

# **Evaluacja *ex-post***

Teoria i praktyka badawcza





POLSKA AGENCJA ROZWOJU PRZEDSIĘBIORCZOŚCI  
POLISH AGENCY FOR ENTERPRISE DEVELOPMENT

# **Ewaluacja *ex-post***

## **Teoria i praktyka badawcza**

pod redakcją  
Agnieszki Haber

Polska Agencja Rozwoju Przedsiębiorczości

Warszawa 2007

Publikacja pod redakcją *Agnieszki Haber*

Redakcja merytoryczna: *Agnieszka Haber, Jacek Pokorski,  
Jędrzej Stasiowski, Rafał Trzcziński, Maciej Szalaj*

Redakcja techniczna i językowa: *Agnieszka Szopińska,  
Aleksandra Jaskólska*

© Copyright by **Polska Agencja Rozwoju Przedsiębiorczości**

ISBN 978-83-60009-47-5

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

Łamanie: F.W.-P. „Prezes 2” – Jerzy Łazarski

Przygotowanie do druku, druk i oprawa: Pasaż sp. z o.o.

ul. Rydlówka 24; 30-363 Kraków

[www.pasaz.com](http://www.pasaz.com)

## Spis treści

<i>Jacek Szlachta</i> Wprowadzenie . . . . .	7
---	---

### **Część I. Ewaluacja *ex-post* – aspekty teoretyczne i organizacyjne**

<i>Karol Olejniczak</i> Teoretyczne podstawy ewaluacji <i>ex-post</i> . . . . .	15
<i>Agnieszka Haber</i> Proces ewaluacji <i>ex-post</i> – konceptualizacja, operacjonalizacja, realizacja badania ewaluacyjnego . . . . .	43
<i>Maciej Szalaj</i> Mechanizmy doboru kryteriów oceny w ewaluacji <i>ex-post</i> . . . . .	59
<i>Agnieszka Sternik</i> Funkcja autoedukacyjna ewaluacji w rozwoju organizacji. Analiza rozwiązań międzynarodowych . . . . .	75

### **Część II. Wybrane zagadnienia metodologiczne**

<i>Mikołaj Jasiński, Michał Kowalski</i> Fałszywa sprzeczność: metodologia jakościowa czy ilościowa? . . . . .	97
<i>Jarosław Górniak, Barbara Worek, Seweryn Krupnik</i> Zastosowanie podejścia badań jakościowych w ewaluacji <i>ex-post</i> . . . . .	115
<i>Agnieszka Haber, Maciej D. Kryszczuk</i> Wybrane problemy doboru technik badawczych w ewaluacji <i>ex-post</i> . . . . .	137
<i>Jarosław Górniak, Karolina Keler</i> Wskaźniki w ewaluacji <i>ex-post</i> programów publicznych . . . . .	155
<i>Roman Konarski, Michał Kotnarowski</i> Zastosowanie metody <i>propensity score matching</i> w ewaluacji <i>ex-post</i> . . . . .	183



## Wprowadzenie

Praca, którą oddajemy w Państwa ręce, adresowana jest do środowiska profesjonalno-naukowego, zajmującego się zarówno teoretycznymi, jak – przede wszystkim – praktycznymi aspektami ewaluacji. Gwałtownie rosnące potrzeby działań ewaluacyjnych przesądzą o możliwości zastosowania zaproponowanych tu rozwiązań na poziomie krajowym i regionalnym. Ze względu na zakres i charakter tekstów zamieszczonych w tej pracy, może ona zostać wykorzystana w trakcie empirycznie prowadzonych ewaluacji oraz w działaniach o charakterze edukacyjnym. Interdyscyplinarność ewaluacji wyznacza także bardzo szeroki profil zawodowy potencjalnych czytelników.

Już sam tytuł tej książki stanowi istotne wyzwanie. Jeszcze do niedawna różnorodne słowniki języka polskiego nie uwzględniały pojęcia „ewaluacja”, które było zastępowane bardziej ogólnym terminem „ocena”. Dlatego we wszystkich dokumentach rządowych oraz oficjalnych tłumaczeniach aktów prawnych Unii Europejskiej wciąż mamy do czynienia ze stosowaniem w języku polskim słowa „ocena” jako odpowiednika angielskojęzycznego terminu *evaluation*. Na szczęście, wszystkie prestiżowe słowniki języka polskiego oraz najważniejsze słowniki angielsko-polskie wydane w XXI wieku uzupełniły polską terminologię naukową o nowe pojęcie: „ewaluacja”. Wynika stąd, że już najwyższy czas, aby w działalności empirycznej oraz w pracach naukowo-badawczych konsekwentnie odchodzić od dotychczas obowiązującego, a zupełnie niepotrzebnego komplikowania spraw terminologicznych. Pozytywnym zjawiskiem jest tu zastosowanie pojęcia „ewaluacja” we wszystkich wytycznych wydawanych przez Ministerstwo Rozwoju Regionalnego na podstawie Narodowych Strategicznych Ram Odniesienia dla Polski na lata 2007–2013.

Oddajemy w ręce czytelników książkę dotyczącą jednego z segmentów ewaluacji, rozpatrywanej na ogół w trzech przekrojach czasowych – mamy bowiem ewaluację wstępną (*ex-ante*), bieżącą (*on-going*) oraz końcową (*ex-post*). Wyróżnia się także ewaluację śródkresową (*mid-term*).

W Polsce, jak dotąd, główne zainteresowanie nauki i praktyki było zorientowane na ewaluację *ex-ante*. Wynikało to z bardzo utylitarne go podejścia: taki rodzaj ewaluacji programów i projektów pojawił się w naszym kraju na szerszą skalę po roku 2003, w związku z wdrażaniem w Polsce europejskiej polityki spójności, co wyznaczało nowe potrzeby dotyczące funduszy strukturalnych i Funduszu Spójności Wspólnoty. Dzięki temu ukazały się już bardzo

interesujące prace badawcze, zawierające wartościową refleksję teoretyczną i tworzące ramy naukowe dla praktycznych działań ewaluacyjnych typu *ex-ante*. W ostatnich latach wykonano wiele takich ewaluacji dla programów zaprojektowanych do realizacji na lata 2004–2006 i 2007–2013. Tworzą one bardzo bogaty katalog materiałów źródłowych, mogących stać się w przyszłości podstawą szerszych analiz naukowych.

Ewaluacja *ex-post* uruchamiana jest po zakończeniu realizacji poszczególnych programów operacyjnych. W Polsce zakres badań naukowych dotyczących ewaluacji *ex-post* jest niestety bardzo ubogi, a działania i doświadczenia praktyczne ograniczają się do programów przedakcesyjnych Unii Europejskiej, realizowanych przed rokiem 2004. Pierwsze ewaluacje zostały wykonane w naszym kraju dla programu Phare, jeszcze w latach dziewięćdziesiątych ubiegłego wieku. Najbardziej wartościowe ewaluacje były podejmowane na zlecenie Polskiej Agencji Rozwoju Przedsiębiorczości dla różnych segmentów Phare edycji 2000, 2001, 2002 oraz 2003, dotyczących Programu Spójności Społeczno-Gospodarczej. Jednak w ciągu najbliższych kilku lat niezbędne będzie przeprowadzenie ewaluacji *ex-post* palety programów Unii Europejskiej, które były wdrażane w Polsce w latach 2004–2006. Dlatego wiedza na ten temat staje się coraz bardziej użyteczna.

Ponieważ brak jest rodzimych, teoretycznych podstaw badań ewaluacyjnych, a dobre praktyki ewaluacyjne *ex-post* są dopiero wypracowywane, niniejsza praca ma na rynku polskim niemal pionierski charakter. Podawana na końcu każdego artykułu bibliografia przedmiotu – w aspekcie działań ewaluacyjnych typu *ex-post* – jest prawdopodobnie jednym z najbogatszych zbiorów takiej literatury w Polsce. Trzon bibliografii stanowią naukowe teksty anglosaskie, uzupełnione różnymi publikacjami i raportami Unii Europejskiej. Taka struktura źródeł i odniesień informacyjnych nie powinna szczególnie dziwić czytelnika, ewaluacja bowiem pojawiła się najwcześniej w teorii i praktyce funkcjonowania gospodarki Stanów Zjednoczonych Ameryki Północnej, potem była rozwijana w Wielkiej Brytanii, a następnie adaptowana i modyfikowana w ramach polityki spójności Unii Europejskiej. Do Polski nowoczesna metodologia ewaluacji *ex-post* trafiła za pośrednictwem Komisji Europejskiej jako obligatoryjnie obowiązujący element europejskiej polityki spójności.

Istotną wartością tej książki jest podjęcie dyskusji o teoretycznych i praktycznych wymiarach ewaluacji *ex-post*. Ewaluacja *ex-post* posiada szczególne cechy, uczestników, charakter oraz funkcje. Także więc zestaw technik (metod) badawczych musi być odmienny niż w przypadku ewaluacji *ex-ante*. Praca ta zawiera także próbę pewnego uporządkowania aparatu pojęciowego. Chaos terminologiczny, wynikający z chałupniczego tłumaczenia terminów pochodzących z literatury anglosaskiej, jest jednym z poważniejszych problemów tej sfery, zarówno w aspekcie naukowo-badawczym, jak zwłaszcza implementacyjnym, gdyż różni autorzy posługują się często bardzo nieudolnymi kalkami terminów angielskojęzycznych.



Najpoważniejszym wyzwaniem, jakie przed nami stoi w tej sferze, jest nie tylko merytoryczne przygotowanie, a następnie przeprowadzenie ewaluacji *ex-post* dla zestawu programów realizowanych w latach 2004–2006, lecz także zasadnicze poszerzenie zasięgu badań ewaluacyjnych typu *ex-post* na różne polityki i programy publiczne, realizowane bez udziału funduszy Unii Europejskiej. Niniejsza praca jest krokiem w tym kierunku.

Książka składa się, poza (1) Wprowadzeniem, z dziewięciu artykułów: (2) Teoretyczne podstawy ewaluacji *ex-post*, (3) Proces ewaluacji *ex-post* – konceptualizacja, operacjonalizacja, realizacja badania ewaluacyjnego, (4) Mechanizmy doboru kryteriów oceny w ewaluacji *ex-post*, (5) Funkcja autoedukacyjna ewaluacji w rozwoju organizacji. Analiza rozwiązań międzynarodowych, (6) Fałszywa sprzeczność: metodologia jakościowa czy ilościowa?, (7) Zastosowanie podejścia badań jakościowych w ewaluacji *ex-post*, (8) Wybrane problemy doboru technik badawczych w ewaluacji *ex-post*, (9) Wskaźniki w ewaluacji *ex-post* programów publicznych oraz (10) Zastosowanie metody *propensity score matching* w ewaluacji *ex-post*.

W artykule Karola Olejniczaka, zatytułowanym „Teoretyczne podstawy ewaluacji *ex-post*”, przedstawiono teoretyczne podstawy działalności ewaluacyjnej, bazując przede wszystkim na bardzo bogatej anglosaskiej bibliografii przedmiotu, uzupełnionej najważniejszymi opracowaniami Komisji Europejskiej. Dla osób szerzej zainteresowanych różnymi aspektami ewaluacji artykuł ten może stanowić bardzo wartościowy przewodnik po literaturze przedmiotu. Na podstawie głębokich studiów literaturowych przeprowadzonych przez Autora wykazano, jakie są najważniejsze odmienności ewaluacji *ex-post*, w porównaniu z innymi jej rodzajami, zwłaszcza z tymi prowadzonymi w układzie *ex-ante*. Autor precyzyjnie dokumentuje, jakie mogą być efekty poprawnie prowadzonej działalności ewaluacyjnej *ex-post*. Przy lekturze tego tekstu pojawia się pytanie o możliwości wypracowania w naszym kraju własnych teoretycznych podstaw ewaluacji, dostosowanych do specyfiki Polski, by nie bazować wyłącznie na bardzo odległych kulturowo wzorcach anglosaskich.

W artykule Agnieszki Haber „Proces ewaluacji *ex-post* – konceptualizacja, operacjonalizacja, realizacja badania ewaluacyjnego” opisano szczegółowo modelowy przebieg tego procesu; omówiono cele i etapy procesu ewaluacji, wyszczególniając jej poszczególne fazy. Wartość dodana artykułu polega, między innymi, na usystematyzowaniu zestawu zadań stojących przed instytucją dokonującą ewaluacji, które zostały omówione krok po kroku na tle tzw. cyklu życia ewaluacji. Dzięki temu czytelnik może sobie wyobrazić, w jaki sposób poprawnie prowadzi się ewaluację *ex-post*. W dalszych pracach, należących do tego ważnego nurtu modelowania procesu ewaluacyjnego, niezbędne będzie rozwinięcie problemu różnych zagrożeń występujących w poszczególnych fazach badania. Prawdopodobnie w przyszłości zostaną także określone zalecane lub obowiązujące standardy w poszczególnych etapach i fazach.

