. Badania pokazują między innymi, że zachowania te mają charakter sieciowy, osoby otyłe mają więcej znajomych wśród osób otyłych, palące wśród palących itd., w związku z tym skuteczne przeciwdziałanie i zwalczanie nałogów powinno uwzględniać te mechanizmy relacyjne. Niedawne badania Christakisa i Fowlera pokazały, że również rzucanie palenia ma charakter grupowy, tym samym kampanie antynikotynowe powinny być kierowane nie do pojedynczych jednostek, ale do grup osób znających się i palących w swoim towarzystwie<sup>46</sup>. Prowadzenie badań ewaluacyjnych tego typu programów powinno opierać się na analizie sieciowej.

Metody analizy sieci były także wykorzystywane w ewaluacji programów szkolnych mających na celu poprawę funkcjonowania szkoły poprzez zwiększenie współpracy pomiędzy nauczycielami<sup>47</sup>. Ogólnie rzecz biorąc o zastosowaniu analiz sieci w ewaluacji można myśleć wszędzie tam, gdzie mamy do czynienia z danymi o charakterze relacyjnym.

---

<sup>42</sup> Durland M.M., Fredericks K.A., „Social Network Analysis in Program Evaluation”, *New Directions for Evaluation* 2005, nr 107.

<sup>43</sup> Specyfika metod sieciowych powoduje, że większość popularnych programów komputerowych służących do analiz statystycznych nie pozwala na analizowanie danych metodami sieciowymi. Jest jednak wiele programów, również darmowych, dostępnych w Internecie, które pozwalają na wizualizowanie i analizowanie danych sieciowych. Do najpopularniejszych należą UCINET i program R (w szczególności pakiety SNA, Network i Statnet), oraz Pajek, NetDraw do wizualizacji, a także SIENA, pnet do bardziej zaawansowanych analiz.

<sup>44</sup> Davies R.J., *SNA tools: their potential and limits*, 2007.

<sup>45</sup> Valente T.W., *Evaluating Health Promotion Programs*, Oxford University Press 2002.

<sup>46</sup> Christakis N.A., Fowler J.H., „The Collective Dynamics of Smoking in a Large Social Network”, *New England Journal of Medicine* 2008, nr 358(21), s. 2249–2258.

<sup>47</sup> Penuel W.R., Sussex W., Korbak Ch., Hoadley Ch., „Investigating the Potential of Using Social Network Analysis in Educational Evaluation”, *American Journal of Evaluation* 2006, vol. 27(4), s. 437–451.

Jednak analiza sieci może być wykorzystywana nie tylko do ewaluacji efektów oddziaływań różnych programów, ale także do ewaluacji samych programów i ich organizacji. W szczególności większe programy mogą być konceptualizowane jako sieci<sup>48</sup>. Praktycznie, każdy projekt, a tym bardziej większy program, są zaangażowane rozmaite osoby, często z różnych organizacji. Relacje między tymi osobami, ich zaangażowanie na poszczególnych poziomach czy w różne części programu, mogą być opisywane za pomocą sieci afiliacji. Wiedza wynikająca z analiz sieciowych może przyczynić się do usprawnienia analizowanego programu.

Przykładem wykorzystania analizy sieci do analizy informacji o bardzo dużym programie jest zastosowanie tych metod przez Polską Agencję Rozwoju Przedsiębiorczości w analizie działania 2.3 Sektorowego Programu Operacyjnego Rozwój Zasobów Ludzkich (SPO RZL). W ramach programu „SPO RZL 2.3 Rozwój kadr nowoczesnej gospodarki” ponad 400 tys. osób (z ponad 100 tys. firm) wzięło udział w około 700 programach szkoleniowych. Na podstawie danych zbieranych w trakcie realizacji programu możliwe jest analizowanie relacji uczestnictwa osób i firm w poszczególnych projektach szkoleniowych. Pozwala to między innymi na badanie relacji pomiędzy szkoleniami (poprzez współuczestnictwo) i na wyciąganie wniosków dotyczących najczęściej występujących kombinacji szkoleń, tematycznego uzupełniania się projektów oraz ich substytucyjności.

Podsumowując, analiza sieciowa może być przydatna na różnych poziomach ewaluacji. Metody te pozwalają analizować nie tylko dane o atrybutach jednostek, ale również o powiązaniach między nimi. Jest to też podejście do badań społecznych, w których jednym z głównych źródeł wyjaśnień działań i sytuacji jednostek są relacje i struktury wzajemnych powiązań, w które jednostki te są zaangażowane. Dlatego też w badaniach ewaluacyjnych analiza sieci może być przydatna zarówno w ocenie efektów podejmowanych działań, których istotnym elementem są relacje, np. wspomnianych wcześniej programów walki z nałogami, jak i przy analizie sposobu organizacji działań i projektów oraz znaczeniu tej organizacji dla efektywności tychże. Warto więc, myśląc o prowadzeniu badań ewaluacyjnych, rozważyć przydatność zastosowania podejścia sieciowego i zbierać dane o charakterze relacyjnym. Mimo trudności, związanych z wielością i skomplikowaniem metod sieciowych i obsługą programów do analizy danych sieciowych, napotykanym przez oso-

---

<sup>48</sup> Davies R.J., „Network Perspectives on the Evaluation of Development Interventions”, Paper for the EDAIS Conference November 24–25, *New Directions in Impact Assessment for Development: Methods and Practice* 2003; Davies R.J., „Scale, Complexity and the Representation of Theories of Change”, *Evaluation* 2004, Vol 10(1), s. 101–121, Sage Publications, London; Davies R.J., „Scale, Complexity and the Representation of Theories of Change Part II”. *Evaluation*, Vol 11(2):133–149, Sage Publications, London 2005.

by, które dopiero zaczynają zajmować się tego typu badaniami, należy mieć nadzieję, że podejście to zacznie być szerzej stosowane również w badaniach ewaluacyjnych.

---

**Dominik Batorski** – doktor socjologii. Obecnie pracuje jako adiunkt w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego na Uniwersytecie Warszawskim. Naukowo zajmuje się przede wszystkim: problematyką relacji i sieci społecznych oraz dynamiką procesów społecznych zachodzących w sieciach, takich jak dyfuzja innowacji czy wpływ społeczny. Pracuje również nad metodami analizy sieci społecznych i metodologią prowadzenia badań psychologicznych i społecznych przez Internet. Jest członkiem Rady Monitoringu Społecznego, przygotowującej cykl badań Diagnoza Społeczna. W badaniach tych odpowiada za moduł poświęcony uwarunkowaniom, sposobom i konsekwencjom korzystania z nowych technologii. W latach 2006–2007 był stypendystą Fundacji na rzecz Nauki Polskiej. Należy do kilku stowarzyszeń naukowych, między innymi International Network for Social Networks Analysis i Polskiego Towarzystwa Socjologicznego.

## Bibliografia

- Barabási A.-L., *Linked: how everything is connected to everything else and what it means for business, science and everyday life*, A Plume Book 2003.
- Burt R.S., *Structural holes: the social structure of competition*, Harvard University Press 1992.
- Burt R.S., *Brokerage and Closure: An introduction to Social Capital*, Oxford University Press 2005.
- Casciaro T., „Seeing things clearly: social structure, personality, and accuracy in social network perception”, *Social Networks* 1998, 20(4).
- Christakis N.A., Fowler J.H., „The Collective Dynamics of Smoking in a Large Social Network”, *New England Journal of Medicine* 2008, nr 358(21).
- Cummings J.N., Cross R., „Structural properties of work groups and their consequences for performance”, *Social Networks* 2003, 25(3).
- Davies R.J., „Network Perspectives on the Evaluation of Development Interventions”, Paper for the ED AIS Conference November 24–25, *New Directions in Impact Assessment for Development: Methods and Practice* 2003.
- Davies R.J., „Scale, Complexity and the Representation of Theories of Change”, *Evaluation* 2004, Vol 10(1), Sage Publications, London.
- Davies R.J., „Scale, Complexity and the Representation of Theories of Change Part II”. *Evaluation*, Vol 11(2), Sage Publications, London 2005.
- Davies R.J., *SNA tools: their potential and limits*, 2007.
- Durland M.M., Fredericks K.A., „Social Network Analysis in Program Evaluation”, *New Directions for Evaluation* 2005, nr 107.
- Durland M.M., Kimberly A.F., „Social Network Analysis in Program Evaluation”, *New Directions for Evaluation* 2006, nr 107, Jossey-Bass.
- Erickson B.H., Nosanchuk T.A., „Applied network sampling”, *Social Networks* 1983, nr 5.

- Frank O., „Network sampling and model fitting”, w: P.J. Carrington, J. Scott, S. Wasserman (red.), *Models and methods in social networks analysis*, Cambridge University Press 2005.
- Freeman L.C., „Centrality in social networks: I. conceptual clarification”, *Social Networks* 1979, nr 1.
- Granovetter M., „The strength of weak ties”, *American Journal of Sociology* 1973, nr 73.
- Hanneman R.A., Riddle M., *Introduction to social network methods*, University of California, Riverside 2005. Opublikowane elektronicznie na: <http://faculty.ucr.edu/~hanneman/>.
- Heider F., „On balance and attribution”, w: P.W. Holland, S. Leinhardt (red.), *Perspectives on social network research*, Academic Press 1979.
- Kilduff M., Tsai W., *Social Networks and Organizations*, Sage, Londyn 2003.
- Knoke, D., Kuklinski J.H., *Network analysis*, Sage 1982.
- Krackhardt D., Kilduff M., „Whether close or far: Social distance effects on perceived balance in friendship networks”, *Journal of Personality and Social Psychology* 1999, nr 76.
- Lin N., Cook K., Burt R.S. (red.), *Social capital: theory and research* 2001.
- Lin N., Fu Y., Hsung R.-M., „The position generator: measurement techniques for investigations of social capital”, w: N. Lin, K. Cook, R.S. Burt (red.), *Social capital: theory and research*, 2001.
- Marsden P.V., „Network data and measurement”, *Annual Review of Sociology* 1990, nr 16.
- Monge P.R., Contractor N.S., *Theories of communication networks*, Oxford University Press, New York 2003.
- Moreno J.L., „Foundations of sociometry: An introduction”, *Sociometry* 1941, nr 4(1).
- Newman M.E.J., „The structure and function of complex networks”, *SIAM Review* 2003, nr 45.
- de Nooy W., Batagelj V., Mrvar A., *Exploratory Social Network Analysis with Pajek*, Cambridge University Press 2005.
- Penuel W.R., Sussex W., Korbak Ch., Hoadley Ch., „Investigating the Potential of Using Social Network Analysis in Educational Evaluation”, *American Journal of Evaluation* 2006, vol. 27(4).
- Robins G., Pattison P., „Random graph models for temporal processes in social networks”, *Journal of Mathematical Sociology* 2000, nr 25.
- Robins G., Pattison P., Elliott P., „Network models for social influence processes”, *Psychometrika* 2001, nr 66.
- Robins G.L., Snijders T.A.B., Wang P., Handcock M., Pattison P., „Recent developments in exponential random graph ( $p^*$ ) models for social networks”, *Social Networks* 2007, nr 295.
- Scott J., *Social Network Analysis: a handbook*, Sage Publications, Londyn 2000.
- Snijders T.A.B., „Models for Longitudinal Network Data”, w: P. Carrington, J. Scott, S. Wasserman (red.), *Models and methods in social network analysis*, Cambridge University Press, New York 2005, rozdział 11.
- Snijders T.A.B., Pattison P.E., Robins G.L., Handcock M.S., „New specifications for exponential random graph models”, *Sociological Methodology* 2006.
- Valente T.W., *Evaluating Health Promotion Programs*, Oxford University Press 2002.
- Wasserman S., Faust K., *Social Networks Analysis: Methods and applications. Structural analysis in the social sciences*, Cambridge University Press 1994.
- Watts D.J., „Networks, dynamics, and the small-world phenomenon”, *American Journal of Sociology* 1999, nr 105(2).
- Watts D.J., *Six degrees: the science of a connected age*, W.W. Norton 2003.
- Wellman B., „Structural analysis: from method and metaphor to theory and substance”, w: *Social structures: a network approach*, B. Wellman i S.D. Berkowitz (red.), Cambridge University Press 1988, tom 15 serii *Contemporary studies in sociology*, rozdział 2.