Tekst Macieja Szałaja „Mechanizmy doboru kryteriów oceny w ewaluacji *ex-post*” podejmuje jeden z kluczowych problemów tych działań, jakim jest prawidłowe określenie kryteriów oceny. W artykule szczegółowo przedstawiono warunki prawidłowego przebiegu tego procesu. Scharakteryzowano wpływ różnorodnych czynników na wybór tych kryteriów, a następnie na tej podstawie przytoczono ich dokładną klasyfikację. Zwrócono uwagę na konsekwencje (niestety, często negatywne) zbyt jednostronnego wiązania ewaluacji *ex-post* z europejską polityką strukturalną. Skomentowane zostały polskie doświadczenia ewaluacji *ex-post*, dotyczące programów przedakcesyjnych Unii Europejskiej. Interesująca jest zawarta w tym artykule refleksja na temat meta-kryterium ewaluacji.

Tekst przygotowany przez Agnieszkę Sternik, „Funkcja autoedukacyjna ewaluacji w rozwoju organizacji. Analiza rozwiązań międzynarodowych”, wskazuje na znaczenie prawidłowej realizacji funkcji edukacyjnej w trakcie procesu ewaluacji. Uczenie się w oparciu o ewaluację, czyli zwiększenie roli informacji zwrotnej i różnorodnych rekomendacji, służy wspomaganie rozwoju organizacji i doskonaleniu prowadzonych przez nią interwencji. Bardzo interesującym dla czytelnika fragmentem tego rozdziału jest zestawienie rozwiązań instytucjonalnych w zakresie ewaluacji stosowanych przez wybrane organizacje międzynarodowe, takie jak: Komisja Europejska, Program Narodów Zjednoczonych ds. Rozwoju, Ministerstwo Spraw Zagranicznych Danii, Wysoki Komisarz Organizacji Narodów Zjednoczonych ds. Uchodźców, Organizacja Narodów Zjednoczonych oraz USAID (*United States Agency for International Development*) oraz sformułowanie wniosków merytorycznych na podstawie przeprowadzonej analizy. Umiejętność wdrażania wniosków wynikających z ewaluacji jest jednym z najważniejszych wymiarów tej działalności. Dlatego organizacje międzynarodowe, zarówno te przedstawione przez Autorkę, jak też inne, tworzą naturalny katalog najlepszych praktyk (*best practices*), które powinny być adaptowane w Polsce.

Mikołaj Jasiński i Michał Kowalski przedstawili artykuł zatytułowany „Fałszywa sprzeczność: metodologia jakościowa czy ilościowa?”. Autorzy wskazują na szczególne ograniczenia ewaluacji *ex-post*, wynikające z pokusy manipulowania danymi, co wiąże się z momentem podejmowania takiego badania. Opisano różnorodne problemy związane z wyborem odpowiedniej metodologii do oceny ewaluacyjnej. Następnie wskazano na możliwości wynikające z tzw. triangulacji, przeprowadzanej w różnych wymiarach, która pozwala na skuteczne połączenie metod ilościowych i jakościowych. Autorzy pokazali różne modele integracji obydwu rodzajów metod oraz opisali ich przydatność do ewaluacji *ex-post*. W końcowej części tego artykułu podjęto kwestię mierzenia efektu netto interwencji, stanowiącą jeden z kluczowych tematów dotyczących ewaluacji i systematycznie rozwijanych w odniesieniu do funduszy strukturalnych Unii Europejskiej. Ze względu na charakter przygotowania zawodowego osób zajmujących się ewaluacją istnieje kon-

flikt pomiędzy zwolennikami stosowania technik ilościowych a sympatykami badań jakościowych. Ponieważ potrzebne są obydwa rodzaje ujęć, opracowanie to ma istotne walory aplikacyjne.

W kolejnym tekście Jarosław Górniak, Barbara Worek i Seweryn Krupnik omówili „Zastosowanie podejścia badań jakościowych w ewaluacji *ex-post*”. Bardzo szczegółowo przedstawiono specyfikę jakościowych badań ewaluacyjnych typu *ex-post* oraz różnorodne korzyści wynikające z ich systematycznego prowadzenia. Zwrócono uwagę na znaczenie stosowania metod jakościowych odnośnie możliwości wyjaśniania i przewidywania zjawisk społeczno-ekonomicznych. Poprawie jakości wnioskowania w ramach ewaluacji *ex-post* mogą służyć m.in. następujące techniki: matryce skutków, matryce dynamiki przypadku oraz sieci przyczynowe. Szczególnie przydatna w takich badaniach może być ewaluacja oparta na teorii programu, której założenia, etapy oraz warianty zostały przez Autorów szczegółowo scharakteryzowane. Artykuł ten poszerza wiedzę na temat znanych w świecie technik ewaluacji, jakie mogą zostać szerzej zastosowane w Polsce w nadchodzących latach.

Agnieszka Haber i Maciej Kryszczuk przedstawili tekst zatytułowany „Wybrane problemy doboru technik badawczych w ewaluacji *ex-post*”. Autorzy scharakteryzowali funkcje ewaluatora, a następnie na tej podstawie opisali zmieniające się parametry badań ewaluacyjnych. Szczegółowo wyjaśnione zostały kryteria doboru technik badawczych, czyli: (1) uzyskanie rzetelnych i trafnych informacji, (2) wyznaczony cel, (3) zakres i charakter poszukiwanych informacji, (4) moment dokonywanej oceny, (5) podmiot zlecający, (6) podmiot realizujący, (7) perspektywa podmiotowo-celowa (kto jest beneficjentem), (8) źródła, jakość i dostępność danych, (9) jakość, dostępność i zakres bazy danych o beneficjentach, (10) zasoby czasowe oraz (11) zasoby finansowe. Scharakteryzowano następnie wybrane techniki badawcze ze względu na najważniejsze czynniki decydujące o ich ewentualnym zastosowaniu. W końcowej części tego artykułu omówiono kontekst instytucjonalny badań ewaluacyjnych, prowadzący niekiedy do zniekształcania lub zasadniczego obniżania jakości uzyskiwanych wyników oraz decydujący o unikalności tego typu aktywności badawczej. W anglosaskiej literaturze przedmiotu prezentowana jest paleta możliwych do zastosowania technik ewaluacyjnych. Aby wybrać metodę odpowiednią do przedmiotu badania, niezbędna jest bardzo staranna analiza, przeprowadzona jeszcze przed podjęciem właściwej ewaluacji. Wartość dodana tego tekstu polega na systemowym przedstawieniu takiej analizy.

Ósmy artykuł, przygotowany przez Jarosława Górniaka i Karolinę Keler, omawia „Wskaźniki w ewaluacji *ex-post* programów publicznych”. Opisano w nim problem wyboru odpowiednich wskaźników, które mają być wykorzystane w procesie ewaluacji. Główna część tego artykułu została poświęcona celom interwencji publicznych. Uwzględniono między innymi ich następujące aspekty: (1) hierarchia celów przedstawiona na tle różnych pojęć angielskich,

odpowiadających polskiemu terminowi „cel”, (2) zadania ewaluacji *ex-post* oraz określanie celów, (3) dwoisty charakter celów będących wyzwaniem lub zadaniami, (4) powiązania ogólnych celów interwencji z celami poszczególnych projektów na poziomie dotyczącym beneficjentów, (5) problem identyfikowania, dla różnych poziomów celów interwencji, konkretnej wartości przewidywanej do osiągnięcia. Opisano rolę diagnozy społecznej, dokonywanej za pomocą wskaźników, w pomiarze użyteczności programu. W końcowej części tego tekstu przedstawiono różnorodne kryteria oceny przyjmowanych do ewaluacji wskaźników. Stosowanie tych kryteriów zdecydowanie pomaga wybrać najlepsze wskaźniki. W zakończeniu wskazano na znaczenie dla ewaluacji *ex-post* dobrze zdefiniowanych i wyrażonych poprzez wskaźniki celów.

Roman Konarski i Michał Kotnarowski przedstawili tekst „Zastosowanie metody *propensity score matching* w ewaluacji *ex-post*”. Metoda *propensity score matching* (PSM) jest jednym z instrumentów ilościowych, służących oszacowaniu efektu podejmowanej interwencji publicznej. Wykorzystanie metody PSM wymaga bardzo starannej selekcji grupy kontrolnej. Autorzy opisują formalną stronę tej metody, a następnie przedstawiają jej zastosowanie do analizy efektu netto komponentu regionalnego programu przedakcesyjnego Phare SSG RZL 2002, dotyczącego rozwoju zasobów ludzkich. W końcowej części Autorzy dokumentują możliwości zastosowania metody PSM w badaniach ewaluacyjnych typu *ex-post*.

Po akcesji do Wspólnoty Europejskiej w bardzo szerokim zakresie podejmuje się w Polsce działania ewaluacyjne, które mają, niestety, bardzo ubogie podstawy teoretyczne. Niezbędne jest pilne rozszerzenie praktyk ewaluacyjnych na wszystkie sfery funkcjonowania administracji publicznej, także te, które nie są związane z funduszami strukturalnymi Unii Europejskiej. Praca ta częściowo wypełnia związaną z tym dotkliwą lukę merytoryczną.

*Jacek Szlachta*

Część I

**Ewaluacja *ex-post***  
**– aspekty teoretyczne i organizacyjne**



Karol Olejniczak

## Teoretyczne podstawy ewaluacji *ex-post*<sup>1</sup>

### Wstęp

Celem niniejszego rozdziału jest wprowadzenie czytelnika w kwestie ewaluacji *ex-post*. Są one omawiane na szerszym tle ewaluacji interwencji publicznych, w odniesieniu do praktyki Unii Europejskiej i USA. Rozdział ma charakter głównie techniczny i przeglądowy. Zostają w nim kolejno omówione: definicja ewaluacji – w tym specyfika i zakres ewaluacji *ex-post*, funkcje ewaluacji *ex-post* w teorii i praktyce UE, paleta możliwych podejść badawczych wraz z zaznaczeniem ich mocnych stron i ograniczeń. Ponadto w rozdziale przeanalizowane są role uczestników procesu badawczego oraz możliwe formy użycia badań ewaluacyjnych (z przywołaniem również kwestii negatywnego wykorzystania).

Mimo swojej technicznej orientacji rozdział stawia także szerszą tezę – trzeba, by ewaluacja *ex-post* korzystała z metod interaktywnych i podejścia realistycznego, ewaluator powinien zaś odchodzić od roli naukowca na rzecz roli edukatora (wymagającej większego zaangażowania oraz dialogu ze zlecającym i interesariuszami interwencji).

### Co to jest ewaluacja *ex-post*?

#### Definicja ewaluacji

Semantycznie ewaluacja to oszacowanie wartości<sup>2</sup>. W takim przypadku termin „ewaluacja” zasadniczo odpowiadałby polskiemu słowu „ocena”.

---

<sup>1</sup> Niniejszy rozdział to fragment pracy doktorskiej pt. *Ewaluacja w zarządzaniu programami INTERREG* przygotowanej pod kierunkiem Prof. Grzegorza Gorzelaka. Rozdział jest efektem projektu badawczego zrealizowanego w 2006 r. na Indiana University, Bloomington, w ramach stypendium badawczego Indiana University i Uniwersytetu Warszawskiego. Jako autor chciałbym podziękować pracownikom IU za opiekę merytoryczną i organizacyjną. Szczególne podziękowania kieruję do Prof. Elinor Ostrom, Prof. Davida Reingold'a oraz Christy Borders. *Last but not least*, dziękuję kadrze Polish Studies Center – w osobach Prof. Justyny Beinek, Prof. Billa Johnstona oraz Andrew Hinnant'a za ciepłe przyjęcie i stworzenie rodzinnej atmosfery podczas mego pobytu na obczyźnie.

<sup>2</sup> Oxford University Press, *Oxford English Dictionary*, Oxford University Press, Oxford 2002.

Jednak zarówno definicje ewaluacji obecne w literaturze tematu, jak i praktyka, podkreślają jej odmienność od oceny<sup>3</sup>.

W literaturze tematu można znaleźć wiele definicji ewaluacji, podkreślających różne aspekty samego procesu ewaluacji, uwypuklających jej zróżnicowane cele czy też odmienne kryteria samego momentu oceny. Najszerzej ujmując, ewaluacja to systematyczne badanie społeczno-ekonomiczne oceniające jakość i wartość programów publicznych<sup>4</sup>.

Podkreśla się szczególnie trzy elementy. Po pierwsze, uwypuklany jest analityczny i systematyczny/systemowy charakter procesu (*systematic inquiry*). Oznacza to stosowanie podejścia naukowego popartego danymi, przestrzeganie rygoru badawczego wynikającego z kanonu badań nauk społecznych (niezależnie od jakościowej lub ilościowej orientacji badania czy łączenia metod badawczych)<sup>5</sup>. Po drugie, zaznacza się, iż ewaluacja bada zarówno jakość danej interwencji (*merit-quality*), jak i jej wartość w kategoriach finansowych, ekonomicznych (*worth-value*)<sup>6</sup>. Wreszcie trzecim elementem podkreślanym w definicjach ewaluacji jest to, iż badanie ewaluacyjne obejmuje ocenę zarówno procesów działania programu/projektu, jak i jego efektów<sup>7</sup>.

Tym, co odróżnia ewaluację od innych narzędzi badania i dyscyplinowania sektora publicznego (takich jak audyt, monitoring, badania *stricte* naukowe) są cztery kwestie:

---

<sup>3</sup> Użycie angielskiego terminu jest bardziej wskazane z kilku powodów. „Ewaluacja” jest procesem. „Ocena” jest tylko jednym z elementów tego procesu. Według niektórych podejść osąd nie zawsze jest kluczowy w procesie ewaluacji, a może być nawet pominięty (akcent stawia się wtedy na interakcje i animacje negocjacji pomiędzy uczestnikami programu). Użycie terminu „ocena” koncentruje uwagę wyłącznie na jednym elemencie, przypisując mu cechy całego procesu, a także wyłącza cały szereg alternatywnych podejść. Wreszcie, po trzecie, „ewaluacja” to nie tylko pojedyncze działanie, ale dziedzina wiedzy. Jak zauważa Kierzkowski (Kierzkowski T., *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2002, s. 10), ze względów semantycznych trudno byłoby więc określać terminem „ocena” zespół koncepcji teoretycznych.

<sup>4</sup> Forunier D.M., „Evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 139–140, Sage Publications, Thousand Oaks–London 2005; Patton M.Q., „Evaluation research”, w: M.S. Lewis-Beck, A. Bryman, T.F. Liao (red.), *The Sage Encyclopedia of Social Science Research Methods*, s. 337–340, Sage Publications, Thousand Oaks 2004; Rossi P.H., Lipsey M.W., Freeman H.E., *Evaluation: A Systematic Approach*, Sage Publications, Thousand Oaks–London 2004; Weiss C.H., *Evaluation: Methods for Studying Programs and Policies*, Prentice Hall, Upper Saddle River 1998.

<sup>5</sup> Langbein L., Felbinger C.L., *Public Program Evaluation. A Statistical Guide*, M.E.Shape, New York–London 2006; Patton M.Q., „Evaluation research”, op.cit.; Rossi P.H., Lipsey M.W., Freeman H.E., *Evaluation...*, op.cit.; Scriven M., „Pros and cons about goal-free evaluation”, w: G.V. Glass (red.), *Evaluation Studies Review Annual*, vol 1., Sage Publications, Beverly Hills 1976, s. 131.

<sup>6</sup> W niektórych przypadkach terminy te są stosowane rozłącznie, w innych natomiast wymienianie. Por. Davidson E.J., *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*, Sage Publications, Thousand Oaks 2005, s. 1–2; Paddock S.C., „Evaluation”, w: J.M. Shafritz (red.), *International Encyclopedia of Public Policy and Administration*, s. 818–823, Westview Press, Boulder 1998.

<sup>7</sup> Bingham R.D., Felbinger C.L., *Evaluation in Practice: A Methodological Approach*, Seven Bridges Press, New York–London 2002, s. 3; Weiss C.H., *Evaluation: Methods...*, op.cit., s. 4.



- łączenie empirii (badanie programu i zjawisk pojawiających się jako konsekwencje programów) z charakterem normatywnym (ocenie struktury działań, ich logiki, ich wartości)<sup>8</sup>,
- utylitaryzm – ewaluacja ma wspomagać konkretnych odbiorców w konkretnej sytuacji<sup>9</sup>, wywołać pozytywne zmiany i poprawiać jakość tak bieżącej, jak i przyszłych, analogicznych interwencji<sup>10</sup>,
- interaktywność i negocjacyjne podejście do procesu badania, będącego jednocześnie procesem uczenia się<sup>11</sup>,
- ograniczona siła wpływu – zależna od chęci przyswajania wiedzy przez instytucje zlecające badanie i ich gotowości do zmian, reform (w odróżnieniu od audytu, wyniki ewaluacji nie mają żadnej prawnej siły dyscyplinującej).

Definicja funkcjonująca w odniesieniu do programów i projektów UE jest dużo bardziej techniczna. Według Komisji Europejskiej ewaluacja to „ocena interwencji publicznej pod kątem jej rezultatów, oddziaływania oraz potrzeb, które ma spełniać/miała zaspokoić”<sup>12</sup>.

## Definicja i zakres ewaluacji *ex-post*

Ewaluacja *ex-post* (nazywana także *a posteriori*, końcową, zamykającą) to badanie ewaluacyjne wykonane po zakończeniu interwencji (programu, działania, projektu itd.). W praktyce UE oznacza to nie później niż 3 lata po zakończeniu danego działania. Ewaluacja *ex-post* jest więc jednym z rodzajów ewaluacji według typologii momentu wykonania badania<sup>13</sup>.

<sup>8</sup> Forunier D.M., „Evaluation”, op.cit., s. 139.

<sup>9</sup> Stufflebeam D.L., „Evaluation models”, *New Directions for Evaluation* 2001, nr 89, s. 11.

<sup>10</sup> Chelimsky E., *Program Evaluation: Patterns and Directions*, American Society for Public Administration, Washington 1985, s. 7; Cronbach L.J. i in., *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*, Jossey-Bass, San Francisco 1981, s. 14; Weiss C.H., *Evaluation: Methods...*, op.cit., s. 5.

<sup>11</sup> Greene J.C., „Understanding social programs through evaluation”, w: N.K. Denzin, Y.S. Lincoln (red.), *Collecting and Interpreting Qualitative Materials*, Sage Publications, Thousand Oaks–London 2003, s. 592.

<sup>12</sup> European Commission, *Focus on Results: Strengthening Evaluation of Commission Activities, SEC2000/1051*, European Commission 2000.

<sup>13</sup> Inne wyróżniane rodzaje ewaluacji ze względu na moment prowadzenia badania to: ewaluacja *ex-ante* (przed rozpoczęciem interwencji), *mid-term* (również w połowie trwania interwencji), ewaluacje *ad hoc* lub *on-going* (bieżące, wykonywane w dowolnym momencie wdrażania programu, zwykle jednak w jego zaawansowanym stadium). Por. European Commission, *MEANS Collection: Evaluating Socio-economic Programmes*, DG Regio, Office for Official Publications of the European Communities, Luxembourg 1999, vol. 1; GHK Consultants, *The Evaluation of Socio-Economic Development. The Evalsed Guide (2nd edition)*, www.evalsed.com, 31.12.2006; Olejniczak K., „Teoria i praktyka ewaluacji w świetle doświadczeń europejskich”, *Studia Regionalne i Lokalne* 2005, nr 4(22). Dla obecnego okresu programowania, lat 2007–13 UE przewiduje trzy rodzaje ewaluacji: *ex-ante*, *on-going*, *ex-post* (Council of the European Union, *Council Regulation (EC) No 1083/2006 of 11 July 2006 laying down on the European Regional Development Fund, the European Social Fund and the Cohesion Fund*, Official Journal of the European Communities L210, 31.7.2006, art. 47–49).

Zgodnie z wytycznymi i praktyką europejską ewaluacja *ex-post* obejmuje pytania z zakresu czterech kryteriów ewaluacyjnych: skuteczności (*effectiveness*), wydajności (*efficiency*), użyteczności (*utility*) oraz trwałości (*sustainability*).

W praktyce ewaluacja *ex-post*<sup>14</sup>:

- Zawiera badanie i ocenę wykorzystania środków – skuteczności, wydajności pomocy. Pozwala odpowiedzieć na następujące pytania: Czy cele interwencji zostały osiągnięte? Jak silny był wpływ czynników zewnętrznych? Jakie były sukcesy interwencji? Jakie problemy napotkano? Czy wybrane instrumenty i rozwiązania pomocy okazały się odpowiednie? Czy podobne efekty można osiągnąć przy wykorzystaniu innych instrumentów?;
- Obejmuje badanie i ocenę *post factum* funkcjonowania systemu wdrażania interwencji (skuteczność instytucjonalna): Czy system zarządzania i wdrażania był skuteczny? Czy instytucje poprawnie wypełniły przypisane im role?;
- Bada i przedstawia ocenę użyteczności rzeczywistych efektów – rezultatów i oddziaływania pomocy (w tym efektów ubocznych, tak pozytywnych, jak i negatywnych), pytając: Czy interwencja spełniła oczekiwania adresatów? Czy przyczyniła się do rozwiązania adresowanych problemów? Czy efekty były korzystne dla różnych grup odbiorców? Czy pojawiły się pozytywne lub negatywne efekty uboczne?<sup>15</sup>;
- Bada trwałość osiągniętych pozytywnych efektów: Czy efekty działań są trwałe, długookresowe? Czy są odczuwalne po zakończeniu interwencji? Jeśli tak, to w jakim stopniu i dla kogo?;
- Identyfikuje czynniki, które przyczyniły się do sukcesu lub niepowodzenia danej interwencji;
- Formułuje konkluzje, które mogą być przenoszone na inne, analogiczne interwencje, identyfikuje najlepsze praktyki;
- Formułuje wnioski dotyczące polityki w zakresie spójności gospodarczej i społecznej.

<sup>14</sup> Council of the European Union, *Council Regulation (EC) No 1260/1999 of 21 June 1999 laying down General Provisions on the Structural Funds*, Official Journal of the European Communities L161, 26.6.1999, vol. 5; Council of the European Union, *Council Regulation (EC) No 1083/2006...*, op.cit.; European Commission, *MEANS Handbook 6, The Ex-post Evaluation of Structural Funds Programmes: A Methodological Framework*, XVI D., Commission of the European Communities, Brussels 1995; European Commission, *MEANS Collection...*, op.cit.

<sup>15</sup> Typowe pytania o pozytywne efekty uboczne to: Jakie środki publiczne zmobilizowane zostały przez środki pomocowe (efekt dodatkowości)? Czy pomoc zmobilizowała jakieś środki prywatne (efekt dźwigni)? Czy zrealizowana inicjatywa pobudziła kolejne inwestycje (efekty impulsu, mnożnikowe)? Czy inwestycja sprzegła się z innymi działaniami publicznymi, wzmacniając je (efekty synergii)? Standardowe pytania o negatywne efekty uboczne to: Czy wykonane inwestycje powstałyby i bez pomocy UE? Jeśli tak, to czy przesunięcie czasowe byłoby znaczące (efekt *deadweight*, tzw. biegu jałowego)? (szerzej zob.: Kierzkowski T., *Ocena (ewaluacja) programów i projektów...*, op.cit.; Olejniczak K., „Teoria i praktyka ewaluacji...”, op.cit.).

## Czemu służy ewaluacja *ex-post*?

### Funkcje ewaluacji

Ewaluacja jest standardem współczesnego zarządzania w sektorze publicznym – stanowi integralny element zarządzania przez cele i działania (*Management by objectives, Activity-Based Management*)<sup>16</sup> oraz promowanego w UE do końca lat 90-tych „Przejrzystego i efektywnego zarządzania” (*Sound and Efficient Management*)<sup>17</sup>.

Według tych podejść podstawowym celem każdej ewaluacji (niezależnie od jej typu czasowego) jest podnoszenie jakości, spójności i skuteczności interwencji publicznych<sup>18</sup>. Ta rola ewaluacji wynika z założenia, iż działania sektora publicznego, choć często prowadzone w warunkach rynkowych, nie mogą być oceniane wyłącznie z perspektywy osiągniętego zysku<sup>19</sup>. Kryteria osądu wartości programów publicznych są zróżnicowane i zależne od wielu czynników, w tym od dominujących przekonań czy aktualnych potrzeb społeczności. 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>20</sup>.

Te zadania mogą być realizowane przez szereg funkcji. W literaturze tematu przedstawiane są różne kategoryzacje<sup>21</sup>, ale zasadniczo można wyróżnić 5 głównych funkcji ewaluacji:

---

<sup>16</sup> Dahler-Larsen P., „Evaluation in public management”, w: E. Ferlie, L.E. Lynn, C. Pollitt (red.), *The Oxford Handbook of Public Management*, s. 615–639, Oxford University Press, Oxford 2005; Heinrich C.J., „Measuring public sector performance”, w: G. Peters, J. Pierre (red.), *Handbook of Public Administration*, s. 25–37, Sage Publications, London 2003; Wollmann H., *Evaluation in Public Sector Reform: Concepts and Practice in International Perspective*, Edward Elgar, Cheltenham 2003.

<sup>17</sup> European Commission, *Focus on Results...*, op.cit.; European Commission, *Implementing Activity-Based Management in the Commission*, SEC(2001)1197/6 & 7, European Commission 2000; European Commission, *White Paper on European Governance – Report of the Working Group Evaluation & Transparency*, Working Group 2B, European Commission 2001, [http://europa.eu.int/comm/governance/index\\_en.htm](http://europa.eu.int/comm/governance/index_en.htm).

<sup>18</sup> Chelimsky E., *Program Evaluation...*, op.cit., s. 7; Council of the European Union, *Council Regulation (EC) No 1083/2006...*, op.cit., art. 47.1; Mark M.M., Henry G.T., Julnes G., *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*, Jossey-Bass, San Francisco 2000, s. 3; Rist R.C., *Program Evaluation and the Management of Government: Patterns and Prospects across Eight Nations*, Transaction Publishers, New Brunswick–London 1990, s. 5.

<sup>19</sup> Stewart J., Ranson S., „Management in the public domain”, w: D. McKevitt, A. Lawton (red.), *Public Sector Management: Theory, Critique and Practice*, s. 54–70, Open University, Sage Publications, London 1994.

<sup>20</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.