Paweł Strawiński

## Quasi-eksperymentalne metody ewaluacji

### Problem ewaluacji

Prowadzone w ramach polityki gospodarczej państwa ukierunkowane programy pomocy pochłaniają niemałą część środków budżetowych. Pod pojęciem programu pomocy jest rozumiane każde działanie podjęte dla dobra lub na rzecz innej jednostki w celu usprawnienia jej funkcjonowania. W odniesieniu do gospodarki oznacza to, że polityka gospodarcza, interwencje prowadzone na rynku, czy pomoc bezpośrednio zaadresowana do wyznaczonych beneficjentów są przykładami programów pomocy. Ze względu na istnienie alternatywnych sposobów wykorzystania dostępnych środków ocena skuteczności oraz efektywności programów jest ważnym zagadnieniem i powinna stanowić podstawę do podejmowania strategicznych decyzji.

Ekonomiczne ramy analizy ewaluacyjnej wyznaczają modele Roya i Rubina. Model Roya<sup>1</sup> formalizuje zasady wyboru ekonomicznego w warunkach ograniczonych zasobów. Tłumaczy on dlaczego jednostki nie podejmują decyzji spontanicznie (losowo), a kierują się pewnym kryterium. Z kolei w pracy Rubina<sup>2</sup> pokazano, że prawidłowy dobór grupy odniesienia uzyskanych rezultatów pozwala na otrzymanie lepszych w sensie statystycznym wyników.

Badacz obserwuje efekt działania programu i dokonuje podsumowania jego skuteczności na podstawie oceny pewnych wielkości, na przykład analizując, czy osoba znalazła pracę, lub rozpatrując zmiany osiąganego przez nią dochodu. Ogólnie miernik efektu programu możemy utożsamić ze zmienną losową. Będzie ona nazywana zmienną wynikową.

Problemem występującym przy ocenie efektów stosowanych polityk jest to, że obserwowana jednostka w danym momencie czasu może uczestniczyć w programie bądź w nim nie uczestniczyć. Wartość zmiennej wynikowej jest przez badacza obserwowana i na jej podstawie jest oceniany zastosowany pro-

---

<sup>1</sup> Roy A., „Some Thoughts on the Distribution of Earnings”, *Oxford Economic Papers. New Series*, vol. 3/2, s. 135–146.

<sup>2</sup> Rubin D., „Matching to Remove Bias in Observational Studies”, *Biometrics* 1973a, vol. 29.

gram. Dla potrzeb oceny programu należy dokonać porównania, np. przeciętnego efektu dla uczestników programu, oraz jednostek w nim nieuczestniczących. Z punktu widzenia matematyki i poprawności formalnej przeprowadzonych obliczeń nie ma to większego znaczenia, do której grupy jednostka należy, jednak rozróżnienie tych dwóch stanów jest kluczowym elementem, niezbędnym do przeprowadzenia poprawnej analizy ekonomicznej<sup>3</sup>. Ocena wpływu działania na obserwowany rezultat zawiera spekulację, w jaki sposób jednostka by funkcjonowała, gdyby nie została poddana oddziaływaniu<sup>4</sup>.

W dalszej części zostanie przedstawione jedno z możliwych rozwiązań problemu, w jaki sposób dokonać ewaluacji działania przy braku bezpośrednio dostępnej grupy porównawczej. Następnie omówimy ramy przyczynowo-skutkowe analizy oraz interesujące badaczy efekty oddziaływania. Dokonamy rozróżnienia pomiędzy przeciętnym efektem oddziaływania (ang. *Average Treatment Effect* – ATE) i przeciętnym efektem oddziaływania na jednostki poddane oddziaływaniu (*Average Treatment on Treated* – ATT). Zaprezentujemy statystyczne oraz ekonometryczne metody łączenia danych (ang. *matching methods*) jako sposób konstrukcji sztucznej grupy porównawczej. Przyjrzymy się technice *propensity score matching*, która jest jednym ze sposobów pozwalającym na przekształcenie danych o charakterze nieeksperymentalnym w dane quasi-eksperymentalne. Przedstawimy różne podejścia i narzędzia, które pozwalają określić stopień skuteczności metody łączenia wykorzystującej wektory *propensity score*. W ostatniej części znajdzie się podsumowanie kluczowych elementów metody *propensity score* oraz krytyka tego podejścia badawczego.

Najważniejsze elementy metody łączenia danych zostaną zilustrowane za pomocą sztucznie wygenerowanych danych przykładowych. Intencją autora jest stworzenie takiego zbioru, na podstawie którego można pokazać podstawowe idee metody łączenia, a zarazem dopasowanego nieidealnie do metody łączenia wykorzystującej wektory *propensity score*.

## Model przyczynowy

W analizie ekonomicznej działania programu jako jego efekt obserwowana jest dla każdej jednostki  $i$  wartość zmiennej wynikowej. Wynosi ona  $Y_{1i}$ , jeżeli jednostka uczestniczy w programie, lub  $Y_{0i}$ , jeżeli w nim nie uczestniczy. Posługując się terminologią nauk eksperymentalnych, w pierwszym przypadku będziemy mówić, że jednostka należy do grupy eksperymentalnej

<sup>3</sup> Strawiński P., „Przyczynowość, selekcja i endogeniczne oddziaływanie”, *Przegląd Statystyczny* 2007, nr 4, s. 49–61.

<sup>4</sup> Caliendo M., Kopeinig S., „Some Practical Guidance for the Implementation of Propensity Score Matching”, *Journal of Economic Surveys* 2008.

lub poddanej oddziaływaniu. W drugim, że należy do grupy kontrolnej lub grupy odniesienia. Aby porównać te dwie wielkości, niezbędne jest uwzględnienie w budowanym modelu zjawiska informacji o stanie uczestnictwa badanej jednostki w programie.

Trudność oceny programów jest problemem brakujących danych<sup>5</sup>. W danej chwili rozpatrywana jednostka może znajdować się w jednym z dwóch alternatywnych stanów, ale nie może być w obu jednocześnie. Może być uczestnikiem programu, bądź w nim nie uczestniczyć. Z każdym ze stanów wiąże się jeden z poziomów obserwowanej zmiennej wynikowej ( $Y_{1i}, Y_{0i}$ ). Jeżeli można by było jednocześnie obserwować obie wartości dla każdej poddanej badaniu jednostki, zniknąłby problem związany z oceną programu.

W wielu pracach empirycznych dokonuje się oceny efektów programu pomocy przez pomiar przeciętnego efektu zastosowania pewnej polityki. Niech zmienna  $P$  opisuje stan uczestnictwa w programie i przyjmuje wartość  $P_i=1$ , gdy jednostka  $i$  jest uczestnikiem programu (należy do grupy eksperymentalnej), a  $P_i=0$  oznacza, że znajduje się poza programem. Wówczas przeciętny efekt działania programu (ang. *average treatment effect*) można zapisać jako:

$$Y_{ATE} = \frac{1}{N} \left[ \sum_{i=1}^N (P_i Y_{1i} + (1 - P_i) Y_{0i}) \right] \quad (1)$$

Efekt programu obliczony w sposób opisany przez wzór (1) jest efektem dla przeciętnej jednostki pochodzącej z badanej populacji, bez uwzględnienia, czy jest ona poddana oddziaływaniu, czy nie. W praktyce jako wielkość efektu przyjmuje się różnicę w średnim wyniku dla grupy eksperymentalnej i wszystkich jednostek nieuczestniczących w eksperymencie. Nie gwarantuje to dokładności, ponieważ jednostki z grupy eksperymentalnej mogą w sposób systematyczny odróżniać się od pozostałych, co przy pomiarze nie jest uwzględnione.

Dużo ciekawszą miarą, niosącą więcej informacji o skuteczności działania, jest przeciętny efekt oddziaływania programu na jednostki poddane oddziaływaniu (ang. *average treatment on treated*). Jest to efekt ograniczony do grupy, na którą oddziaływano.

$$Y_{ATT} = E(Y_{1i} - Y_{0i} | P_i = 1) = E(Y_{1i} | P_i = 1) - E(Y_{0i} | P_i = 1) \quad (2)$$

Pokazuje on ilościowy efekt oddziaływania programu wobec jednostek w nim uczestniczących. Jest równy przeciętnej zmianie wartości zmiennej wynikowej wskutek wystąpienia oddziaływania. W kontekście badań programów pomocy informuje ona, jaki jest zysk bądź strata z faktu uczestniczenia w programie dla jego beneficjentów. Należy zaznaczyć, że wielkość

<sup>5</sup> Heckman J., Ichimura H., Todd P., „Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *The Review of Economic Studies* 1997, vol. 64/4, s. 605–654.



$E(Y_{1i} | P_i = 1)$  jest oczekiwaną wartością zmiennej wynikowej dla beneficjentów programu i jest wartością obserwowaną, natomiast  $E(Y_{0i} | P_i = 1)$  jest wielkością hipotetycznego efektu, gdyby beneficjenci programu pozostawali poza jego oddziaływaniem. Ta druga wartość nie jest obserwowana w rzeczywistości. W literaturze polskiej jest określana mianem stanu kontrfaktycznego<sup>6</sup> (ang. *counterfactual*).

Przyjmijmy na chwilę założenie, że istnieją dwa identyczne pod względem wszystkich cech równoległe światy, w których może być prowadzony program wsparcia drobnych przedsiębiorców. W świecie 0 program nie jest prowadzony, przeciętny zysk firm sektora MSP wynosi  $Y_{i0}$ . Z kolei w świecie 1 każda firma uczestniczy w programie i osiąga zysk w wysokości  $Y_{i1}$ . Ponieważ dwa światy są identyczne pod względem wszystkich pozostałych cech, stan świata 0 może służyć jako stan kontrfaktyczny dla świata 1 i na odwrót. W takim przypadku przeciętny efekt oddziaływania programu na jednostki poddane oddziaływaniu jest równy  $Y_{i1} - Y_{i0}$ . Jediną różnicą pomiędzy światami (stanami) jest fakt, że w jednym z nich jest prowadzony program. Ewaluator o tym wie i może tę informację wykorzystać. Niestety, w rzeczywistości równoległe światy nie istnieją, istnieje jeden będący ich mieszkanką.

Rodzajem badania, które dzięki specyficznej konstrukcji pozwala uniknąć problemów metodologicznych związanych z analizą danych przekrojowych, jest kontrolowany eksperyment. Jest to często stosowane narzędzie badawcze w naukach eksperymentalnych, takich jak biologia, medycyna czy fizyka doświadczalna. Kontrolowany eksperyment polega na zaplanowaniu badania w taki sposób, aby na podstawie uzyskanych wyników można było wyciągnąć prawidłowe wnioski odnoszące się do całej populacji. Dane eksperymentalne pozwalają na skonstruowanie grup porównawczych. W eksperymencie jednostki podlegają randomizacji i przez to są całkowicie arbitralnie lub losowo przydzielane do grupy eksperymentalnej czy grupy kontrolnej. Dzięki losowości wynik oddziaływania jest całkowicie niezależny od stanu oddziaływania i nie występuje zjawisko selekcji. Ponadto, prawidłowo zaprojektowany eksperyment w klarowny sposób rozróżnia czynniki na endogeniczne, czyli takie, na które wpływa rezultat eksperymentu, i egzogeniczne, czyli takie, których wartość nie jest powiązana z jego rezultatem<sup>7</sup>. Przy założeniu braku możliwości przepływu jednostek między grupami, zbiór jednostek poddanych oddziaływaniu jest statystycznie równoważny pod każdym względem zbiorowi niepoddanemu oddziaływaniu<sup>8</sup>. W takim przypadku przeciętny

<sup>6</sup> Konarski R., Kotnarowski M., „Zastosowanie Metody Propensity Score Matching w ewaluacji ex-post”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka*, PARP, Warszawa 2007.