<sup>21</sup> Bachtler J., „Quod erat demonstrandum? Ewaluacja polityki regionalnej”, *Studia Regionalne i Lokalne* 2001, nr 4(7); Batterbury S., „Principles and purposes of European Union Cohesion policy evaluation”, *Regional Studies* 2006, nr 40(2), s.179–188; Chelimsky E., „The coming transformation in evaluation”, w: E. Chelimsky, W.R. Shadish (red.), *Evaluation for the 21st Century: A Handbook*, Sage Publications, Thousand Oaks–London 1997; Mark M.M., Henry G.T., Julnes G., *Evaluation: An*

- 1) Poprawianie planowania (*improving planning*) oznacza:
  - zagwarantowanie racjonalnego uzasadnienia dla danej interwencji poprzez zweryfikowanie jego spójności zewnętrznej, przystawalności celów do potrzeb i problemów;
  - zapewnienie wydajnej, racjonalnej i skutecznej alokacji zasobów poprzez zweryfikowanie logiki interwencji, racjonalności wyznaczonych celów, spójności zewnętrznej planowanej strategii.
- 2) Poprawa wdrażania i kontrola jakości (*improving performance & quality*) obejmuje:
  - poprawienie istniejącej struktury, procedur i procesu zarządzania interwencją publiczną (*improving programme management*);
  - minimalizowanie negatywnych efektów ubocznych zarządzania opartego na wynikach<sup>22</sup>;
  - podnoszenie efektywności dostarczania działań poprzez weryfikowanie, na ile interwencja realizuje zakładane cele, jaka jest jakość produktów (*improving programme delivery*).
- 3) Wzmocnienie odpowiedzialności (*improving accountability*) to:
  - informowanie szerszej opinii publicznej o skuteczności działań publicznych – na ile cele zostały osiągnięte, czy pojawiły się dodatkowe efekty;
  - informowanie szerszej opinii publicznej o wydajności podjętych działań – jak efektywnie użyto przypisanych zasobów publicznych.
- 4) Wspieranie procesów uczenia się oznacza:
  - wyjaśnianie czynników i procesów warunkujących działania, ich sukces lub porażkę (w tym także czynników ubocznych);
  - weryfikowanie teorii leżących u podstaw danych polityk, programów, projektów;
  - wyciąganie wniosków dla innych programów i projektów, promowanie dobrych wzorców;
  - budowanie zdolności ewaluacyjnych (*building capacity*).
- 5) Wzmacnianie partnerstwa i poczucia współwłasności (*improving partnership & ownership*) obejmuje:
  - wzmacnianie powiązań, relacji, współpracy, interakcji między uczestnikami danej inicjatywy;
  - aktywizowanie interesariuszy – włączanie ich w działania interwencji, dyskusję nad jej oceną i kierunkami rozwoju;
  - uprawnianie interesariuszy – włączanie we współdecydowanie tych grup, które były marginalizowane (*empowerment*).

---

*Integrated Framework...*, op.cit.; Tavistock Institute, GHK, IRS, *The Evaluation of Socio-Economic Development. The Evalsed Guide (1st edition)*, www.evalsed.com 2003, downloaded 21.08.2006.

<sup>22</sup> Szerzej na temat negatywnych skutków, czyli tzw. dysfunkcji zarządzania opartego na wynikach zob.: Olejniczak K., „Teoria i praktyka ewaluacji...”, op.cit., s. 38; na podstawie: Smith P., „On the unintended consequences of publishing performance data in the public sector”, *International Journal of Public Administration* 1995, nr 18, s. 277–310.

Poszczególne funkcje ewaluacji w praktyce nie są rozdzielne. Pojedyncze badanie ewaluacyjne może, a nawet powinno spełniać kilka funkcji. To, na jakie funkcje zostaje położony akcent, zależy w dużej mierze od typu czasowego ewaluacji, skali badanej interwencji (program, projekt) oraz od zastosowanego podejścia badawczego (szerzej o tym – w odniesieniu do ewaluacji *ex-post* – w dalszej części rozdziału).

Moment wykonania badania ewaluacyjnego wyraźnie ogranicza niektóre z powyższych funkcji i uwypukla inne. W rezultacie pewne ewaluacje mają charakter normatywny (*normative evaluation*) – poprawiają zarówno racjonalność działań, jak i bieżące zarządzanie programem, służą głównie instytucjom wdrażającym daną interwencję; inne natomiast mają charakter podsumowujący (*summative evaluation*) – opisowy i wyjaśniający, dokonują bowiem całościowej oceny, służą aktorom zewnętrznym<sup>23</sup>. W obecnych rozporządzeniach UE na temat funduszy strukturalnych te dwie funkcje są nazywane odpowiednio operacyjną i strategiczną<sup>24</sup>.

Warto także zwrócić uwagę, że istnieje pewne napięcie między niektórymi funkcjami. I tak, w przypadku łączenia funkcji poprawy wdrażania i budowania partnerstwa z funkcją wzmacniania odpowiedzialności, trzeba dokonać wyraźnego wyboru (*trade-off*)<sup>25</sup>. Odpowiedzialność ma w sobie silny element kontroli (rozliczenie z efektów, błędów, opóźnień przed opinią publiczną). Natomiast bieżące usuwanie problemów (funkcja poprawy zarządzania) działa najlepiej w sytuacji otwartej, szczerzej i autokrytycznej dyskusji o problemach i blokadach, prowadzonej w gronie operatorów interwencji i ewaluatorów. Wizja upublicznienia tej dyskusji blokuje uczestników procesu w owych rozmowach. Aspekt rozliczeniowy ogranicza też bieżące procesy uczenia się – wymagają one zaufania i otwartości pomiędzy uczestnikami, czego nie można osiągnąć, wiedząc, że przebieg całego badania (a więc i autokrytyczne wypowiedzi) będzie upubliczniony i stanie się podstawą ostatecznej oceny. Konkluzja (warta dodatkowego sprawdzenia empirycznego) jest taka, że zbyt wiele odpowiedzialności, zwłaszcza egzekwowanej na wczesnym etapie rozwoju interwencji może blokować jej wdrażanie. Z tego płynie wniosek praktyczny, że funkcja odpowiedzialności powinna być budowana przede wszystkim w ewaluacji *ex-post*. Ewaluacje *on-going* winny się zaś koncentrować na poprawie zarządzania, partnerstwa i wiedzy operacyjnej.

---

<sup>23</sup> Scriven M., *Evaluation Thesaurus*, Sage Publications, Newbury Park–London 1991, s. 168–169, 340.

<sup>24</sup> Council of the European Union, *Council Regulation (EC) No 1083/2006...*, op.cit., art. 45.2.

<sup>25</sup> Eser T.W., Nussmueller E., „Mid-term evaluations of Community Initiatives under European Union Structural Funds: a process between accounting and common learning”, *Regional Studies* 2006, nr 40(2), s. 249–258; Stern E., *What do we know about the utilization of evaluation?*, I konferencja ewaluacyjna – Ewaluacja programów o charakterze społeczno-gospodarczym finansowanych z funduszy strukturalnych, Ministerstwo Gospodarki i Pracy, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 16 września 2005, s. 25.

## Cel i funkcje ewaluacji *ex-post*

Ewaluacja *ex-post* ma charakter podsumowujący (strategiczny), a jej główne funkcje to:

- wzmocnienie odpowiedzialności – przede wszystkim zbadanie relacji przyczynowo-skutkowej między podjętą interwencją a efektami i procesami wywołanymi w danym układzie, a w rezultacie pokazanie sukcesu lub porażki interwencji;
- wsparcie procesów uczenia się – szczególnie weryfikowanie teorii, założeń i paradygmatów leżących u podstaw interwencji.

Z uwagi na ciągłość programów publicznych i przesunięcie czasowe typowe dla funduszy strukturalnych ewaluacje *ex-post* w małym stopniu pomagają w planowaniu i wdrażaniu (w momencie, gdy dostępne są wnioski z ewaluacji *ex-post* programy kolejnej edycji są już zwykle w 1/3 wdrożone). Przełożenie na planowanie i poprawę systemu wdrażania może dotyczyć dalszych edycji programu – ale z reguły wykorzystuje się wnioski bardziej ogólne (kierunek strategii, sens procedur), a nie kwestie operacyjne, bo te i tak już w międzyczasie uległy zmianie.

Programy publiczne – szczególne interwencje UE – są wieloaspektowe, wdrażane na wielu poziomach. To sprawia, że liczne instytucje i grupy społeczne (tzw. interesariusze, *stakeholders*) zarówno włączają się w sam proces wdrażania programów, jak i odczuwają ich efekty. Każdy z tych aktorów procesu ewaluacji może stawiać inne pytania (patrz zakres ewaluacji *ex-post*) i każdemu badanie *ex-post* daje co innego.

Dla Komisji Europejskiej wyniki ewaluacji *ex-post* spełniają funkcję odpowiedzialności. Pokazują, czy sensownie zagospodarowano środki UE (pozwalają też rozliczyć aktorów krajowych). Po drugie, wypełniają funkcję uczenia się; pomagają zweryfikować paradygmaty polityki (np. niedawna zmiana w podejściu do funduszy strukturalnych i zaakcentowanie wspierania najbardziej konkurencyjnych regionów), czyli odpowiedzieć na pytanie: czy założenia teoretyczne, na których opierają się programy, sprawdziły się w praktyce? Należy jednak zaznaczyć, że jest to wpływ kumulatywny (wniosków z wielu badań). Ponadto, ewaluacje *ex-post* dają Komisji możliwość porównań pomiędzy skutecznością efektów programów w różnych krajach UE (pytania o to, które z praktyk okazały się najbardziej skuteczne, jakie wzorce warto promować i powielać w innych krajach).

Dla instytucji zarządzających danym programem ewaluacja *ex-post* jest swoistym sposobem rozliczenia się przed Komisją Europejską i przed społeczeństwem ze swojej działalności, z racjonalności zarządzania i efektywności gospodarowania finansami publicznymi (pytania o to, czy dobrze spożytkowano pieniądze, jaki był ich rzeczywisty wpływ na dany układ społeczno-gospodarczy, czy rzeczywiście pomogły). Wnioski z tego typu badań mogą również służyć refleksji nad racjonalnością danego programu (pomagają

odpowiedzieć na pytanie o sens i skuteczność danej pomocy, a także o alternatywne rozwiązania na przyszłość). Wreszcie, takie ewaluacje są również narzędziem uczenia się (pytania o to, co warto poprawić w przyszłości, co warto powtórzyć w kolejnych edycjach programu). Tu jednak wiele zależy od tzw. pamięci instytucjonalnej – stabilności zaangażowanej kadry oraz, oczywiście, umiejętności gromadzenia i wykorzystywania doświadczeń w ramach danej instytucji.

Dla instytucji pośredniczących, beneficjentów końcowych (realizujących projekty), wyniki ewaluacji *ex-post* mogą być głównie źródłem inspiracji na przyszłość (czego dotyczyły i jak zostały wykonane najbardziej efektywne i innowacyjne projekty). Stanowią też okazję ujrzenia swojego projektu w szerszym kontekście: jeśli projekt był dobry i pokazano go jako jedną z najlepszych praktyk, jego ewaluacja staje się również swoistą promocją.

Wreszcie, dla szerokiej rzeszy beneficjentów ostatecznych i społeczeństwa wyniki ewaluacji *ex-post* są źródłem informacji o skuteczności i zasadności gospodarowania „groszem publicznym” (pytania o to, jaki był rzeczywisty efekt interwencji publicznej, w czym rzeczywiście pomogła, czy nie zaszkodziła). Dla naukowców wyniki te stanowią także cenne źródło danych do weryfikowania teorii działania całych polityk.

Do kwestii wykorzystania ewaluacji powrócimy jeszcze w dalszej części tego rozdziału przy omawianiu możliwych form użycia konkretnych badań.

### **Jakie są podejścia badawcze, uczestnicy i użycie ewaluacji *ex-post*?**

Podejścia do ewaluacji *ex-post* są zróżnicowane, a co za tym idzie, ewaluatorzy przyjmują różne role. Zróżnicowanie to wynika z dwóch odmiennych filozofii badawczych, współistniejących w ewaluacji – pozytywizmu (czy konstrukcjonizmu) i, zbliżonego do niego, realizmu. Ponadto wybór podejścia badawczego jest warunkowany celem badania (czyli tym, jakie funkcje ma ono spełnić) i wielkością badanej interwencji (program *vs.* projekt).

#### **Podejścia badawcze**

Na podstawie literatury tematu, a także praktyki (amerykańskiej i europejskiej) można wyróżnić kilka głównych podejść badawczych do ewaluacji (w tym ewaluacji *ex-post*).

**Podejście eksperymentalne i *quasi*-eksperymentalne** (*experimental & quasi-experimental design*) opiera się na paradygmacie pozytywistycznym i skupia się na badaniu relacji przyczynowo-skutkowych między zaplanowanymi działaniami danej interwencji a osiągniętymi zmianami w danej

populacji, na danym obszarze<sup>26</sup>. Ewaluator jest naukowcem – „strażnikiem prawdy”, kierującym się przy ocenie obiektywnymi kryteriami naukowymi. W badaniu wykorzystywane są metody ilościowe, eksperymentalne i analiza statystyczna. Rdzeń ewaluacji stanowi badanie sytuacji przed interwencją i po niej, w formie eksperymentu (jeśli zastosowany jest losowy wybór grupy) lub *quasi*-eksperymentu (gdy grupa kontrolna nie jest losowa)<sup>27</sup>.

Zasadniczą przewagą tego modelu nad pozostałymi jest możliwość jasnego zidentyfikowania – w oparciu o obiektywnie naukową analizę – relacji przyczynowo-skutkowej (lub jej braku) między działaniem interwencji a jej efektem. Za słabą stronę uznaje się natomiast trudności w zastosowaniu tego modelu przy dużych, zróżnicowanych interwencjach realizowanych na rozległym obszarze (problem stworzenia poprawnej grupy kontrolnej i kontrolowania wszystkich zmiennych), a także brak wyjaśnienia samego procesu, który spowodował ewentualne efekty netto. Ponadto, w przypadku Polski poważnym ograniczeniem jest brak danych (fragmentaryczność, krótkie sekwencje czasowe, brak danych kontekstowych, zmieniające się sposoby i definicje kategorii danych, np. zmiany NTS).

**Podejście wyłączające cele programu** (*goal-free evaluation model*) opiera się na założeniu, że prawdziwa wartość programu może być oceniona na podstawie rzeczywistych zmian w otoczeniu, a nie na podstawie tego, co starano się zmienić<sup>28</sup>. Oznacza to, że w całym badaniu pomija się zapisane cele, koncentrując się na otwartej obserwacji i identyfikacji rzeczywistych

<sup>26</sup> Bingham R.D., Felbinger C.L., *Evaluation in Practice...*, op.cit.; Campbell D.T., Stanley J.C., *Experimental and Quasi-Experimental Designs for Research*, Rand McNally College Publishing Company, Chicago 1967; Shadish W.R., Cook T.D., Campbell D.T., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston 2001.

<sup>27</sup> Przy określaniu relacji przyczynowo-skutkowej zastosowany może być jeden z kilku podtypów tego modelu – w zależności od dostępnych danych, charakteru grupy docelowej, możliwości kontrolowania zmiennych (Shadish W.R., Cook T.D., Campbell D.T., *Experimental and Quasi-Experimental Designs...*, op.cit.). Po pierwsze, wyróżnia się podejście pre-eksperymentalne (*reflexive design*) w dwóch opcjach: porównanie danej grupy poddanej programowi przed i po programie (*one group pretest-posttest design*) lub tej samej grupy w kilku momentach przed programem z jednorazowym badaniem po programie (*simple time-series design*). Po drugie, wyróżnia się właściwe podejście *quasi*-eksperymentalne w dwóch opcjach: badanie grupy docelowej poddanej interwencji i grupy porównawczej, nie poddanej interwencji (*pretest-posttest comparison group*), oraz wreszcie „prawdziwie naukowe” podejście, tzw. klasyczny plan eksperymentalny, w którym pomiary początkowe i końcowe są wykonywane i porównywane między grupą poddaną interwencji (eksperymentalną) oraz specjalnie stworzoną na potrzeby badania grupą kontrolną, dla której wszystkie zmienne są kontrolowane przez badaczy (*pretest-posttest control group*), a dobór próby odbywa się na zasadzie losowej. Ten ostatni model można prowadzić wyłącznie w warunkach laboratoryjnych. Nie ma więc on zastosowania do programów publicznych rozwoju regionalnego. Dodatkowo wyróżnia się również plan czynnikowy (*factorial design*) – podtyp badania eksperymentalnego, dający możliwość jednoczesnego badania wpływu kilku zmiennych zależnych na zmienną niezależną i wykrywania relacji między nimi. W takim przypadku próbuje się kontrolować i mierzyć siłę oddziaływania poszczególnych czynników, stosując dodatkowe testy statystyczne (Bingham R.D., Felbinger C.L., *Evaluation in Practice...*, op.cit.).

<sup>28</sup> Scriven M., „Pros and cons...”, op.cit.



zmian wywołanych interwencją (zarówno pozytywnych, jak i negatywnych)<sup>29</sup>. Obserwacje te są następnie odnoszone do potrzeb i problemów grupy docelowej<sup>30</sup>. Ewaluator wartościuje i ocenia, kierując się swoim doświadczeniem, krytyczną interpretacją tego, co ważne, a przede wszystkim kryteriami wypracowanymi na bazie analizy alternatywnych efektów programu i rzeczywistych potrzeb odbiorców (dobre działanie to takie, które spełniło potrzeby odbiorców). Autor tego podejścia, M. Scriven, podkreśla użycie głównie metod ilościowych, metody jakościowe są jednak również dopuszczalne. Zaletą tego podejścia jest całościowa analiza efektów programu – mówiąca o efektach ubocznych i wiążąca je z oczekiwaniami grupy docelowej. Pomijanie zapisanych celów programu wywołuje jednak opór ze strony zarządzających programem, wymaga też dużo szerszej analizy kontekstowej, co może być trudne w przypadku napiętego harmonogramu badawczego.

**Podejście uczestniczące** (*participatory evaluation*), opierając się na filozofii konstrukcjonizmu i teoriach uczenia się organizacji, stawia ewaluacji dwa cele. Po pierwsze, zrozumienie interwencji poprzez uwzględnienie różnych perspektyw jej postrzegania i włączenie interesariuszy w proces ewaluacji (szczególnie w momencie oceny). Po drugie, poprawę interwencji poprzez wypracowanie konsensusu wśród interesariuszy i kumulowanie wiedzy instytucjonalnej<sup>31</sup>. W założeniu, siła ewaluacji (pozytywna zmiana, którą wywołuje ewaluacja) jest proporcjonalna do tego, w jakim stopniu są reprezentowane opinie różnych interesariuszy, w jakim stopniu interesariusze są włączani w badanie (w określenie jego celów, ocenę efektów, identyfikację problemów), na ile rozumieją proces i utożsamiają się z nim (*sense of ownership*).

W tym podejściu ewaluatorzy przede wszystkim pełnią rolę animatorów dyskusji, doradców wprowadzających nowe perspektywy i wiedzę, asystentów technicznych, gromadzących i podsumowujących dane i informacje do dyskusji oraz sugerujących narzędzia, koordynatorów procesu, a także negocjatorów pomiędzy włączonymi w badanie uczestnikami ewaluacji.

Propagatorzy tego podejścia różnie rozmieszczają akcenty w kwestii roli ewaluatora. Niektórzy autorzy zaznaczają, że ewaluator włącza uczestników w proces formułowania kryteriów, jednak ostatecznej oceny dokonuje sam, niezależnie od nich. Co więcej, poszukuje konsensusu, ale nie stara się go

---

<sup>29</sup> W tym celu wykorzystywana jest tzw. metoda *Modus Operandi* – ciąg przyczynowy. Ewaluator tworzy pełną listę potencjalnych efektów programu (identyfikując alternatywne rozwiązania i sprzeczne efekty), zawęża potencjalne wybory poprzez określenie, które z potencjalnych przyczyn występowały zanim pojawiły się efekty oraz który ciąg przyczynowy w całości pasuje do zaobserwowanych zmian i zdarzeń.

<sup>30</sup> Alkin M.C., *Evaluation Roots: Tracing Theorists' Views and Influences*, Sage Publications, Thousand Oaks–London 2004, s. 33.

<sup>31</sup> Cousins J.B., Whitmore E., „Framing participatory evaluation”, *New Directions for Evaluation* 1998, nr 80, s. 5–23; King J.A., „Participatory evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 291–294, Sage Publications, Thousand Oaks–London 2005.

budować, gdy ten nie istnieje<sup>32</sup>. Inni natomiast twierdzą, że ewaluator identyfikuje potrzeby informacyjne uczestników, zbiera informacje merytoryczne, a następnie ułatwia jedynie dialog i negocjacje pomiędzy interesariuszami dokonującymi ostatecznej oceny. Stara się więc aktywnie budować konsensus i wiedzę<sup>33</sup>.

Szczególną odmianą podejścia partycypatywnego jest „ewaluacja skoncentrowana na użyciu” (*utilization-focused evaluation*)<sup>34</sup>. Takie badanie ma pomóc konkretnej grupie adresatów w konkretny sposób. Ewaluatorzy mają zatem za zadanie zmaksymalizować wykorzystanie badania przez konkretnych użytkowników, dostosowując wszystkie jego elementy do zakładanego użycia<sup>35</sup>.

Głównymi narzędziami badania są metody jakościowe i interaktywne: wywiady, panele dyskusyjne, warsztaty – zarówno na etapie tworzenia pytań badawczych, kryteriów oceny programu, zbierania opinii, jak i samej oceny. Niektórzy autorzy silnie podkreślają również rolę studiów przypadku<sup>36</sup>. Siłą tego podejścia jest aktywizowanie szerokiej rzeszy odbiorców badania oraz proces uczenia się. Trudność stanowi natomiast zachowanie przez ewaluatorów obiektywizmu w badaniu i integralności badania<sup>37</sup>, a także zapewnienie,

<sup>32</sup> Stake R., „A modest commitment to the promotion of democracy”, *New Directions for Evaluation* 2000, nr 85, s. 104.

<sup>33</sup> Alkin M.C., *Evaluation Roots...*, op.cit., s. 39; Torres R., Preskill H., „Evaluation and organizational learning: past, present and future”, *New Directions for Evaluation* 2001, nr 22, s. 387–395.

<sup>34</sup> Termin został wprowadzony przez Michela Pattona (Patton M.Q., *Utilization-Focused Evaluation*, Sage Publications, Beverly Hills 1978; Patton M.Q., *Utilization-Focused Evaluation: The New Century Text*, Sage Publications, Thousand Oaks 1997). Zbliżone merytorycznie koncepcje to: *responsive evaluation* promowana przez Roberta Stake’a (Stake R., „Responsive evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 376–379, Sage Publications, Thousand Oaks–London 2005) czy *user-oriented evaluation* Marvina Alkina (Alkin M.C., *Evaluation Roots...*, op.cit., s. 49).

<sup>35</sup> Ewaluator nie ocenia sam, a współpracuje i pomaga przyszłym użytkownikom ewaluacji w wypracowaniu kryteriów oceny i w samej ocenie (Alkin M.C., *Evaluation Roots...*, op.cit., s. 50). Wymaga to zidentyfikowania grupy potencjalnych użytkowników, włączenia ich w proces ustalania celów i form badania (w tym do dyskusji nad metodami), aktywnej współpracy z użytkownikami w procesie zbierania danych, ich interpretacji, oceny i wypracowywania rekomendacji, wspólnego stworzenia strategii rozpowszechniania wyników (Patton M.Q., *Utilization-Focused Evaluation: The New Century Text*, op.cit.). Ewaluator działa analogicznie jak w podejściu partycypatywnym, ale szczególnie akcent jest kładziony na elastyczność i dostosowywanie się do potrzeb użytkowników (Owen J., Lambert F., „Evaluation and the information needs of organizational leaders”, *American Journal of Evaluation* 1998, nr 19, s. 355–365; Stake R., „Responsive evaluation”, op.cit.). Stosowana zasada to *active-reactive-adaptive*: bądź aktywny (w identyfikowaniu potencjalnych użytkowników, włączaniu ich, precyzowaniu pytań), odpowiadaj (na zmieniającą się sytuację badania), dostosowuj się (pytania badawcze i metody wraz ze zmieniającą się sytuacją i pogłębianym rozumieniem programu). Por. Patton M.Q., „Utilization-focused evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, Sage Publications, Thousand Oaks–London 2005, s. 431.

<sup>36</sup> Guba E.G., Lincoln Y.S., *Fourth Generation Evaluation*, Sage Publications, Newbury Park 1989, s. 185.

<sup>37</sup> Stufflebeam D.L., „Empowerment evaluation, objectivist evaluation, and evaluation standards: where the future of evaluation should not go and where it needs to go”, *Evaluation Practice* 1994, nr 15, s. 321–338.

by ewaluacja nie pomijała kwestii kontrowersyjnych czy niewygodnych dla głównych interesariuszy, ugruntowując jedynie *status quo*<sup>38</sup>. Osia krytyki jest również to, że ewaluacja, zamiast być obiektywnym procesem oceny, staje się animowaną, wielogłosową dyskusją nad programem (w praktyce zdominowaną przez głównych decydentów).

**Podejście realistyczne** (*realistic/realist evaluation*)<sup>39</sup> ma swoje korzenie w paradygmacie realizmu, a co za tym idzie, skupia się na kontekście i procesach interwencji – w jakich warunkach, w jaki sposób i w stosunku do jakich grup dana interwencja działa lub nie<sup>40</sup>. Rdzeniem badania jest skonstruowanie, a następnie przetestowanie teoretycznego modelu działania interwencji<sup>41</sup>. Procedura badawcza składa się z czterech etapów. Punktem wyjścia stanowi zidentyfikowanie mechanizmu interwencji – teorii lub założeń, którymi się kierowano (czyli założonej przez operatorów relacji przyczynowo-skutkowej), i stworzenie wstępnej hipotezy działania interwencji. Krokiem drugim jest opis kontekstu działań, w tym zarówno systemu relacji społecznych i interpersonalnych, jak i instytucjonalnych. Etap trzeci to zidentyfikowanie efektów, czyli zamierzonych i niezamierzonych konsekwencji działań. Wreszcie, ostatnim krokiem jest stworzenie modelu wiążącego trzy wcześniejsze elementy i przeanalizowanie relacji między nimi, zróżnicowania efektów pod kątem weryfikacji wstępnej hipotezy działania interwencji<sup>42</sup>. Ewaluator w tym podejściu pełni rolę analityka odtwarzającego logikę relacji. Podejście realistyczne nie określa preferowanych metod – mogą być one zarówno jakościowe, jak i ilościowe – choć niektórzy z autorów preferują metody ilościowe i *quasi*-eksperymenty<sup>43</sup>. Metody dobiera się w zależności od skali badanych działań, charakteru relacji itd. Z reguły ewaluatorzy korzystają z modeli i map logicznych, które są następnie testowane i weryfikowane

<sup>38</sup> Chelimsky E., „The coming transformation...”, op.cit., s. 17.