<sup>7</sup> Meyer B., „Natural and Quasi-Experiments in Economics”, *Journal of Business and Economic Statistics* 1995, vol. 13/2, s. 151–161.

<sup>8</sup> Blundell R., Costa-Diás M., „Evaluation Methods for Non-Experimental Data”, *Fiscal Studies* 2000, vol. 21/4, s. 427–468.

efekt oddziaływania jest równy efektowi oddziaływania na jednostki poddane oddziaływaniu i wynosi

$$Y_{ATE} = Y_{ATT} = \bar{Y}_1 - \bar{Y}_0 \quad (3)$$

W praktyce występuje szereg czynników utrudniających przeprowadzenie tego typu eksperymentów i zaburzających ich wyniki. Po pierwsze, z reguły eksperymenty społeczne są dość czasochłonne. W związku z tym są narażone na problem odpadania jednostek na skutek rezygnacji. Trudno jest oczekiwać, że ten problem dokładnie w takim samym stopniu będzie dotyczył grupy eksperymentalnej poddanej oddziaływaniu i grupy kontrolnej. Po drugie, prowadząc kontrolowany eksperyment, badacz nie jest w stanie odizolować grupy uczestników od jej otoczenia ekonomicznego. Kontakty ze światem zewnętrznym mogą wpływać na decyzje jednostek zaangażowanych w programie i jednostek z grupy kontrolnej. Tutaj również założenie o jednakowym wpływie czynników zewnętrznych na obie subpopulacje wydaje się być mało realistyczne.

Schematu badań obowiązującego w naukach eksperymentalnych nie można w prosty sposób przenieść na grunt nauk społecznych. Dobór uczestników do eksperymentalnego programu pomocy jest zawsze problematyczny i pozostaje kontrowersyjną decyzją. Może on naruszyć zasady zachowania równowagi ekonomicznej i konkurencyjności. Jednostki, które zostaną wybrane, mogą dzięki temu odnieść korzyść i uzyskać przewagę ekonomiczną nad pozostałą częścią populacji. Eksperymentowanie na organizmie społecznym prowadzi ewentualnie do powstania czy wzmocnienia nierówności, co przeczy idei pomocy. Ze względów etycznych, społecznych i politycznych zastosowanie reguł eksperymentu medycznego do badania oddziaływania czynników ekonomicznych wydaje się niemożliwe.

Poważnym zagadnieniem w badaniach społecznych, oceniających skuteczność różnych programów pomocy, jest konstrukcja grupy porównawczej. Zazwyczaj nie jest ona dostępna wprost, z wyjątkiem spontanicznie powstających naturalnych eksperymentów, więc badacze nie mają możliwości odniesienia rezultatów działania programu z powodu braku grupy referencyjnej. Jednym z rozwiązań problemu jest wykorzystanie metod quasi-eksperymentalnych, które pozwalają przekształcić dane zebrane podczas badania przekrojowego w równoważnik grupy porównawczej.

## Metody łączenia danych

Zadaniem metody łączenia obserwacji jest przekształcenie zbioru danych w sposób umożliwiający oszacowanie przeciętnego wpływu, jaki wywiera poddanie jednostki działaniu programu. Z reguły do szacowania efektów wykorzystuje się techniki nieparametryczne lub semi-parametryczne. Techniki

parametryczne polegają na tym, że badacz *a priori* zakłada, że zmienna losowa opisująca badane zjawisko zachowuje się tak, jakby pochodziła z arbitralnie wybranego przez badacza rozkładu prawdopodobieństwa. Znaczną liczbę rozkładów prawdopodobieństwa można przedstawić jako funkcję zbioru parametrów. Zadaniem badacza jest znalezienie takich wartości nieznanymi parametrów, aby optymalizować funkcję kryterium dopasowania dostępnej próby empirycznej do wybranego rozkładu.

Z uwagi na konieczność określenia rozkładu prawdopodobieństwa badanego zjawiska metody parametryczne są efektywne jedynie w przypadku wyboru prawidłowego rozkładu. Techniki nieparametryczne znajdowania estymatorów, w większym stopniu niż techniki parametryczne bazują na informacji zawartej w dostępnej badaczowi próbie. Nie uwzględnia się rozkładu zjawiska, zastępując go estymatorem próbkowym. Wadą tego podejścia jest budowanie całego rozkładu na podstawie jednej realizacji, oraz ograniczona możliwość wyprowadzenia wzorów analitycznych.

Heckman, Ichimura i Todd<sup>9</sup> precyzują warunki, jakie muszą spełniać dane empiryczne zbierane w warunkach nieeksperymentalnych oraz metody łączenia danych, aby wypełnić kryteria stawiane przed idealnym badaniem społecznym:

- 1) jednostki poddane oddziaływaniu i jednostki z grupy kontrolnej powinny mieć takie same rozkłady cech nieobserwowanych;
- 2) powinny mieć takie same rozkłady cech obserwowanych;
- 3) dane w obu subpopulacjach powinny być zbierane w ten sam sposób;
- 4) uczestnicy programu i członkowie grupy kontrolnej powinni pochodzić z tego samego środowiska społeczno-ekonomicznego.

Spełnienie kryteriów (3) i (4) nie jest ograniczeniem utrudniającym przeprowadzenie ewaluacji, gdyż każde poprawnie sformułowane badanie powinno je spełniać. Kryterium (2) jest możliwe do spełnienia podczas różnych etapów badania, odpowiednio zaprojektowanego. Alternatywnie, podczas opracowywania wyników mogą zostać użyte zaawansowane techniki ekonometryczne, takie jak przeważenie danych czy budowa funkcji kontrolującej. Problematyczne jest spełnienie kryterium (1). Zakłada się, że jest ono spełnione przy prawdziwym warunku (2).

Metody łączenia obserwacji (ang. *matching methods*) pozwalają na rozwiązanie problemu oceny prowadzonych programów pomocy w inny sposób niż metody parametryczne. Główną przeszkodą utrudniającą zastosowanie metod parametrycznych jest brak adekwatnej grupy porównawczej, wobec której można odnieść uzyskane w badaniu wyniki. Łączenie obserwacji pozwala na skonstruowanie poprawnej grupy porównawczej zastępującej brakujące dane,

---

<sup>9</sup> Heckman J., Ichimura H., Todd P., „Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *The Review of Economic Studies* 1997, vol. 64/4, s. 605–654.

bez konieczności przyjmowania dodatkowych założeń *a priori*<sup>10</sup>. Ich zadaniem jest naśladowanie eksperymentu poprzez taki wybór jednostek do grupy kontrolnej, by była jak najbardziej zbliżona pod względem obserwowanych charakterystyk do grupy poddanej oddziaływaniu.

Metody łączenia koncentrują się na sposobach konstrukcji grupy referencyjnej. Ich istotą jest dobór grupy porównawczej z puli potencjalnych jednostek kontrolnych na podstawie wartości wybranej miary podobieństwa. Są to metody wykorzystujące techniki nieparametryczne w celu transformacji rzeczywistych danych zebranych podczas badania przekrojowego w zbiór danych quasi-eksperymentalnych. Do przekształconych danych można stosować techniki wykorzystywane w ewaluacji danych eksperymentalnych, czyli wystarczy porównać średnią wartość zmiennej mierzącej wynik działania.

Głównym zadaniem metod nieparametrycznych jest maksymalne i efektywne wykorzystanie dostępnej informacji. Poprawna konstrukcja sztucznej grupy odniesienia zapewnia zredukowanie obciążenia i zwiększenie precyzji uzyskanych wyników<sup>11</sup>. Podobnie jak szereg innych elementów teorii ekonometrii, jej korzenie i pierwsze zastosowania tkwią w biologii i naukach medycznych.

### ***Propensity score matching***

Wśród metod łączenia danych można wyróżnić dwa podejścia. Historycznie pierwszą metodą było łączenie według wartości współzmiennych (ang. *covariates matching*). Współzmiennymi nazywa się zmienne używane w procedurze łączenia. Ich nazwa wywodzi się stąd, że różne ich wartości determinują zarówno jeden ze stanów oddziaływania programu, jak i wartość zmiennej mierzącej efekt ilościowy programu. Zatem ich wartość zmienia się wspólnie z wartością zmiennej określającej stan oddziaływania. Przykładowymi współzmiennymi, często wykorzystywanymi przy ewaluacji programów rynku pracy, z zastosowaniem metod mikroekonometrycznych, są cechy beneficjentów. Z reguły selekcja uczestników odbywa się na podstawie takich cech, jak wiek, płeć czy miejsce zamieszkania. Zmienną wynikową mierzącą efekt programu mogą być np. osiągnięte zarobki w pewnym okresie po programie (np. rok po zakończeniu). Z teorii rynku pracy wiadomo, że płeć, wiek czy miejsce zamieszkania osoby są silnymi determinantami wysokości uzyskiwanych zarobków<sup>12</sup>. Zatem te wszystkie cechy w przypadku programu pomocy nadają się na współzmiennie.

<sup>10</sup> Blundell R., Costa-Diás M., op. cit.

<sup>11</sup> Rubin D., „Matching to Remove Bias in Observational Studies”, *Biometrics* 1973, vol. 29, s. 159–183.

<sup>12</sup> Wielkość miejscowości jest standardową zmienną stosowaną w badaniach rynku pracy jako przybliżenie wielkości (pojemności) lokalnego rynku pracy.

Łączenie według wartości współzmiennych polega na łączeniu w pary obserwacji o identycznych lub zbliżonych wartościach współzmiennych. Jeżeli wartości zmiennych zostaną przedstawione jako punkty przestrzeni wielowymiarowej, to statystyczna metoda łączenia polega na podziale tej przestrzeni na klatki według wartości współzmiennych. Następnie łączone są nie obserwacje, a utworzone klatki w pary. Niestety, jeżeli wśród współzmiennych występują zmienne ciągłe lub gdy jest duża liczba współzmiennych, to metoda okazuje się niepraktyczna w zastosowaniu, gdyż duża liczba klatek pozostaje pusta lub są one zbyt duże, co zmniejsza efektywność łączenia i zwiększa wariancję estymatora wielkości oddziaływania. Druga metoda, bardziej współczesna, wykorzystuje wektory *propensity score* albo inną miarę odległości między wektorami.

### Przykład

Istnieje zbiór danych zawierający 150 obserwacji, których cechy są opisane za pomocą kilku zmiennych. Po pierwsze, zmienna wynikowa, którą oznaczmy jako  $Y$ . Na podstawie jej wartości będzie oceniana skuteczność programu. Dodatkowo dostępna jest zmienna  $P$  informująca, czy jednostka została poddana działaniu programu (wtedy  $P = 1$ ), czy znajduje się w puli jednostek, z której można wybrać grupę referencyjną ( $P = 0$ ). Arbitralnie przyjęto, że jednostek uczestniczących w programie jest 50, a jednostek niepoddanych działaniu programu 100. Dodatkowo dla każdej jednostki dysponujemy wartościami ciągłych współzmiennych  $X_1$  oraz  $X_2$ , dyskretnych współzmiennych  $X_3$  oraz  $X_4$  oraz zmiennej ( $X_5$ ), której wartość wpływa na zmienną wynikową  $Y$ , ale nie ma wpływu na stan poddania oddziaływaniu. Współzmiennie  $X_1, X_2, X_3, X_4$  są to zmienne, których wartości jednocześnie mają wpływ na stan poddania oddziaływaniu  $P$ , oraz wartość zmiennej wynikowej  $Y$  (ang. *outcome variable*).

Aby obliczyć przeciętny efekt programu dany przez (1), należy w równaniu  $Y$  zawrzeć informacje zarówno o wielkościach współzmiennych  $X_1, X_2, X_3, X_4$ , jak i o stanie poddania oddziaływaniu  $P$ .