<sup>39</sup> Mathison (*Encyclopedia of Evaluation*, Sage Publications, Thousand Oaks–London 2005, s. 359–366) odróżnia *realist evaluation* od *realistic evaluation*. Wynika to raczej z podkreślenia wkładu dwóch różnych grup autorów (G.T. Henry i M.M. Mark ze Stanów Zjednoczonych vs. R. Pawson i N. Tilley z Wielkiej Brytanii), nie zaś z samego *meritum* podejścia. Bardzo zbliżonym podejściem jest również *theory-based evaluation*; por. Rossi P.H., Freeman H.E., Lipsey M.W., *Evaluation: A Systematic Approach. 6th edition*, Sage Publications, Thousand Oaks–London 1999; Chen H.-T., *Theory-Driven Evaluations*, Sage Publications, Newbury Park 1990.

<sup>40</sup> Henry G.T., „Realist evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 359–362, Sage Publications, Thousand Oaks–London 2005; Pawson R., Tilley N., „Realistic evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 362–367, Sage Publications, Thousand Oaks–London 2005.

<sup>41</sup> Chen H.-T., *Theory-Driven Evaluations*, op.cit.; Chen H.-T., *Theory-Driven Evaluations*, w: S. Mathison (red.), *Encyclopedia of evaluation*, s. 415–419, Sage Publications, Thousand Oaks–London 2005.

<sup>42</sup> Te cztery kroki badawcze ujmowane są w modelu CMOC (*Context–Mechanism–Outcomes–Pattern Configuration*). Por. Pawson R., Tilley N., „Realistic evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 362–367, Sage Publications, Thousand Oaks–London 2005, s. 365.

<sup>43</sup> Chen H.-T., Rossi P.H., „Evaluating with sense: the theory-driven approach”, *Evaluation Review* 1983, nr 7, s. 283–302.

w oparciu o badania *quasi*-eksperymentalne, wywiady, panele dyskusyjne czy przegląd literatury<sup>44</sup>.

Zaletą tego podejścia jest przystawanie do wieloaspektowej rzeczywistości dużych interwencji i programów publicznych. Również stosowanie map logicznych stanowi narzędzie bardzo komunikatywne. Natomiast za wadę czy też ograniczenie można uznać silne oparcie na wiedzy i doświadczeniu samych ewaluatorów (ich rozeznanie w dotychczasowych badaniach, weryfikacja modelu polityki). Co więcej, w przypadku krótszych ewaluacji problematyczne może okazać się zidentyfikowanie i ocena kontekstu, relacji, interakcji programu, wskutek czego odtworzona logika interwencji byłaby ułomna.

Odnosząc powyższe rozważania do celu i funkcji ewaluacji *ex-post*, należy uznać, że teoretycznie najbardziej pożądane wydaje się zastosowanie eksperymentalnych i *quasi*-eksperymentalnych podejść badawczych, gdyż jasno weryfikują one relację przyczynowo-skutkową interwencji, a co za tym idzie, pozwalają określić efekt netto. W praktyce interwencji UE, szczególnie zaś na poziomie programu, są one trudne do zastosowania ze względu na zróżnicowane, wieloaspektowe cele, prowadzone równocześnie inne interwencje (i związane z tym problem kontrolowania zmiennych), słabość danych statystycznych. Podejścia te nie wyjaśniają też, jak już zaznaczono, procesów stojących za efektami – mają więc ograniczone zastosowanie praktyczne w reformach i udoskonalaniu projektów czy programów. Nie należy również zapominać, że w ewaluacji analiza jest oddzielona od oceny. Ocena uwzględnia nie tylko kwestie obiektywne naukowo i stopień wypełnienia zakładanych celów, lecz także efekty uboczne oraz ich postrzeganie przez danych interesariuszy. W związku z tym wydaje się, że w praktyce badań *ex-post* interwencji UE szersze możliwości zastosowania ma podejście realistyczne, korzystające zarówno z metod ilościowych, jak i jakościowych, *quasi*-eksperymentalnych planów badawczych i perspektyw interesariuszy czy głównych ekspertów.

### Uczestnicy badania i ich role

Współczesna literatura tematu podkreśla, że badanie ewaluacyjne jest procesem „produkcji wiedzy”, systemowego uczenia się jednostek, zespołów, organizacji<sup>45</sup>. Proces ten składa się z szeregu interakcji, a często, w odróżnieniu od czystych badań naukowych, także negocjacji. Niektórzy autorzy zaznaczają nawet, iż prowadzone badanie wchodzi w układ zależności wewnętrznej

<sup>44</sup> Henry G.T., Julnes G., Mark M.M., „Realist evaluation: an emerging theory in support of practice”, *New Directions for Evaluation* 1998, nr 78; Pawson R., *Evidence-Based Policy: A Realist Perspective*, Sage Publications, London 2006; Pawson R., Tilley N., „Realistic evaluation”, Sage Publications, London 1997.

<sup>45</sup> Leviton L.C., „Evaluation use: advances, challenges and applications”, *American Journal of Evaluation* 2003, nr 24, s. 525–535; Preskill H., Torres R., „The learning dimension of evaluation use”, *New Directions for Evaluation* 2000, nr 88, s. 25–37.

i międzyorganizacyjnych, a w rezultacie samo tworzy mikrosystem społeczny: mikrosystem interesów, interakcji, osób<sup>46</sup>.

Uczestnicy badania, czyli ewaluatorzy, zlecający oraz szersza grupa interesariuszy, przyjmują w tym procesie (świadomie lub nie) różne role. Tradycyjnie analizy koncentrują się na strategiach ewaluatora jako strony determinującej badanie. Morabito<sup>47</sup> podaje cztery możliwe strategie (edukatora, konsultanta, animatora i doradcy), wskazując na „klimat zaufania” jako ich kluczowy element. Inną ciekawą typologię proponuje Valovirta, opisując rolę ewaluatora poprzez pryzmat dwóch wymiarów kontekstowych (konsensus–konfliktu wokół badanej interwencji; presji do zmian w interwencji lub jej braku). W rezultacie wyróżnia cztery możliwe role ewaluatora<sup>48</sup>: mediator (konflikt i brak presji zmian), arbiter rozstrzygający konflikt (konflikt i silna presja zmian), ekspert budzący świadomość (konsensus i brak potrzeby zmian), reformator (konsensus i potrzeba zmian).

Adaptując powyższe podziały do realiów interwencji europejskich, można wyróżnić następujące role ewaluatorów:

- Naukowiec – unika interakcji, a zlecających i innych interesariuszy traktuje wyłącznie jako przedmiot badania. Ewaluator bada i ocenia program *ex cathedra* – z perspektywy naukowej, bazując wyłącznie na „obiektywnych” danych. W tej roli ewaluator ma tendencje do odnoszenia interwencji, a zwłaszcza jej efektów, do szerszych teorii i procesów. Akcent w badaniu jest stawiany na warsztat naukowy. W przypadku rozważania kontekstu programu (konfliktów, presji zmian wokół programu) taka rola z uwagi na posiadany autorytet dobrze współgra z rolą arbitra lub eksperta. Mała elastyczność ogranicza jednak wykorzystywanie wniosków w praktyce.
- Konsultant – ogranicza się on do wytycznych kontraktowych (dla niego klientem, głównym adresatem i punktem odniesienia podejmowanych działań jest więc zlecający). Metodologia zależy od oczekiwań klienta, z tym, że główny akcent jest stawiany na koszty metod w porównaniu do ich efektów i szybkości użycia, a nie na ich naukową poprawność i stopień skomplikowania. Gdy trzeba uwzględnić kontekst programu (konflikt, presja zmian) ta rola dobrze łączy się z rolą mediatora i eksperta (w zależności od posiadanych kompetencji i doświadczenia).
- Animator–doradca – prowadzi głównie badania interaktywne, na bieżąco wspierając użytkowników w ich technicznych działaniach, organizując pole dialogu. Uwzględnia punkty widzenia różnych uczestników interwencji (zlecającego i interesariuszy). Ma jednak tendencje do koncentrowania się na procesach i kwestiach bieżących, a ogranicza szersze oceny efek-

<sup>46</sup> Johnson B., „Towards a theoretical model of evaluation utilization”, *Evaluation and Program Planning* 1998, nr 21, s. 93–110.

<sup>47</sup> Morabito S.M., „Evaluator roles and strategies for expanding evaluation process influence”, *American Journal of Evaluation* 2002, nr 23, s. 321–330.

<sup>48</sup> Valovirta V., „Evaluation utilization as argumentation”, *Evaluation* 2002, nr 8, s. 75.

tów interwencji (szczególnie długofalowych) do wniosków bezpośrednich uczestników. W kontekście konfliktów i presji wokół programu ta rola dobrze łączy się z rolą mediatora i reformatora.

- Edukator – ewaluator, który prócz cech opisanych w poprzednim punkcie (animatore–doradca) posiada również duży autorytet naukowy i wiedzę szerszą od operatorów danej interwencji. Dzięki temu wprowadza nowe punkty widzenia, spojrzenie na sprawy wykraczające poza bieżące zarządzanie i poza program. Podobnie jak animator, włącza uczestników interwencji w ocenę jej efektów, jednak zasila dyskusję szerszymi naukowymi analizami (prosząc o interpretację i sugerując kierunek dyskusji). W kontekście konfliktów i presji wokół programu ta rola łączy się dobrze z rolą mediatora i reformatora, gdyż ewaluator ma dobre narzędzia budowy zaufania i dialogu oraz autorytet.

Przedstawianie ewaluacji jako procesu uczenia się rodzi również pytanie o rolę interesariuszy w tym procesie. Prowadzone w literaturze analizy dotyczą zarówno kontekstu nauki, kwestii instytucjonalnych odbiorców wiedzy, jak i ich personalnej charakterystyki. Mocno upraszczając, można stwierdzić (co w zasadzie jest wnioskiem trywialnym), że im poważniej włącza się interesariuszy w badanie, tym lepiej się uczą, przyswajają nową perspektywę, tym bardziej otwarci są na zmiany, krytyczną ocenę<sup>49</sup>. Ponadto Boyle i Lemaire<sup>50</sup> zwracają uwagę na cechy odbiorców ewaluacji, którzy, aby efektywnie się uczyć, powinni być przygotowani do odbioru i użycia wiedzy (być otwartymi, mieć świadomość, czym jest ewaluacja, mieć sprecyzowane oczekiwania). Johnson<sup>51</sup> zauważa dwie kwestie. Po pierwsze, uczestnictwo jest bardziej intensywne w organizacjach o strukturze organicznej niż w tych o strukturze hierarchicznej – mechanistycznej. Po drugie, uczestnictwo jest częstsze w przypadku osób „uczących się”, otwartych w porównaniu do osób o typowo biurokratycznym usposobieniu.

Jeśli odnieść kwestię ról i interakcji do ewaluacji *ex-post*, wydaje się, że najbardziej wskazaną jest rola ewaluatora–edukatora, gdyż włącza efekty badań naukowych w dialog społeczny i wieloaspektową rzeczywistość interwencji. Wymaga ona jednak wysokich kompetencji, zdolności komunikacyjnych (przystępne przekazanie podejścia badawczego, nawiązanie kontaktu z odbiorcami, animacja dyskusji wokół oceny i interpretacji wniosków) oraz zbudowania zaufania. Z drugiej strony, od zlecającego wymaga to otwartości

---

<sup>49</sup> Alkin M.C., „Utilization of evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 143, 434–436, Sage Publications, Thousand Oaks–London 2005; Greene J.C., „Understanding...”, op.cit.; Mark M.M., Henry G.T., „The mechanisms and outcomes of evaluation influence”, *Evaluation* 2004, nr 10, s. 35–57; Patton M.Q., *Utilization-Focused Evaluation: The New Century Text*, op.cit.; Weiss C.H., „Have we learned anything new about the use of evaluation?”, *American Journal of Evaluation* 1998, nr 19, s. 21–33.

<sup>50</sup> Boyle R., Lemaire D., *Building Effective Evaluation Capacity: Lessons from Practice*, Transaction Publishers, New Brunswick 1998; za: Stern E., *What do we know...*, op.cit.

<sup>51</sup> Johnson B., „Towards a theoretical model...”, op.cit.

(w tym otwartości na krytykę) oraz dłuższego czasu kontraktu, od innych interesariuszy zaś – wysokiej aktywności (czy raczej responsywności, bo przecież kontakty nawiązuje ewaluator).

Interakcje uczestników badań w praktyce UE są złożone; zależą od poziomu, na którym wykonywane jest badanie, źródła finansowania interwencji oraz przedmiotu badania (tzw. ewaluandy, ang. *evaluand*), wielkości interwencji publicznej (projekt, działanie, program) i jej tematyki. Upraszczając, możemy wyróżnić trzy płaszczyzny<sup>52</sup>:

- płaszczyzna instytucji UE,
- płaszczyzna programów krajowych lub unijnych realizowanych na poziomie krajowym,
- płaszczyzna projektów sektora publicznego.

Płaszczyzna pierwsza to instytucje UE. Zapisy Rozporządzenia Finansowego, regulującego zasady całego budżetu UE<sup>53</sup>, nakładają obowiązek poddania ewaluacji każdego działania, które ma wpływ na budżet UE. W rezultacie każda z Dyrekcji Komisji w ramach każdej z polityk zleca i nadzoruje badania ewaluacyjne. Od 1996 roku Komisja zleciła ponad 1 500 takich badań. Ewaluatorami są konsorcja badawcze złożone z firm doradczych i instytutów badawczych, zwykle pełniące rolę konsultantów.

Płaszczyzna druga to programy realizowane na poziomie krajowym. Chodzi tu przede wszystkim o fundusze strukturalne. W tym przypadku obowiązek ewaluacji nakładają Rozporządzenia dotyczące funduszy strukturalnych. Według nich ewaluacje *ex-post* całych celów leżą w gestii Komisji Europejskiej przy współpracy z państwem członkowskim i Władzą Zarządzającą danym programem. W praktyce w ewaluację *ex-post* od strony zlecenia i nadzoru nad badaniem włączone są instytucje nadzorujące program (zwykle Ministerstwo<sup>54</sup>). W przypadku programów *stricte* krajowych (finansowanych wyłącznie ze źródeł krajowych) zlecającym jest instytucja publiczna kierująca programem. Praktyka ewaluacji środków krajowych jest w Polsce dość ograniczona. Jak na razie, główny przykład takich ewaluacji stanowią badania zlecane przez Polską Agencję Rozwoju Przedsiębiorczości<sup>55</sup>.

<sup>52</sup> Europejskie praktyki ewaluacyjne na pierwszych trzech z tych płaszczyzn zostały szerzej opisane w: Bachtler J., Poverari L., Taylor S., *Culture, Commitment and Capacity. A comparative review of Structural Funds Evaluation in the European Union*, Edinburgh Conference on Evaluation, Edinburgh 2000; Bachtler J., *The Evaluation of Regional Policy in the European Union – Main Report*, European Policies Research Centre, Glasgow 2000; Toulemonde J., „Evaluation culture(s) in Europe: differences and convergence between national practices”, *The Danish Evaluation Society Annual Conference*, 2000; Olejniczak K., „Teoria i praktyka ewaluacji...”, op.cit.

<sup>53</sup> Council of the European Union, *Council Regulation (EC, Euratom) No 1605/2002 of 25 June 2002 on the Financial Regulation applicable to the general budget of the European Communities*, Official Journal of the European Communities L248, 16.09.2002.

<sup>54</sup> Poszczególne departamenty odpowiedzialne za zarządzanie programem lub Krajowa Jednostka Oceny (komórka w Ministerstwie Rozwoju Regionalnego odpowiedzialna za koordynację badań ewaluacyjnych funduszy strukturalnych w Polsce).

<sup>55</sup> Zob. <http://www.parp.gov.pl/index/index/110>.

Perspektywy są jednak szersze – ewaluacje mogą zlecać inne agencje rządowe, Urzędy Marszałkowskie. Wykonawcy ewaluacji programów to firmy doradcze lub instytuty badawcze, szeroka zaś gama interesariuszy programu jest przedmiotem badania (wywiady, ankiety), a coraz częściej włącza się ją w proces oceny programu i jego efektów. Przy badaniach całych pakietów interwencji ewaluatorzy często przyjmują rolę naukowców (przy stosowaniu modeli ekonometrycznych), przy zróżnicowanych interwencjach – doradców, a nawet edukatorów (skala badań daje możliwość porównań europejskich, promocji najlepszych praktyk).

Wreszcie, płaszczyzna projektów wymaga ścisłych interakcji wykonawca–zlecający. Ponieważ skala badania jest dużo mniejsza, pojawia się możliwość badań całej populacji projektu, grup kontrolnych, badań porównawczych czy nawet *benchmarking* (porównania do najlepszych projektów krajowych, europejskich). Przy stworzeniu dobrej relacji zlecający–badający oraz zaangażowaniu i zainteresowaniu ze strony interesariuszy, ewaluator może występować w roli edukatora i reformatora.

## Wykorzystanie wyników ewaluacji

Na koniec należy wskazać kwestię, która obok rzetelności badania wydaje się być jedną z najbardziej kluczowych w ewaluacji – **kwestię wykorzystania badania**.

W literaturze tematu podawanych jest wiele czynników warunkujących wykorzystanie ewaluacji (np. kontekst polityczny, moment badania, cechy użytkowników itd.). Użycie czy wykorzystanie ewaluacji (*evaluation utilization, evaluation use*) może być rozpatrywane w przynajmniej czterech wymiarach:

1) tego, co jest używane (*source of use*) – z jednej strony mówimy o użyciu wniosków, rezultatów badania prezentowanych najczęściej w formie raportu ewaluacyjnego (*findings use*), z drugiej, o użyciu samego procesu badania – interakcji i dyskusji, które mają miejsce podczas prowadzenia badania między ewaluatorami, operatorami programu, instytucjami i grupami osób zaangażowanymi w program (*process use*)<sup>56</sup>;

2) przez kogo jest używane<sup>57</sup> – w perspektywie skali użycia (ewaluacja może być używana przez jednostki, grupy i/lub całe organizacje, instytucje) i w perspektywie stopnia powiązania użytkowników z badaną interwencją;

---

<sup>56</sup> Alkin M.C., Taut S.M., „Unbundling evaluation use”, *Studies in Educational Evaluation* 2003, nr 29, s. 6; Kirkhart K.E., „Reconceptualizing evaluation use: an integrated theory of influence”, *New Directions for Evaluation* 2000, nr 88, s. 5–25.

<sup>57</sup> Można zidentyfikować następujących użytkowników ewaluacji za C.H. Weissem („Evaluation for decisions: is anybody there? Does anybody care?”, w: E. Stern (red.), *Evaluation Research Methods*, Sage Publications, London 2005, 292–295):



3) w jaki sposób i z jaką świadomością jest używane (*way and intention*) – z pozytywnych rodzajów wyróżnia się tu użycie instrumentalne (użycie rekomendacji) i koncepcyjne (zmiany w myśleniu o programach), po nich jest zaś użycie neutralne (symboliczne, czyli użycie ewaluacji do uzasadnienia *de facto* podjętych już decyzji oraz uzasadniony brak użycia, gdy jakość ewaluacji jest niska i ewaluacja nic nie wnosi), typologię zamyka cały szereg błędnych, niewłaściwych form wykorzystania ewaluacji: pozorne użycie (*sham use*), które zachodzi, gdy badanie zostało potraktowane wyłącznie jako spełnienie formalnego wymogu administracyjnego; błędne użycie, gdy użytkownik, mając słabe rozeznanie w standardach nauk społecznych, bezkrytycznie używa w dobrej wierze słabego raportu ewaluacyjnego; nadużycie – manipulacja, zatajenie (*abuse, mischievous use*) – czyli manipulowanie raportem i wynikami badania przez jego odbiorców, prezentowanie wyników w sposób wybiórczy lub utajnienie całego raportu<sup>58</sup>.

4) w jakiej perspektywie czasowej jest używane (*time*) – efekty ewaluacji mogą nastąpić bezpośrednio po zakończeniu ewaluacji i wpłynąć na badaną interwencję lub też pojawić się z opóźnieniem, oddziałując na kolejne edycje danego programu/projektu lub przyszłe inicjatywy o analogicznym charakterze<sup>59</sup>.

- personel i zarządzający daną interwencją,
- interesariusze – aplikujący o projekty, beneficjenci, beneficjenci końcowi,
- szersza publika europejska – personel i zarządzający analogicznymi interwencjami,
- grupy interesu (*interest groups*) – stowarzyszenia, związki, organizacje firm, klientów, regionów, gmin związane z tematyką danej interwencji,
- sieci tematyczne danej kwestii (*issue networks*) – eksperci, osoby, instytucje badawcze, *think tanks* znające daną tematykę, śledzące jej ewolucję, jak też zainteresowane wpływaniem na kształt danej polityki,
- główni interesariusze danej polityki – instytucje tworzące daną politykę, zestawy interwencji, a właściwie personel tychże instytucji, odpowiednich dyrekcji Komisji Europejskiej, krajowych ministerstw (decydenci wyższej rangi i personel przygotowujący sprawozdania),
- mass-media – dziennikarze prasowi i telewizyjni, internetowe serwisy informacyjne,
- szeroka opinia publiczna.

<sup>58</sup> Szerzej zob.: Alkin M.C., Coyle K., „Thoughts on evaluation utilization, misutilization and non-utilization”, *Studies in Educational Evaluation* 1988, nr 14, s. 331–340; Cousins B.J., „Commentary: Minimizing evaluation misuse as principled practice”, *American Journal of Evaluation* 2004, nr 25, s. 391–397; Alkin M.C., Taut S.M., „Unbundling...”, op.cit., s. 5; Leviton L.C., Hughes E.F.X., „Research on the utilization of evaluations: a review and synthesis”, *Evaluation Review* 1981, nr 5, s. 525–548; Mark M.M., Henry G.T., „The mechanisms...”, op.cit.; Rich R.F., „Uses of social science information by federal bureaucrats: knowledge for action versus knowledge for understanding”, w: C.H. Weiss (red.), *Using Social Research in Public Policy Making*, s. 199–211, Lexington Books, Lexington 1977; Shulha L.M., Cousins B.J., „Evaluation use: theory, research, and practice since 1986”, *Evaluation Practice* 1997, nr 18, s. 195–208; Weiss C.H., Murphy-Graham E., Birkeland S., „An alternate route to policy influence: how evaluations affect D.A.R.E.”, *American Journal of Evaluation* 2005, nr 26, s. 13–14.

<sup>59</sup> Alkin M.C., „Utilization of evaluation”, op.cit.; Shulha L.M., Cousins B.J., „Evaluation use...”, op.cit.

Odnosząc zaprezentowane typologie do ewaluacji *ex-post* należy stwierdzić, że:

- najczęściej używane są wnioski badania; sam proces w niewielkim stopniu może pomóc w zmianie interwencji, gdyż ta jest już zakończona; pewne ograniczone użycie procesowe może mieć miejsce, gdy interwencja jest kontynuowana (wtedy np. używa się przy kolejnym działaniu stworzonych w czasie ewaluacji sieci kontaktów i konsultacji);
- użycie instrumentalne (np. natychmiastowe zatrzymanie lub zmiana strategii w kolejnej edycji projektu) jest możliwe raczej przy badaniach interwencji o małej skali (np. projektów), gdy cykl wdrażania jest krótszy, a sieć instytucji i procedur mniejsza;
- w większej skali, w rzeczywistości programów funduszy strukturalnych (gdy zasady funkcjonowania programu zmieniają się) ewaluacja *ex-post* rzadko dostarcza technicznych rekomendacji; tak, jak już wyjaśniano, koncentruje się ona na zmianie paradygmatu polityki/interwencji;
- ewaluacja *ex-post* najczęściej jest używana koncepcyjnie (wdrażający program/projekt zdają sobie sprawę z jej rzeczywistych efektów, są w stanie zweryfikować zasadność swojego myślenia o przyczynach skutkach interwencji), użycie takie trudno jednak zidentyfikować;
- w przypadku ewaluacji programów efekt badania jest rzadko natychmiastowy – zmiana paradygmatów, sposobu myślenia wymaga czasu i ma charakter kumulatywny (kolejne badania dostarczają dowodów i orientacja programów jest powoli zmieniana); zmiany w myśleniu i ich efekt (zmiany w celach i podejściu do interwencji) są dużo szybsze na poziomie ewaluacji projektów;
- badania *ex-post* są wyjątkowo narażone na błędne użycie, gdyż poprawna analiza relacji przyczynowo-skutkowych (oraz efektu netto) w skomplikowanych układach społecznych jest trudna i często niejednoznaczna, prowadzona z użyciem skomplikowanych metod; wymaga to uwagi ze strony zlecającego oraz wysokich kompetencji ze strony wykonawcy;
- ponadto, badania *ex-post* z uwagi na „rozliczeniowy charakter” (pokazanie sukcesu lub porażki interwencji) są często narażone na nadużycie i zatajenie (unikanie publikacji niewygodnych – z perspektywy operatorów interwencji – wniosków, świadczących np. o porażce programu, jego niedopasowaniu, ograniczonej racjonalności); by temu zapobiec, trzeba ustalać już na początku badania szeroką formę publikacji wyników;
- wreszcie, odbiorcy badania stanowią liczną grupę, co zależy głównie od wielkości badanej interwencji; wyniki ewaluacji polityk i programów są uważnie śledzone i wykorzystywane do dyskusji, jak też do dalszych badań na poziomie europejskim, przez całe spektrum aktorów (Komisję, ministerstwa, grupy interesu, ekspertów, mass-media itd.); wynikami badań projektów interesują się najczęściej instytucje wdrażające dane działania, lokalni interesariusze, bezpośredni beneficjenci.