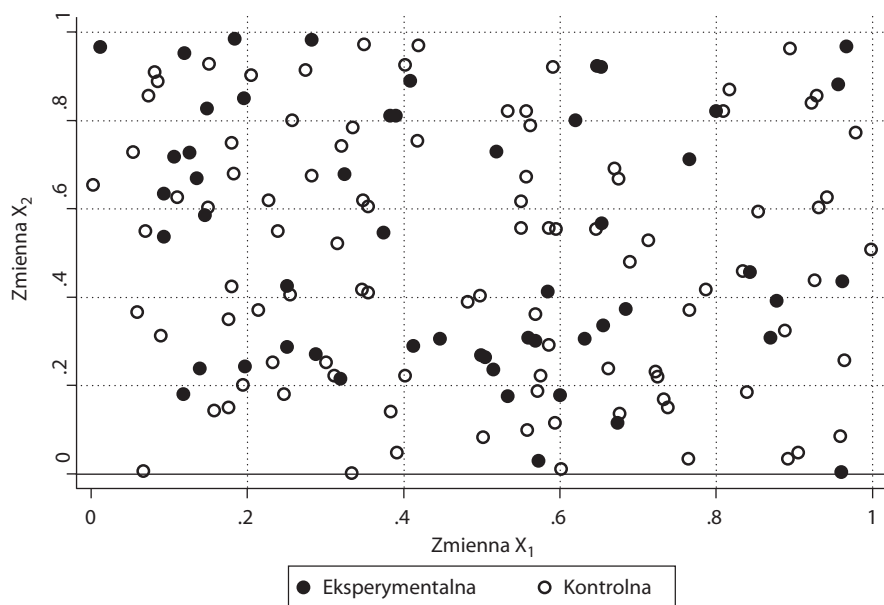
$$Y = f(X_1, X_2, X_3, X_4, P)$$

Jednak spowoduje to problem endogeniczności, ponieważ wartość współzmiennych  $X_1, X_2, X_3, X_4$  będzie jednocześnie wpływać na wynik  $Y$  i na stan poddania oddziaływaniu  $P$ .

Obrazowe przedstawienie punktów przestrzeni wielowymiarowej jest trudne, z tego powodu przykład zostanie ograniczony do dwóch wymiarów. Każdemu z wymiarów odpowiada jedna współzmienna używana do łączenia danych.

W celu ułatwienia prezentacji wszystkie wartości zmiennych znormalizowano do przedziału  $[0; 1]$ . Na rysunku 1 przedstawiono przykładowe wartości dwóch zmiennych ciągłych ( $X_1, X_2$ ). Z każdym z punktów jest

związana pewna wartość zmiennej wynikowej  $Y$ . Łączenie w klatki polega na podziale całej dziedziny na klatki, w przypadku dwóch wymiarów będą to prostokąty. Dokonano równomiernego podziału dziedziny, czyli w pierwszej klatce wyznaczonej przez zmienną  $X_1$  znalazły się obserwacje o wartościach z przedziału  $(0; 0,2)$  itd. Analogicznie utworzono klatki względem wartości zmiennej  $X_2$ .



Rysunek 1. Współzmiennie ciągłe. Podział równomierny na klatki

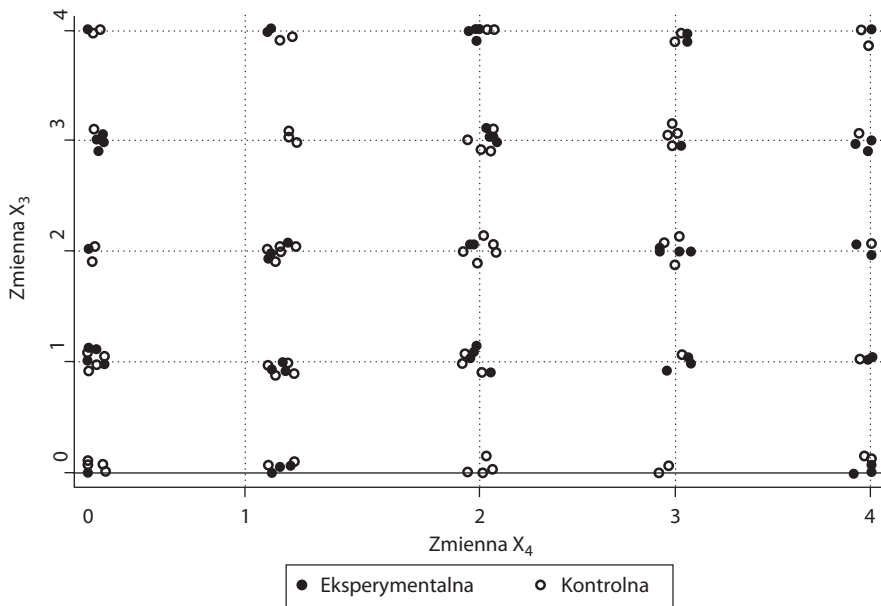
Źródło: Obliczenia własne.

Po połączeniu uzyskano 25 klatek. W klatkach znajdują się jednostki z grupy eksperymentalnej (wypełniony okrąg) i grupy kontrolnej (pusty okrąg). Jednostki znajdujące się w tej samej klatce można uznać za podobne pod względem wartości zmiennych.

W przypadku, gdy zmienne mają charakter dyskretny (zmienne  $X_3$  oraz  $X_4$ ), sytuacja jest dużo prostsza. Dane można połączyć w sposób dokładny (patrz rysunek 2). Na rysunku wprowadzono losowe zaburzenia wartości punktów, by wszystkie były widoczne.

Problem znalezienia elementu porównawczego w bliskim sąsiedztwie jest problemem wielowymiarowym. Metodą pozwalającą na uniknięcie konieczności porównywania wszystkich cech badanych jednostek jest użycie zamiast cech pewnej miary podobieństwa obserwacji. Najpopularniejszą i najszerzej

stosowaną jest wektor *propensity score*<sup>13</sup>. *Propensity score* jest wartością pochodnej funkcji prawdopodobieństwa doboru obserwacji do programu obliczoną względem wektora cech jednostki. Wartość wektora dla jednostki jest równoważna prawdopodobieństwu zdarzenia, że jednostka należy do grupy eksperymentalnej poddanej działaniu<sup>14</sup>.



Rysunek 2. Współzmiennie dyskretne. Podział na klatki

Źródło: Obliczenia własne.

Termin *propensity score matching* odnosi się do klasy metod wielowymiarowych, stosowanych w badaniach porównawczych w celu konstrukcji grup eksperymentalnych i dołączonych do nich grup kontrolnych o podobnych rozkładach dużej liczby współzmiennych<sup>15</sup>. Łączenie w oparciu o współzmiennie, gdy jest ich dużo, staje się niewykonalne. W badaniach *stricte* eksperymentalnych do obliczenia przeciętnych efektów stosowana jest randomizacja. Łączenie danych, wykorzystujące wektory *propensity score*, jest odpowiednikiem

<sup>13</sup> Rosenbaum P., Rubin D., „The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika* 1983, vol. 70/1, s. 41–55; Rosenbaum P., Rubin D., „Constructing a Control Group Using Multivariate Matched Sampling Method That Incorporate the Propensity Score”, *The American Statistician* 1985, vol. 39, s. 33–38.

<sup>14</sup> Conniffe D., Gash V., O’Connel P., „Evaluating State Programs: «Natural Experiments» and Propensity Scores”, *The Economic and Social Review* 2000, vol. 31/4.

<sup>15</sup> Rubin D., Thomas N., „Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates”, *Journal of the American Statistical Association* 2000, vol. 95, s. 573–585.

randomizacji zastosowanym do danych pochodzących z nieeksperymentalnych badań przekrojowych. Tym, co odróżnia randomizację od łączenia danych, jest to, że pierwsza metoda pozwala na zbilansowanie rozkładów wszystkich współzmiennych, zarówno obserwowanych, jak i niedostępnych dla badacza, natomiast *propensity score matching* pozwala na zbilansowanie wyłącznie względem obserwowanych zmiennych. Ważną własnością techniki *propensity score* jest to, że w prosty sposób można ją połączyć z modelami opartymi na regresji liniowej<sup>16</sup>.

Metoda łączenia danych oparta jest na dwóch założeniach<sup>17</sup>. Pierwsze z nich jest nazywane warunkową niezależnością (ang. *Conditional Independence Assumption*). Zakłada się, że badacz dysponuje zbiorem czynników warunkujących  $W$ , nazywanych współzmiennymi (ang. *covariates*), takich, że w każdej warstwie<sup>18</sup>, zdefiniowanej przez podzbiór  $w \in W$ , rozkład obserwowanych cech jednostek jest taki sam w grupie, na którą oddziaływano, i w grupie kontrolnej.

$$(Y_0, Y_1) \perp\!\!\!\perp P \mid W \quad (4)$$

Oznacza to, że realizacje zmiennej wynikowej są warunkowo, względem zastosowanych w procedurze łączenia współzmiennych  $W$ , niezależne od statusu uczestnictwa w programie. Znajomość wartości zmiennej  $w$  dla jednostki  $i$  nie powinna wskazywać, który stan, eksperymentalny czy kontrolny, jest bardziej prawdopodobny dla tej jednostki. Zatem, dla ustalonej wartości  $w$  jest równie prawdopodobne, że jednostka pochodzi z grupy eksperymentalnej i że pochodzi z grupy kontrolnej. Wobec tego, selekcja jednostki do wskazanej grupy zależy wyłącznie od czynników obserwowanych. Przy spełnionym założeniu dobrana grupa będzie zachowywać się tak, jakby była rzeczywistą grupą kontrolną. Drugim założeniem jest, że wszystkie jednostki poddane działaniu programu mają swój odpowiednik znajdujący się poza programem.

$$0 < \Pr(P_i = 1 \mid W) < 1 \quad (5)$$

Oznacza to, że dla każdej wartości ze zbioru zmiennych warunkujących  $W$  i dla każdego uczestnika programu istnieje z niezerowym prawdopodobieństwem co najmniej jedna jednostka znajdująca się poza programem o identycznych cechach. Ponadto, cechy jednostek nie determinują w sposób jednoznaczny, czy jednostka należy do grupy eksperymentalnej, czy do grupy kontrolnej.

Z założeń (4) i (5) wynika, że gdy znana jest wartość wektora zmiennych warunkujących  $w$ , to oczekiwana wartość zmiennej wynikowej jest niezależ-

<sup>16</sup> Rosenbaum P., *Observational Studies*, Springer Verlag 2002.

<sup>17</sup> Rosenbaum P., Rubin D., „The Central Role...”, op. cit.

<sup>18</sup> Warstwa jest odpowiednikiem klatki.



na od statusu uczestnictwa w programie. Zatem obliczając różnicę w średnich wartościach pomiędzy grupami, można uzyskać oszacowanie przeciętnego efektu oddziaływania na jednostki poddane oddziaływaniu.

Jeżeli przydzielenie do grupy eksperymentalnej lub grupy kontrolnej jest niezależne od obserwowanych cech jednostki, czyli spełnione jest założenie (4) o warunkowej niezależności, to obciążenie powodowane przez porównanie rezultatu w dwóch grupach, wynikające z różnych wartości współzmiennych, może zostać wyeliminowane poprzez ich dopasowanie<sup>19</sup>. Rosenbaum i Rubin<sup>20</sup> pokazali, że zamiast cech  $W$  można zastosować wektor *propensity score*  $PS(w)$ .

$$PS(w) = \Pr(P = 1 | w) \quad (6)$$

Użycie wektorów redukuje problem wielowymiarowy do jednowymiarowego. Zatem analogicznie do (4) zachodzi

$$(Y_0, Y_1) \perp\!\!\!\perp P | PS(w) \quad (7)$$

W takim przypadku uzyskana wartość oszacowania będzie nadal nieobciążonym estymatorem efektu, jednak uzyskany w ten sposób estymator nie musi być efektywny<sup>21</sup>. Nieefektywność metody *propensity score* ujawnia się w sytuacji ewaluacji danych pochodzących z eksperymentu<sup>22</sup>. Jest to związane z niepotrzebnym zastosowaniem przekształcenia danych.

Ta ogólna metoda ma wiele implementacji, różniących się sposobem obliczenia miary podobieństwa. Po pierwsze, w metodzie *propensity score* można wybrać różne postaci funkcyjne dla prawdopodobieństwa poddania oddziaływaniu. Po drugie, można zastosować jedną z dostępnych alternatywnych miar podobieństwa obserwacji.

Po obliczeniu miary podobieństwa jednostek są one łączone za pomocą wybranego przez badacza algorytmu. Łączenie może odbywać się według zasady 1 do 1 (w pary), 1 do wielu (do każdej jednostki poddanej oddziaływaniu dobiera się więcej niż jedną jednostkę z grupy kontrolnej) lub wykorzystując jądrowy estymator gęstości (dla każdej obserwacji obliczana jest wartość dystrybuanty i dobierany jest odpowiednik z drugiej grupy o identycznej lub najbardziej zbliżonej wartości). Na koniec obliczane są różnice w zmiennej wynikowej dla połączonych obserwacji. Te różnice są oszacowaniami szukanых efektów.

<sup>19</sup> Hirano K., Imbens G., Ridder G., „Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score”, *Econometrica* 2003, vol. 71/4, s. 1161–1189.

<sup>20</sup> Rosenbaum P., Rubin D., „The Central Role...”, op. cit.

<sup>21</sup> Heckman J., Ichimura H., Smith J., Todd P., „Characterizing Selection Bias Using Experimental Data”, *Econometrica* 66/4, s. 1017–1098.

<sup>22</sup> Hahn J., „On the Role of the Propensity Score in the Efficient Semiparametric Estimation of Average Treatment Effects”, *Econometrica* 1998, vol. 66/2, s. 315–331.

**Przykład (cd.)**

Dla wcześniej opisanych danych utworzono wektor *propensity score* składający się ze zmiennych  $X_1, X_2, X_3, X_4$ . Przyjęto, że rozkład prawdopodobieństwa będzie opisany przez funkcję logistyczną. Ponieważ dwie z wybranych zmiennych mają charakter dyskretny, nie należy ich bezpośrednio używać do obliczenia wartości *propensity score*. Należy je przekształcić w odpowiednią liczbę<sup>23</sup> zmiennych zero-jedynkowych.

W tabeli 1 znajdują się parametry modelu prawdopodobieństwa doboru jednostek do programu. Ważne są nie tyle statystyczne właściwości modelu, a jego zdolność bilansowania próby.

Tabela 1. Wyniki oszacowania parametrów wektora *propensity score*<sup>24</sup>

Logistic regression	Number of obs	=	150
	LR chi2(10)	=	8.34
	Prob > chi2	=	0.5954
Log likelihood = -91.305639	Pseudo R2	=	0.0437

Program	Coef.	Std. Err.	z	P >  z	[95% Conf. Interval]	
X1	-.4235658	.6660653	-0.64	0.525	-1.72903	.8818982
X2	.0874898	.6530546	0.13	0.893	-1.192474	1.367453
X3_1	.4766884	.5656819	0.84	0.399	-.6320279	1.585405
_X3_2	.1316621	.5958279	0.22	0.825	-1.036139	1.299463
_X3_3	.7350328	.5787153	1.27	0.204	-.3992283	1.869294
_X3_4	.442798	.62718	0.71	0.480	-.7864521	1.672048
_X4_1	-.0507517	.5476217	-0.09	0.926	-1.12407	1.022567
_X4_2	-.9729579	.5419965	-1.80	0.073	-2.035252	.0893358
_X4_3	-.5416392	.6005802	-0.90	0.367	-1.718755	.6354763
_X4_4	.2939577	.5671461	0.52	0.604	-.8176282	1.405544
_cons	-.6043907	.7206802	-0.84	0.402	-2.016898	.8081165

Podczas estymacji *propensity score* należy podjąć dwie decyzje. Pierwsza dotyczy wyboru formy funkcyjnej modelu, druga wyboru zmiennych użytych

<sup>23</sup> Odpowiednia liczba jest o jeden mniejsza niż liczba kategorii zmiennej dyskretnej.

<sup>24</sup> Wszystkie obliczenia przeprowadzono w pakiecie Stata10 z wykorzystaniem dodatkowego modułu PSMATCH2.

do obliczenia wartości wektora. Jak pokazują liczne prace empiryczne<sup>25</sup>, z reguły metoda *propensity score* jest odporna na wybór formy funkcyjnej wektora. Dzieje się tak dlatego, że w praktyce różnice wartości wektora wynikające z zastosowania różnych rozkładów są tak małe, że mogą być pominięte.

Heckman, Ichimura i Todd<sup>26</sup> oraz Heckman i Smith<sup>27</sup> proponują dwie strategie doboru zmiennych do konstrukcji wektora *propensity score*. Pierwsza metoda polega na takim doborze zmiennych, by maksymalizować własności predykcyjne modelu. Obserwacja jest klasyfikowana jako „1” (sukces), jeżeli oszacowana wartość *propensity score* jest większa niż udział jednostek poddanych oddziaływaniu w próbie, w przeciwnym przypadku jest klasyfikowana jako „0” (porażka). Metoda maksymalizuje poprawną klasyfikację obserwacji, zakładając, że koszt błędnej klasyfikacji jest identyczny dla grupy eksperymentalnej i kontrolnej<sup>28</sup>. Należy jednak pamiętać, że celem konstrukcji wektorów *propensity score* jest doprowadzenie do zbilansowania próby, a nie budowa idealnego narzędzia klasyfikującego.

W drugiej metodzie wykorzystywana jest statystyczna istotność zmiennych. Analiza rozpoczyna się od „małego” modelu, zawierającego niewielką liczbę oczywistych zmiennych warunkujących (takich, jak np. charakterystyki demograficzne), a następnie, zgodnie z zasadą od szczegółowego do ogólnego, dodawane są kolejne zmienne. Jeżeli dodana zmienna będzie istotna statystycznie, to jest dołączana do wektora, w odwrotnej sytuacji jest pomijana.

Ważnym zagadnieniem jest dobór zmiennych do wektora *propensity score*. Ekonomiczna teoria, prowadzone badania i wszelkie informacje o programie powinny wskazywać na dobór zmiennych<sup>29</sup>. Założenie warunkowej niezależności wymaga, by zmienne użyte do łączenia były niezależne od stanu poddania oddziaływaniu. W innym przypadku pojawi się problem endogeniczności wywołanej przez sprzężenie zwrotne. Heckman, LaLonde i Smith<sup>30</sup> stwierdzili, że dane dotyczące grupy eksperymentalnej i kontrolnej powinny pochodzić z tego samego źródła. We wcześniejszej pracy Heckman, Ichimura

---

<sup>25</sup> Por. np.: Sianesi B., „An Evaluation of the Swedish System of Active Labor Market Programms in the 1990s”, *The Review of Economics and Statistics* 2004, vol. 86/1; Angrist J., Hahn J., „When to Control for Covariates? Panel Asymptotics for Estimation of Treatment Effects”, *The Review of Economics and Statistics* 2004, vol. 86/1.

<sup>26</sup> Heckman J., Ichimura H., Todd P., „Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *The Review of Economic Studies* 1997, vol. 64/4.

<sup>27</sup> Heckman J., Smith J., Clements N., „Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impact”, *The Review of Economic Studies* 1997, vol. 64/4.

<sup>28</sup> Heckman J., Ichimura H., Smith J., Todd P., „Characterizing...”, op. cit.

<sup>29</sup> Smith J., Todd P., „Does Matching Overcome LaLonde’s Critique of nonexperimental estimators?”, *Journal of Econometrics* 2005, vol. 125, s. 365–375.

<sup>30</sup> Heckman J., LaLonde R., Smith J., „The Economics and Econometrics of Active Labour Market Programs”, w: O. Ashenfelter, D. Card (red.) *Handbook of Labour Economics* 1999, vol 3a, Elsevier.

i Todd<sup>31</sup> pokazali, że pominięcie istotnej zmiennej podczas konstrukcji wektora może spowodować obciążenie uzyskanego estymatora. Ponadto wskazali, że konieczny jest dobry dobór elementów do macierzy warunkującej  $W$ . Jeżeli zostaną uwzględnione tylko takie zmienne, które w sposób statystycznie istotny jednocześnie wpływają na partycypację w programie i jego wynik, to zostanie zagwarantowana poprawna forma funkcyjna  $PS(W)$ . Z kolei Heckman, Ichimura i Smith<sup>32</sup> uzasadnili, że w macierzy  $W$  nie powinny znaleźć się elementy silnie skorelowane ze statusem uczestnictwa w programie  $P$ . Takie zmienne są mocnymi determinantami jednego ze stanów uczestnictwa i zbilansowanie zmiennej tego rodzaju może być kłopotliwe, z uwagi na niską liczbę obserwacji o podobnej wartości cechy w grupie odniesienia. Z drugiej strony idealnymi zmiennymi warunkującymi są te, które są silnie skorelowane ze zmienną wynikową.

## Ocena jakości łączenia

Ważnym elementem w procesie budowy każdego modelu jest ocena jego jakości. Niestety, w przypadku metod łączenia nie wypracowano jeszcze powszechnie akceptowanych narzędzi analitycznych umożliwiających ewaluację poprawności procedury. Niemniej istnieje kilka sposobów weryfikacji zaproponowanych przez różnych autorów.

Łączenie danych z wykorzystaniem wybranej techniki w większym stopniu wpływa na charakterystyki grupy kontrolnej niż eksperymentalnej. Grupa eksperymentalna ulega modyfikacji tylko w przypadku, gdy dla poszczególnych jednostek należących do tej grupy nie udaje się znaleźć bliskiego odpowiednika w grupie kontrolnej. Z kolei charakterystyki grupy kontrolnej przed i po wykonaniu łączenia różnią się w sposób znaczący. Celem łączenia jest wybranie ze zbioru potencjalnych jednostek kontrolnych tych najbardziej zbliżonych do jednostek z grupy eksperymentalnej. Zatem grupa kontrolna po łączeniu powinna być bardziej zbliżona pod względem charakterystyk użytych do łączenia do grupy eksperymentalnej niż przed zastosowaniem tej procedury.

Przy szacowaniu efektów ilościowych ważne jest, aby uzyskany estymator był nieobciążony. Potencjalne obciążenie wynikające z zastosowania procedury łączenia można przedstawić jako sumę trzech elementów: obciążenia wynikającego z niepełnego wspólnego przedziału określoności (ang. *common support*), obciążenia wynikającego z nieprawidłowego przeważenia obserwacji, oraz wynikającego z selekcji na podstawie czynników nieobserwowanych. W metodzie łączenia zakłada się, że cała selekcja dokonuje się na podstawie wartości zmiennych obserwowanych, więc *a priori* obciążenie trzeciego typu

<sup>31</sup> Heckman J., Ichimura H., Todd P., op cit.

<sup>32</sup> Heckman J., Ichimura H., Smith J., Todd P., „Characterizing...”, op cit.

nie występuje. Prawidłowa konstrukcja wektora *propensity score* gwarantuje eliminację obciążenia drugiego typu, ale pod warunkiem, że nie występuje obciążenie pierwszego typu. Proponowane testy diagnostyczne badają rozmiar obciążenia wynikającego z braku spełnienia założenia o wspólnym przedziale określoności.

Sukces procedury łączenia zależy od tego, czy w próbie znajdują się odpowiednie obserwacje pozwalające na skonstruowanie zbioru warunkującego  $W$  w taki sposób, aby spełnione były założenia (4) i (5). Jednym z czynników warunkujących sukces wszystkich metod łączenia, a nie tylko *propensity score matching*, jest wspólny przedział określoności dla wartości cech, lub wartości funkcji obliczonej na podstawie wartości cech, które są wykorzystywane podczas konstrukcji miary podobieństwa. Przy spełnionym założeniu o wspólnym przydziale określoności Hirano, Imbens i Ridder<sup>33</sup> wyprowadzili wzory na efektywne estymatory dla przeciętnych efektów oddziaływania. W przypadku wektorów *propensity score*, którym przypisywana jest wartość z przedziału  $[0; 1]$ , chodzi o to, by rzeczywiste realizacje wartości wektora dla grupy eksperymentalnej i skonstruowanej grupy kontrolnej należały do tego samego przedziału. Dodatkowo, część wspólna przedziału wartości wektora dla grupy eksperymentalnej i grupy kontrolnej powinna być jak największa. Gdy wspólny przedział jest stosunkowo wąski lub tylko niewielka część wszystkich obserwacji należy do wspólnego przedziału, występuje inny problem. Pojawia się pytanie o reprezentatywność i wiarygodność uzyskanego wyniku. Jest on prawidłowy dla jednostek ze wspólnego przedziału, ale można go rozszerzyć na całą grupę tylko wobec spełnienia założenia o jednakowym wpływie oddziaływania na jednostki z grupy eksperymentalnej (ang. *SUTVA*)<sup>34</sup>.

#### Przykład (cd.)

Jeżeli wrócimy do rysunku 1 i rysunku 2, to widać, że w każdej klatce zawierającej jednostkę z grupy eksperymentalnej znajduje się co najmniej jedna jednostka z grupy kontrolnej. Zatem nie występuje ewidentne niedopasowanie. W tabeli 2 przedstawiono wartości wektora *propensity score*. Są one zbliżone w obu grupach, jednak dla części jednostek z grupy eksperymentalnej nie ma dobrych odpowiedników z grupy kontrolnej, gdyż brak jest obserwacji o wartości *propensity score* większej niż 0,54.

Tabela 2. Wartości wektora *propensity score*

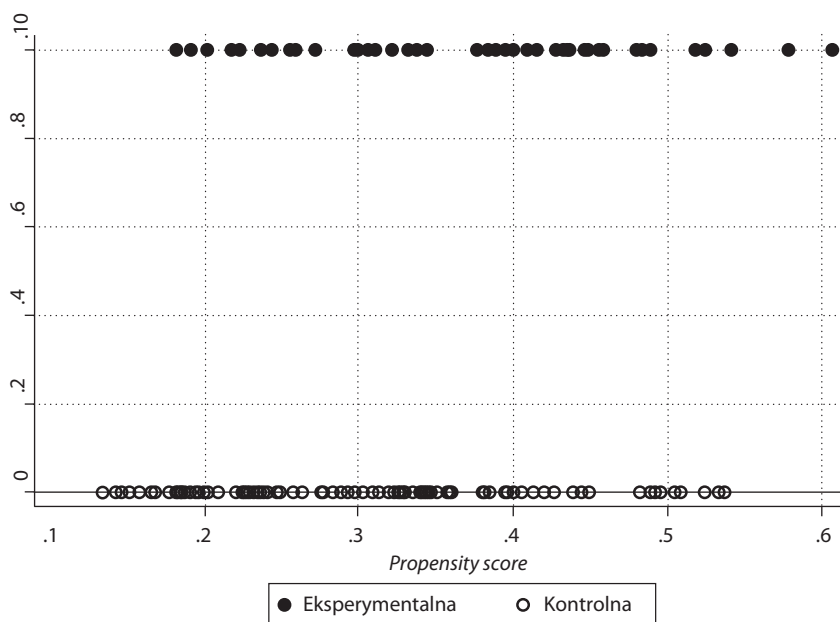
Grupa	Obs	Średnia	Odch. Std.	Min.	Max
Eksperymentalna	50	.368711	.1057437	.1819396	.6062373
Kontrolna	100	.3156445	.1089556	.1342	.5370342

<sup>33</sup> Hirano K., Imbens G., Ridder G., op cit.

<sup>34</sup> Rosenbaum P., Rubin D., „Constructing a Control Group...”. op cit.

Można również przedstawić graficznie rozkład wartości wektora *propensity score* w dwóch grupach.

Brak dobrych odpowiedników nazywany jest problemem braku wspólnego przedziału określoności.



Rysunek 3. Wektor *propensity score*.

Źródło: Obliczenia własne.

Poważnym ograniczeniem metod nieeksperymentalnych jest brak gwarancji istnienia wspólnego przedziału określoności dla cech jednostek uczestniczących w programie i jednostek z grupy kontrolnej. Jeżeli przedziały określoności znacznie się różnią, to złamana zostaje zasada wspólnego przedziału określoności (ang. *common support*). W przypadku homogenicznego oddziaływania nie stanowi to problemu w sensie obciążenia wyniku, jedynie znika efektywność oszacowania, z uwagi na utratę części obserwacji<sup>35</sup>. Gdy oddziaływanie jest heterogeniczne, oszacowane wielkości parametrów dla obu subpopulacji mogą mieć inne znaczenie i interpretację. Z tego powodu niezbędnym elementem analizy wyników jest porównanie rozkładu cech warunkujących  $W$  w obu subpopulacjach.

Efekty przeciętne, które uzyskuje się, między innymi wykorzystując metody łączenia danych, są dobrze zdefiniowane wyłącznie na wspólnym przedzia-

<sup>35</sup> Blundell R., Costa-Diás M., op cit.

le określoności<sup>36</sup>. W literaturze sugerowany jest szereg metod weryfikujących to założenie. Na przykład Lechner<sup>37</sup> proponuje wizualną analizę rozkładów *propensity score*. Bardziej formalne metody polegają na porównaniu minimalnej i maksymalnej wartości wektora *propensity score* w grupie eksperymentalnej i kontrolnej. Obserwacje spoza wspólnego przedziału są wówczas eliminowane z analizy. Jest to szczególnie istotne dla łączenia z wykorzystaniem estymatora jądrowego gęstości rozkładów, który uwzględnia całą informację z grupy kontrolnej. Przy doborze najbliższego sąsiada i tak zawsze jest wybierana najbliższa obserwacja z drugiej grupy. W związku z tym ewentualne obciążenie wynikające z braku wspólnego przedziału określoności jest prawie nieistotne. Problematyczne jest stosowanie tego kryterium w stosunku do obserwacji leżących blisko granicy wspólnego przedziału. W takim przypadku może okazać się, że najbliższe odpowiedniki nie należą do wspólnego przedziału określoności i przez to, po zastosowaniu kryterium, na krańcach wspólnego przedziału dopasowanie będzie gorsze.

Innym problemem może być niska liczba obserwacji w pewnych przedziałach wartości wektora *propensity score* lub gdy wspólny przedział określoności nie jest spójny, tzn. np. wspólny przedział to  $[0.25; 0.8]$ , ale w przedziale  $[0.4; 0.6]$  nie ma obserwacji. Wtedy można podzielić dziedzinę na dwie części i przeprowadzić łączenie osobno w każdej, a następnie obliczyć średni efekt oddziaływania, wykorzystując średnią ważoną liczebnościami. Ale stosując iteracyjnie tę regułę, dojdziemy do absurdalnego rozwiązania, w którym każda wartość będzie znajdować się w osobnym przedziale. Czyli rozwiązaniem granicznym metody iteracyjnej jest dokładne łączenie (ang. *exact matching*). Pozostaną połączone jedynie obserwacje o identycznych wartościach *propensity score*. Gdy co najmniej jedna ze współzmiennych jest ciągła, to nie da się wykluczyć sytuacji, w której nie zostanie utworzona żadna para. Zatem decyzję o ewentualnym podziale dziedziny na części musi arbitralnie podjąć badacz.

Inne rozwiązanie proponują Smith i Todd<sup>38</sup>. Sugerują zastosowanie metody *trimming* do wyznaczenia wspólnego przedziału określoności. Polega ona na tym, że wspólny przedział jest zdefiniowany tylko w tych obszarach, gdzie gęstość rozkładu obserwacji w grupie eksperymentalnej i grupie kontrolnej przekracza wybraną wartość progową  $q > 0$ .

Kolejnym aspektem wymagającym uwagi jest jakość procedury dopasowania. Celem procedury łączenia jest zbliżenie grupy poddanej oddziaływaniu i grupy kontrolnej pod względem cech użytych do łączenia<sup>39</sup>. Zamiast wszyst-

<sup>36</sup> Caliendo M., Kopeinig S., „Some Practical Guidance for the Implementation of Propensity Score Matching”, *Journal of Economic Surveys* 2008.

<sup>37</sup> Lechner M., „An Evaluation of Publicsector Sponsored Continuous Vocational Training Programs in East Germany”, *The Journal of Human Resources*, vol 35/2, s. 347–375.

<sup>38</sup> Smith J., Todd P., op. cit.

<sup>39</sup> Frölich M., „Finite-Sample Properties of Propensity Score Matching and Weighting Estimators”, *The Review of Economics and Statistics* 2004, vol. 86/1, s. 77–90.

kich współzmiennych  $W$  do łączenia używana jest jedna wartość  $PS(W)$ , wobec tego należy sprawdzić zbilansowanie rozkładu poszczególnych zmiennych w grupie eksperymentalnej i skonstruowanej grupie kontrolnej.

Współczesna literatura proponuje dwa sposoby pomiaru skuteczności łączenia<sup>40</sup>. Po pierwsze, można zbadać sposób połączenia na poziomie jednostkowym w skali mikro. Polega to na sprawdzeniu, jak blisko jest każdy element grupy poddanej oddziaływaniu do odpowiadającego mu elementu grupy kontrolnej pod względem cech użytych do łączenia. Rosenbaum i Rubin<sup>41</sup>, jako miarę dopasowania, zaproponowali porównanie standaryzowanego obciążenia przed i po zastosowaniu procedury łączenia. Jest to powszechnie stosowana w badaniach empirycznych metoda walidacji procedury łączenia. Jeżeli obciążenie spada poniżej 5%, jest ono uważane za nieistotnie małe.

Innym sposobem również proponowanym przez Rosenbauma i Rubina<sup>42</sup> jest zastosowanie testu  $t$  dla dwóch prób porównującego średnie wartości zmiennych, które są wykorzystywane w procedurze do obliczenia wartości wektora *propensity score*. Po jego zastosowaniu nie powinno być istotnych statystycznie różnic pomiędzy średnimi wartościami cech użytych do łączenia.

#### PRZYKŁAD (cd.)

Dokonyamy teraz porównania grupy eksperymentalnej i kontrolnej pod względem wartości zastosowanych współzmiennych.

Tabela 3. Porównanie wartości zmiennych użytych do konstrukcji wektora *propensity score*

Zmienna	Próba	Średnia		% redukcji		test t	
		Ekspers.	Kontrol.	% obc	obc	t	p >  t
X1	Niełączona	.46127	.49123	-10.8		-0.62	0.535
	Łączona	.47824	.48429	-2.2	79.8	-0.11	0.916
X2	Niełączona	.51749	.49097	9.2		0.53	0.594
	Łączona	.50725	.57038	-22.0	-138.0	-1.05	0.298
X3_1	Niełączona	.24	.22	4.7		0.27	0.784
	Łączona	.25532	.3617	-25.1	-431.9	-1.11	0.269

<sup>40</sup> Lee M.-J., *Micro-Econometrics for Policy, Program, and Treatment Effects*, Oxford University Press, 2005.

<sup>41</sup> Rosenbaum P., Rubin D., op. cit.

<sup>42</sup> Ibidem.



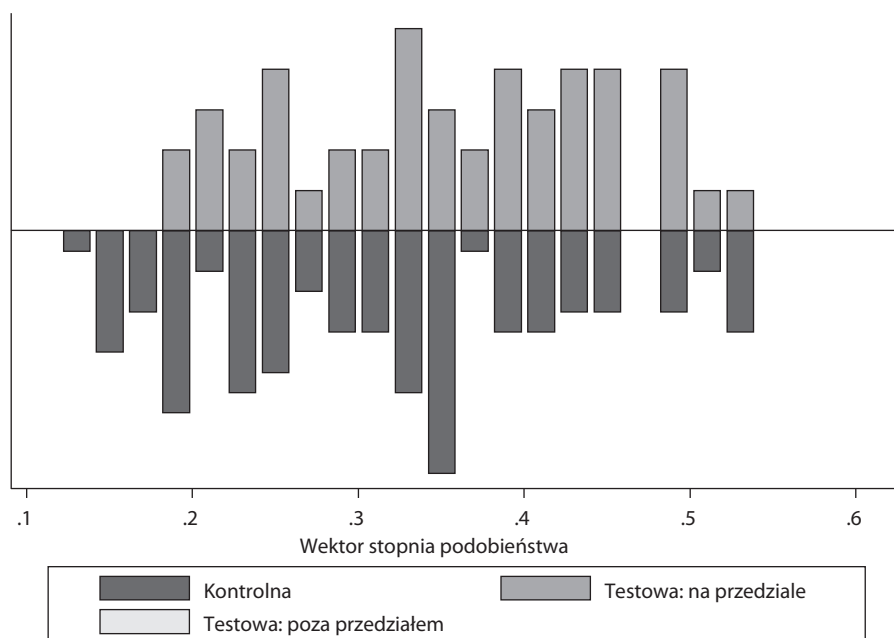
Zmienna	Próba	Średnia		% redukcji		test t	
		Ekspers.	Kontrol.	% obc	obc	t	p >  t
X3_2	Niełączona	.18	.24	-14.7		-0.83	0.406
	Łączona	.19149	.21277	-5.2	64.5	-0.25	0.800
X3_3	Niełączona	.24	.18	14.7		0.86	0.390
	Łączona	.19149	.12766	15.6	-6.4	0.84	0.404
X3_4	Niełączona	.16	.16	0.0		0.00	1.000
	Łączona	.17021	.12766	11.5	.	0.57	0.567
X4_1	Niełączona	.22	.18	9.9		0.58	0.562
	Łączona	.23404	.2766	-10.6	-6.4	-0.47	0.640
X4_2	Niełączona	.18	.34	-36.8		-2.06	0.041
	Łączona	.19149	.19149	0.0	100.0	0.00	1.000
X4_3	Niełączona	.14	.17	-8.2		-0.47	0.639
	Łączona	.14894	.12766	5.8	29.1	0.30	0.768
X4_4	Niełączona	.22	.13	23.7		1.42	0.158
	Łączona	.19149	.19149	0.0	100.0	0.00	1.000

Jak widać w tabeli 3, rozkłady poszczególnych zmiennych są w różnym stopniu zbilansowane. W przypadku wszystkich zmiennych różnice w wartości średnich są statystycznie istotne (test istotności jest w ostatniej kolumnie). Jednak obciążenie wynikające z różnych cech jednostek jest zredukowane poprzez dobranie grupy podobnej pod względem wektora *propensity score* tylko wobec zmiennych  $X_1$  oraz  $X_4$ . Połączenie danych spowodowało zwiększenie różnicy między grupami względem zmiennej  $X_2$  oraz  $X_3$ . Zatem potencjalnie można poprawić efektywność dopasowania, eliminując te zmienne z wektora.

Drugim sposobem jest sprawdzenie zbilansowania (ang. *balance*) obu subpopulacji w skali makro na poziomie grupy, bez zwracania uwagi na cechy pojedynczych jednostek. Polega ono na zbadaniu, czy rozkład wartości wektora *propensity score*, lub wybranej współzmiennnej, jest podobny pod względem przyjętej miary podobieństwa w grupie eksperymentalnej i grupie kontrolnej. To drugie podejście jest dużo częściej stosowane w praktyce ze względu na swoją prostotę. Stosując to podejście, zakłada się, że wszystkie odchylenia na poziomie jednostkowym pomiędzy jednostkami w całej populacji się znoszą (ang. *even-out*).

**Przykład (cd.)**

Często porównanie zbilansowania jest badane z wykorzystaniem analizy graficznej. Wówczas na osi poziomej wykresu znajdują się wartości wektora *propensity score*, a na pionowej wartość funkcji gęstości.



Rysunek 4. Rozkład wartości wektora *propensity score*.

Rozkład będzie wtedy zbilansowany, jeśli, po pierwsze, nie ma dużej liczby jednostek z grupy testowej znajdujących się poza wspólnym przedziałem określoności. Po drugie, gdy w każdym przedziale jednostek w grupie eksperymentalnej jest nie więcej niż w grupie kontrolnej. Należy wziąć pod uwagę, że rysunek 4 jest w inny sposób skalowany dla obu grup.

Heckman, Ichimura i Todd<sup>43</sup> pokazali, że dobór zmiennych do macierzy warunkujące  $W$ , na podstawie której obliczony jest wektor *propensity score*, w praktyce wpływa na wielkość obciążenia uzyskanego oszacowania. Im więcej zmiennych warunkujących zostanie uwzględnionych, tym mniejsze jest obciążenie. Ale tak jak w przypadku regresji liniowej dodanie nieistotnej zmiennej nie powoduje obciążenia, ale zwiększa wariancję uzyskanego oszacowania.

Frölich<sup>44</sup> egzaminował własności różnych technik łączenia wykorzystujących wektory *propensity score* w próbach skończonych. Dla większości esty-

<sup>43</sup> Heckman J., Ichimura H., Smith J., Todd P., „Characterizing Selection...”, op. cit.

<sup>44</sup> Frölich M., op. cit.

matorów występują efekty *trade-off* pomiędzy obciążeniem a precyzją oszacowania. W dużych próbach sytuacja staje się prostsza, gdyż wariancja dąży do zera.

Inną decyzją, którą musi podjąć badacz, jest wybór, co zrobić z niedokładnie połączonymi obserwacjami<sup>45</sup>. Jeżeli duża liczba słabo połączonych par zostanie odrzucona, wówczas zmniejszy się liczebność grupy kontrolnej i dla pewnych obserwacji z grupy eksperymentalnej nie będzie dobrej pary, zatem zostaną one pominięte w analizie. Zredukuje to problem zbilansowania, ale równocześnie zmniejszy się efektywność, z uwagi na pominięcie istotnej części informacji. Z drugiej strony uwzględnienie wszystkich par, bez względu na jakość połączenia, pogorszy zbilansowanie. Jako rozwiązanie Rosenbaum zaproponował, aby porównywać nie zbilansowanie poszczególnych cech, a zastosować podejście na poziomie grupy i zwracać uwagę na zbilansowanie rozkładów cech w całej grupie.

Asymptotycznie wszystkie techniki łączenia danych oparte o wektory *propensity score* powinny dawać takie same rezultaty, gdyż wraz ze wzrostem liczebności próby metoda zbliża się do porównywania obserwacji dokładnie połączonych. Jednakże w małych próbach wybór algorytmu może zaważyć na wyniku<sup>46</sup>. W takim przypadku należy dokonać wyboru między obciążeniem uzyskanego oszacowania efektu a jego wariancją. Nie ma jednego najlepszego sposobu estymacji sprawdzającego się w każdej sytuacji<sup>47</sup>. Jakość poszczególnych technik i wybór estymatora powinny być uzależnione od statystycznych właściwości zbioru danych.

Imbens<sup>48</sup> podkreśla dwa kluczowe jego zdaniem elementy jakości łączenia. Jeżeli do budowy wektora *propensity score* zostaną użyte zmienne, które są słabo skorelowane ze stanem uczestnictwa w programie, to mogą one zwiększyć wariancję oszacowania. Po drugie, nie powinny być uwzględniane zmienne, których wartość jest uzależniona od statusu uczestnictwa w programie.

Angrist i Hahn<sup>49</sup> porównali efektywność łączenia na podstawie współzmiennych (ang. *covariates*) i metody *propensity score*. Hahn<sup>50</sup> udowodnił, że pierwsza jest asymptotycznie efektywna, jednak z uwagi na wielowymiarowość macierzy zmiennych używanych do łączenia, łączenie może okazać się

<sup>45</sup> Lee M.-J., op. cit., s. 89.

<sup>46</sup> Heckman J., Ichimura H., Todd P., „Matching as an Econometric...”, op. cit.

<sup>47</sup> Zhao Z., „Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence”, *The Review of Economics and Statistics* 2004, vol. 86/1, s. 91–107.

<sup>48</sup> Imbens G., „Nonparametric Estimation of the Average Treatment Effects Under Exogeneity: A Review”, *The Review of Economics and Statistics* 2004, vol. 86/1, s. 4–29.

<sup>49</sup> Angrist J., Hahn J., „When to Control for Covariates? Panel Asymptotics for Estimation of Treatment Effects”, *The Review of Economics and Statistics* 2004, vol. 86/1, s. 58–72.

<sup>50</sup> Hahn J., op. cit.

bardzo trudne. PSM jest preferowane, gdy współzmiennie użyte do łączenia wyjaśniają niską część całkowitej wariancji, oraz komórki przez nie stworzone są niewielkie.

#### Przykład (cd.)

Po przeprowadzeniu całej diagnostyki obliczono przeciętny efekt oddziaływania na jednostki poddane oddziaływaniu.

Tabela 4. Obliczenie wielkości efektu

Zmienna	Próba	Eksperym.	Kontrolna	Różnica	Odch. std.	t
wynik	Niełączona	-.010821198	.061218912	-.07204011	.049301042	-1.46
	ATT	-.026268326	.058813968	-.085082295	.068184558	-1.25

psmatch2: Stan oddziaływania	psmatch2: Wspólny przedział		Razem
	Poza	Na	
Poza program	0	100	100
W programie	3	47	50
Razem	3	147	150

Z wyników przeprowadzonej analizy wynika, że jednostki z grupy eksperymentalnej mają niższą wartość zmiennej wynikowej niż jednostki z dobranej grupy kontrolnej. Jednak ta nierówność, po skorygowaniu o różnice w cechach badanych jednostek, nie jest statystycznie istotna.

## Podsumowanie

Omówione tu zostały metody analizy ekonomicznej wykorzystujące łączenie danych w oparciu o technikę *propensity score*. Ich celem jest odtworzenie na podstawie danych przekrojowych warunków eksperymentalnych poprzez konstrukcję hipotetycznej grupy odniesienia (stanu kontrfaktycznego) pozwalającego na porównanie wyników działania programu z sytuacją, w której nie był on przeprowadzony. Odbywa się to poprzez taki wybór jednostek do grupy kontrolnej, by była jak najbardziej zbliżona pod względem obserwowanych charakterystyk do grupy poddanej oddziaływaniu.

Metody łączenia są przydatnym narzędziem w analizie ekonometrycznej, jednak nie zapewniają ogólnego rozwiązania problemu ewaluacji programów pomocy<sup>51</sup>. Ostatnio w literaturze ekonometrycznej pojawiła się grupa prac inspirowanych artykułem Dehejja i Wahba<sup>52</sup>. Wskazują one, że metody estymacji oparte o łączenie za pomocą wektorów *propensity score* mogą być bardziej efektywne niż tradycyjnie używane techniki parametryczne. Dehejja i Wahba<sup>53</sup> pokazali, że proponowane przez nich estymatory mają niższe obciążenie niż tradycyjnie stosowane. Druga grupa prac, odwołująca się do podstawowego artykułu LaLonde'a<sup>54</sup> reprezentowana przez pracę Heckmana, Ichimury i Todda<sup>55</sup>, Heckmana, Ichimury, Smitha i Todda<sup>56</sup> opowiada się za stosowaniem metod parametrycznych, wskazując na niedociągnięcia metod nieparametrycznych. Smith i Todd<sup>57</sup> wskazują na ważną rolę specyficznego zbioru danych skonstruowanego przez Dehejję i Wahbę, oraz na brak odporności ich metody na niewielkie zaburzenia zbioru danych, Odpowiadając na sformułowane zastrzeżenia, profesor Dehejja<sup>58</sup> wskazał na dwa istotne elementy pomijane przez innych autorów. Po pierwsze, wymagana jest oddzielna specyfikacja wektora *propensity score* dla każdego porównania i każdego zbioru danych. Po drugie, musi zostać przeprowadzona analiza odporności wyników, ale nie w stosunku do zbioru danych, a w stosunku do zaproponowanej specyfikacji modelu.

Liczni autorzy wskazują na słabe punkty metod łączenia. Kluczowym elementem tej techniki jest prawidłowe zdefiniowane warunkowej niezależności (4). Słuszność tego założenia powinna być rozpatrywana oddzielnie dla każdego przypadku<sup>59</sup>. Metoda również w dużym stopniu zależy od tego, czy wektory *propensity score* są zdefiniowane na tym samym przedziale. Leuven i Sianesi<sup>60</sup> zaprojektowali użyteczne narzędzia analityczne, sprawdzające

<sup>51</sup> Smith J., Todd P., „Does Matching Overcome LaLonde's Critique of nonexperimental estimators?”, *Journal of Econometrics* 2005, vol. 125.

<sup>52</sup> Dehejja R., Wahba S., „Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Program”, *Journal of American Statistical Association* 1999, vol. 94, no 448, s. 1053–1062.

<sup>53</sup> Dehejja R., Wahba S., „Propensity Score Matching for NonExperimental Causal Studies”, *Review of Economics and Statistics* 2002, vol. 86/1, s. 195–210.

<sup>54</sup> LaLonde R., „Evaluating the Econometric Evaluations of Training Programs with Experimental Data”, *The American Economic Review* 1986, vol. 76/4.

<sup>55</sup> Heckman J., Ichimura H., Todd P., „Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *The Review of Economic Studies* 1997, vol. 64/4.

<sup>56</sup> Heckman J., Ichimura H., Smith J., Todd P., „Characterizing Selection...”, op. cit.

<sup>57</sup> Smith J., Todd P., op. cit.

<sup>58</sup> Dehejja R., „Practical Propensity Score Matching: A Reply to Smith and Todd”, *Journal of Econometrics* 2005, vol. 125, s. 335–364.

<sup>59</sup> Blundell R., Costa-Díaz M., „Alternative Approaches to Evaluation in Empirical Microeconomics”, *IFS Working Paper* 2002, nr 10/02.

<sup>60</sup> Leuven E., Sianesi B., *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*, 2003.

to założenie i porównujące rozkłady wektorów *propensity score*. Jeżeli dziedzina ich określoności dla jednostek poddanych oddziaływaniu znacznie różni się od dziedziny dla jednostek poza programem, można wprowadzić poprawkę zaproponowaną przez Cohrana i Rubina<sup>61</sup>, tzw. *caliper*. Istotny jest fakt, iż poprawki dopasowujące dziedziny mogą wykluczyć większość obserwacji. W takim przypadku uzyskane oszacowania będą mało wiarogodne.

Oceniając metody łączenia, należy również zwrócić uwagę na opinię jej zwolenników. Zdaniem Coniffe i innych<sup>62</sup> brak podobieństwa grupy poddanej oddziaływaniu do grupy kontrolnej nie oznacza, że metody parametryczne są lepsze w modelowaniu zachodzącego zjawiska. Trzeba odnotować, jakie ekonomiczne znaczenie mają wyniki zastosowanych metod parametrycznych w takim przypadku.

Friedlander i Robins<sup>63</sup> pokazują, że statystyczne procedury łączenia danych poprawiają dokładność wyników uzyskiwanych na podstawie danych przekrojowych, ale poprawa jest nieznaczna. Wymieniają szereg czynników powodujących, że estymacja oparta na łączeniu niekoniecznie zwiększa dokładność oszacowań. Po pierwsze, podczas szacowania uwzględniane są jedynie obserwowane charakterystyki jednostek. W rzeczywistości mogą występować znaczne i systematyczne różnice w charakterystykach nieobserwowanych pomiędzy grupą eksperymentalną i kontrolną. Jeśli mają one wpływ na stan poddania oddziaływaniu, wówczas nie jest spełnione jedno z założeń metody. Po drugie, techniki łączenia nie muszą działać w sposób efektywny. Mogą przykładać zbyt małą wagę do najbardziej istotnych pod względem teoretycznym czynników. Po trzecie, założenie, że grupa zakwalifikowana do programu, ale niepoddana jego działaniu, zachowywałaby się identycznie jak grupa kontrolna, może nie być prawdziwe.

Michalopoulos i inni<sup>64</sup> porównali parametryczne i nieparametryczne techniki ewaluacji w zastosowaniu do danych eksperymentalnych. Przedstawili, że sztuczna konstrukcja grupy odniesienia powoduje obciążenie, jednak jest ono duże wyłącznie w dłuższym horyzoncie czasowym. Jeżeli dane wykorzystane do ewaluacji były zebrane w niewielkim odstępie czasowym od wprowadzenia programu, to powstałe obciążenie jest do pominięcia. Po drugie, ma znaczenie, w jaki sposób powstaje grupa kontrolna.

---

<sup>61</sup> Cohran W., Rubin D., „Controlling Bias in Observational Studies”, *Sankhya* 1975, vol 35, s. 417–446.

<sup>62</sup> Coniffe D., Gash V., O’Connell P., „Evaluating State Programs: «Natural Experiments» and Propensity Scores”, *The Economic and Social Review* 2000, vol. 31/4.

<sup>63</sup> Friedlander D., Robins P., „Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods”, *The American Economic Review* 1995, vol 85/4, s. 923–937.

<sup>64</sup> Michalopoulos C., Bloom H., Hill C., „Can Propensity Score Method Match the Findings From a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?”, *The Review of Economics and Statistics* 2004, vol. 86/1, s. 156–179.

Jednostki do niej dobrane powinny pochodzić ze zbliżonego środowiska co jednostki z grupy eksperymentalnej<sup>65</sup>. Po trzecie, nie ma jednej najlepszej techniki eliminacji obciążenia, jakość dopasowania jest determinowana przez własności zbioru danych<sup>66</sup>.

---

**Paweł Strawiński** – doktor nauk ekonomicznych, adiunkt w Katedrze Statystyki i Ekonometrii. Ukończył studia magisterskie na Wydziale Nauk Ekonomicznych Uniwersytetu Warszawskiego oraz uzyskał dyplom Master of Science in Economics na Katholieke Universiteit Leuven (Belgia). W roku 2006 obronił rozprawę dokorską pt. „Ewaluacja programów pomocy publicznej”. Jest autorem licznych publikacji naukowych. Jego badawcze zainteresowania koncentrują się na zastosowaniach współczesnych technik mikroekonometrycznych do analizy zagadnień związanych z funkcjonowaniem rynku edukacyjnego oraz rynku pracy, a także interakcjami pomiędzy tymi segmentami gospodarki.

## Bibliografia

- Angrist J., Hahn J., „When to Control for Covariates? Panel Asymptotics for Estimation of Treatment Effects“, *The Review of Economics and Statistics* 2004, vol. 86/1.
- Blundell R., Costa-Díaz M., „Alternative Approaches to Evaluation in Empirical Microeconomics“, *IFS Working Paper* 2002, nr 10/02.
- Blundell R., Costa-Díaz M., „Evaluation Methods for Non-Experimental Data“, *Fiscal Studies* 2000, vol. 21/4.
- Caliendo M., Kopeinig S., „Some Practical Guidance for the Implementation of Propensity Score Matching“, *Journal of Economic Surveys* 2008.
- Cohran W., Rubin D., „Controlling Bias in Observational Studies“, *Sankhya* 1975, vol. 35.
- Conniffe D., Gash V., O’Connell P., „Evaluating State Programs: «Natural Experiments» and Propensity Scores“, *The Economic and Social Review* 2000, vol. 31/4.
- Dehejia R., „Practical Propensity Score Matching: A Reply to Smith and Todd“, *Journal of Econometrics* 2005, vol. 125.
- Dehejia R., Wahba S., „Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Program“, *Journal of American Statistical Association* 1999, vol. 94, no 448.
- Dehejia R., Wahba S., „Propensity Score Matching for Non-Experimental Causal Studies“, *Review of Economics and Statistics* 2002, vol. 86/1.
- Friedlander D., Robins P., „Evaluating Program Evaluations: New Evidence on Commonly Used Non-Experimental Methods“, *The American Economic Review* 1995, vol. 85/4.

---

<sup>65</sup> Heckman J., Ichimura H., Todd P., „Matching...“, op. cit.

<sup>66</sup> Por.: Frölich M., „Finite-Sample Properties of Propensity Score Matching and Weighting Estimators“, *The Review of Economics and Statistics* 2004, vol. 86/1; Imbens G., „Nonparametric Estimation of the Average Treatment Effects Under Exogeneity: A Review“, *The Review of Economics and Statistics* 2004, vol. 86/1.

- Frölich M., „Finite-Sample Properties of Propensity Score Matching and Weighting Estimators”, *The Review of Economics and Statistics* 2004, vol. 86/1.
- Hahn J., „On the Role of the Propensity Score in the Efficient Semiparametric Estimation of Average Treatment Effects”, *Econometrica* 1998, vol. 66/2.
- Heckman J., Ichimura H., Smith J., Todd P., „Characterizing Selection Bias Using Experimental Data”, *Econometrica* 66/4.
- Heckman J., Ichimura H., Todd P., „Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *The Review of Economic Studies* 1997, vol. 64/4.
- Heckman J., LaLonde R., Smith J., „The Economics and Econometrics of Active Labour Market Programs”, w: O. Ashenfelter, D. Card (red.) *Handbook of Labour Economics* 1999, vol. 3a, Elsevier.
- Heckman J., Smith J., Clements N., „Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impact”, *The Review of Economic Studies* 1997, vol. 64/4.
- Hirano K., Imbens G., Ridder G., „Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score”, *Econometrica* 2003, vol. 71/4.
- Imbens G., „Nonparametric Estimation of the Average Treatment Effects Under Exogeneity: A Review”, *The Review of Economics and Statistics* 2004, vol. 86/1.
- Konarski R., Kotnarowski M., „Zastosowanie Metody Propensity Score Matching w ewaluacji ex-post”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka*, PARP, Warszawa 2007.
- LaLonde R., „Evaluating the Econometric Evaluations of Training Programs with Experimental Data”, *The American Economic Review* 1986, vol. 76/4.
- Lechner M., „An Evaluation of Publicsector Sponsored Continuous Vocational Training Programs in East Germany”, *The Journal of Human Resources*, vol. 35/2.
- Lee M.-J., *Micro-Econometrics for Policy, Program, and Treatment Effects*, Oxford University Press, 2005.
- Leuven E., Sianesi B., *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*, 2002.
- Michalopoulos C., Bloom H., Hill C., „Can Propensity Score Method Match the Findings From a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?”, *The Review of Economics and Statistics* 2004, vol. 86/1.
- Meyer B., „Natural and Quasi-Experiments in Economics”, *Journal of Business and Economic Statistics* 1995, vol. 13/2.
- Rosenbaum P., *Observational Studies*, Springer Verlag 2002.
- Rosenbaum P., Rubin D., „The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika* 1983, vol. 70/1.
- Rosenbaum P., Rubin D., „Constructing a Control Group Using Multivariate Matched Sampling Method That Incorporate the Propensity Score”, *The American Statistician* 1985, vol. 39.
- Rubin D., „Matching to Remove Bias in Observational Studies”, *Biometrics* 1973a, vol. 29.
- Rubin D., „The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies”, *Biometrics* 1973b, vol. 29.
- Rubin D., Thomas N., „Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates”, *Journal of the American Statistical Association* 2000, vol. 95.
- Roy A., „Some Thoughts on the Distribution of Earnings”, *Oxford Economic Papers. New Series*, vol. 3/2.
- Sianesi B., „An Evaluation of the Swedish System of Active Labor Market Programms in the 1990s”, *The Review of Economics and Statistics* 2004, vol. 86/1.



- Smith J., Todd P., „Does Matching Overcome LaLonde’s Critique of nonexperimental estimators?”, *Journal of Econometrics* 2005, vol. 125.
- Smith J., Todd P., „Rejoinder”, *Journal of Econometrics* 2005, vol. 125.
- Strawiński P., „Przyczynowość, selekcja i endogeniczne oddziaływanie”, *Przegląd Statystyczny* 2007 nr 4/2007.
- Zhao Z., „Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence”, *The Review of Economics and Statistics* 2004, vol. 86/1.