## Podsumowanie

Ewaluacja *ex-post* to badanie społeczno-ekonomiczne, wykonane po zakończeniu interwencji publicznej (programu, działania, projektu itd.), oceniające jej rzeczywistą jakość oraz wartość w odniesieniu do potrzeb i problemów, które miała ona rozwiązać.

Ewaluacje *ex-post* pozwalają odpowiedzieć na pytania o stopień osiągnięcia celów interwencji, jej sukcesy i ograniczenia, wpływ czynników zewnętrznych, odpowiedniość użytych instrumentów, skuteczność (*post factum*) systemu wdrażania, rolę poszczególnych aktorów, rzeczywistą użyteczność – efekty oddziaływania pomocy (zarówno te zakładane, jak i te nieprzewidziane), a także trwałość tychże efektów.

W praktyce sektora publicznego ewaluacje *ex-post* spełniają przede wszystkim dwie funkcje – wzmacniają odpowiedzialność (pokazują racjonalność wykorzystania pieniędzy publicznych w podnoszeniu dobrobytu społecznego) oraz wspierają procesy uczenia się (czyli rozumienie procesów społeczno-gospodarczych i identyfikowanie metod pobudzania rozwoju).

Z uwagi na to, że rdzeniem ewaluacji *ex-post* jest zbadanie relacji przyczynowo-skutkowej (tzn. czy dane pieniądze miały wpływ na zidentyfikowane szanse, problemy, grupy docelowe) i, w idealnej sytuacji, określenie efektu netto, w badaniach preferuje się metody *quasi*-eksperymentalne. Jednak w skomplikowanej rzeczywistości interwencji społeczno-gospodarczych (mnogość celów, interesów, działań, zróżnicowanie i skala grup docelowych) taki plan badawczy może być stosowany punktowo i bardziej sprawdza się podejście realistyczne. Ciekawe, lecz wymagające zdecydowanie dłuższych badań, wydaje się również stosowanie podejścia wyłączającego cele programu – pozwala skoncentrować badanie na poszukiwaniu rzeczywistych efektów interwencji.

Najbardziej odpowiednią rolą dla ewaluatora jest chyba edukator – osoba o dużej wiedzy merytorycznej, wspierająca się na badaniach (w tym badaniach europejskich), ale nie stroniąca od używania technik interaktywnych, dialogu, wprowadzająca nowe perspektywy postrzegania interwencji, jej racjonalności, działań jej uczestników. Edukator rozszerza perspektywę patrzenia na program, projekt, na strategię działań, użyteczność z perspektywy nie tylko bieżących zysków, lecz także długookresowej konkurencyjności. Rola ta jest szczególnie ważna na poziomie ewaluacji projektów i programów krajowych, zbyt często ograniczanych perspektywą operatorów, doraźnych procedur i mechanicznego wykorzystania środków. Należy jednak pamiętać, że takie podejście może dać efekty tylko przy zaangażowaniu i otwartości drugiej strony – zlecających i interesariuszy interwencji.

Wreszcie, odnosząc się do kwestii wykorzystania badań *ex-post*, należy stwierdzić, że pomagają one formować kierunki myślenia o interwencjach publicznych, ich logice i zasadności, a w mniejszym stopniu wpływają na

bieżące, operacyjne działania. Stawiają jednak kluczowe pytanie o sens podjętych interwencji, ich realną, długookresową, a nie doraźną, użyteczność – i odpowiadają na nie. Wnioski z ewaluacji *ex-post* mają charakter kumulatywny (dowody z kilku badań sprawiają, że zmienia się np. myślenie o interwencjach na rynku pracy) i działają w przesunięciu czasowym. Dają więc postawy myślenia strategicznego – pozwalają gromadzić dowody na temat słuszności paradygmatów polityk, myślenia o rozwoju regionalnym, zasobach ludzkich i gospodarce w perspektywie wieloletniej.

---

**Karol Olejniczak** jest absolwentem studiów europejskiej administracji publicznej (European Institute of Public Administration, Holandia) oraz studiów regionalnych (EUROREG – Uniwersytet Warszawski). Obecnie, pod kierunkiem Prof. Grzegorza Gorzelaka, przygotowuje pracę doktorską z zakresu wykorzystania ewaluacji w programach sektora publicznego (na wydziale zarządzania WSPiZ im. L. Koźmińskiego). Badania akademickie nad ewaluacją prowadził w USA, Wielkiej Brytanii i Irlandii. Autor ma również szerokie doświadczenie praktyczne – realizował projekty ewaluacyjne zarówno w kraju, jak i za granicą (Austria, Holandia, Wielka Brytania). Prowadzi wykłady i warsztaty z zakresu ewaluacji interwencji publicznych, jest autorem publikacji na temat ewaluacji, był polskim ekspertem biorącym udział w przygotowaniu podręcznika *Evalsed*. Pracuje jako asystent w Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG) na Uniwersytecie Warszawskim.

## Bibliografia

- Alkin M.C., *Evaluation Roots: Tracing Theorists' Views and Influences*, Sage Publications, Thousand Oaks–London 2004.
- Alkin M.C., „Utilization of evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 143, 434–436, Sage Publications, Thousand Oaks–London 2005.
- Alkin M.C., Coyle K., „Thoughts on evaluation utilization, misutilization and non-utilization”, *Studies in Educational Evaluation* 1988, nr 14, s. 331–340.
- Alkin M.C., Taut S.M., „Unbundling evaluation use”, *Studies in Educational Evaluation* 2003, nr 29, s. 1–12.
- Bachtler J., *The Evaluation of Regional Policy in the European Union – Main Report*, European Policies Research Centre, Glasgow 2000.
- Bachtler J., „Quod erat demonstrandum? Ewaluacja polityki regionalnej”, *Studia Regionalne i Lokalne* 2001, nr 4(7).
- Bachtler J., Poverari L., Taylor S., *Culture, Commitment and Capacity. A comparative review of Structural Funds Evaluation in the European Union*, Edinburgh Conference on Evaluation, Edinburgh 2000.
- Bachtler J., Taylor S., Olejniczak K., *A Study of the Mid Term Evaluations of INTERREG programmes for the programming period 2000 until 2006*, INTERACT Programme Secretariat, Vienna 2005.
- Balthasar A., „The effects of institutional design on the utilization of evaluation. Evidenced using qualitative comparative analysis (QCA)”, *Evaluation* 2006, nr 12, s. 353–371.

- Batterbury S., „Principles and purposes of European Union Cohesion policy evaluation”, *Regional Studies* 2006, nr 40(2), s.179–188.
- Berk R.A., Rossi P.H., *Thinking about Program Evaluation*, Sage Publications, Thousand Oaks–London 1999.
- Bingham R.D., Felbinger C.L., *Evaluation in Practice: A Methodological Approach*, Seven Bridges Press, New York–London 2002.
- Boyle R., Lemaire D., *Building Effective Evaluation Capacity: Lessons from Practice*, Transaction Publishers, New Brunswick 1998.
- Bryman A., *Social Research Methods*, Oxford University Press, Oxford 2004.
- Campbell D.T., Stanley J.C., *Experimental and Quasi-Experimental Designs for Research*, Rand McNally College Publishing Company, Chicago 1967.
- Chelimsky E., *Program Evaluation: Patterns and Directions*, American Society for Public Administration, Washington 1985.
- Chelimsky E., „The coming transformation in evaluation”, w: E. Chelimsky, W.R. Shadish (red.), *Evaluation for the 21st Century: A Handbook*, Sage Publications, Thousand Oaks–London 1997.
- Chen H.-T., *Theory-Driven Evaluations*, Sage Publications, Newbury Park 1990.
- Chen H.-T., *Theory-Driven Evaluations*, w: S. Mathison (red.), *Encyclopedia of evaluation*, s. 415–419, Sage Publications, Thousand Oaks–London 2005.
- Chen H.-T., Rossi P.H., „Evaluating with sense: the theory-driven approach”, *Evaluation Review* 1983, nr 7, s. 283–302.
- Council of the European Union, *Council Regulation (EC) No 1260/1999 of 21 June 1999 laying down General Provisions on the Structural Funds*, Official Journal of the European Communities L161, 26.6.1999.
- Council of the European Union, *Council Regulation (EC, Euratom) No 1605/2002 of 25 June 2002 on the Financial Regulation applicable to the general budget of the European Communities*, Official Journal of the European Communities L248, 16.09.2002.
- Council of the European Union, *Council Regulation (EC) No 1083/2006 of 11 July 2006 laying down on the European Regional Development Fund, the European Social Fund and the Cohesion Fund*, Official Journal of the European Communities L210, 31.7.2006.
- Cousins B.J., „Utilization effects of participatory evaluation”, w: T. Kellaghan, D.L. Stufflebeam, A. Wingate (red.), *International Handbook of Educational Evaluation*, vol. 2, s. 245–266, Kluwer Academic, Dordrecht 2003.
- Cousins B.J., „Commentary: Minimizing evaluation misuse as principled practice”, *American Journal of Evaluation* 2004, nr 25, s. 391–397.
- Cousins B.J., Leithwood K.A., „Current empirical research in evaluation utilization”, *Review of Educational Research* 1986, nr 56, s. 331–364.
- Cousins J.B., Whitmore E., „Framing participatory evaluation”, *New Directions for Evaluation* 1998, nr 80, s. 5–23.
- Cronbach L.J. i in., *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*, Jossey-Bass, San Francisco 1981.
- Dahler-Larsen P., „Evaluation in public management”, w: E. Ferlie, L.E. Lynn, C. Pollitt (red.), *The Oxford Handbook of Public Management*, s. 615–639, Oxford University Press, Oxford 2005.
- Davidson E. J., *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*, Sage Publications, Thousand Oaks 2005.
- EPEC, *Study on the Use of Evaluation Results in the Commission. Final report*, European Policy Evaluation Consortium (EPEC), Paris 2005.
- Eser T.W., Nussmueller E., „Mid-term evaluations of Community Initiatives under European Union Structural Funds: a process between accounting and common learning”, *Regional Studies* 2006, nr 40(2), s. 249–258.

- European Commission, *MEANS Handbook 1, Organising Intermediate Evaluation in the Context of Partnership*, XVI D., Commission of the European Communities, Brussels 1995.
- European Commission, *MEANS Handbook 2, Identifying the Structuring Effects of Community Interventions*, XVI D., Commission of the European Communities, Brussels 1995.
- European Commission, *MEANS Handbook 3, Measuring the Employment Effects of Community Structural Interventions*, XVI D., Commission of the European Communities, Brussels 1995.
- European Commission, *MEANS Handbook 4, Applying the Multi-Criteria Method to the Evaluation of Structural Programmes*, XVI D., Commission of the European Communities, Brussels 1995.
- European Commission, *MEANS Handbook 5, Quality Assessment of Evaluation Reports: A Framework*, XVI D., Commission of the European Communities, Brussels 1995.
- European Commission, *MEANS Handbook 6, The Ex-post Evaluation of Structural Funds Programmes: A Methodological Framework*, XVI D., Commission of the European Communities, Brussels 1995.
- European Commission, *MEANS Handbook 7, The Ex-ante Evaluation of Structural Funds Programmes: A Methodological Framework*, XVI D., Commission of the European Communities, Brussels 1995.
- European Commission, *MEANS Collection: Evaluating Socio-economic Programmes*, DG Regio, Office for Official Publications of the European Communities, Luxembourg 1999.
- European Commission, *Focus on Results: Strengthening Evaluation of Commission Activities*, SEC2000/1051, European Commission 2000.
- European Commission, *Implementing Activity-Based Management in the Commission*, SEC-(2001)1197/6 & 7, European Commission 2000.
- European Commission, *Putting Strategic Evaluation into Practice within Commission*, C2001/3661, European Commission 2001.
- European Commission, *White Paper on European Governance – Report of the Working Group Evaluation & Transparency*, Working Group 2B, European Commission 2001, [http://europa.eu.int/comm/governance/index\\_en.htm](http://europa.eu.int/comm/governance/index_en.htm).
- European Commission, *Commission Regulation (EC, Euratom) No 2342/2002 of 23 December 2002 laying down detailed rules for the implementation of Council Regulation (EC, Euratom) No 1605/2002 on the Financial Regulation applicable to the general budget of the European Communities*, Official Journal of the European Communities L 357/1, European Commission, 31.12.2002.
- European Commission, *Evaluation Standards and Good Practice*, C2002/5267, European Commission 2002.
- European Commission, *Evaluating EU activities – A practical guide for the Commission services*, DG Budget, Office for Official Publications of the European Communities, Luxembourg 2004.
- Forunier D.M., „Evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 139–140, Sage Publications, Thousand Oaks–London 2005.
- Frankfort-Nachmias C., Nachmias D., *Metody badawcze w naukach społecznych*, przekł. E. Hornowska, Wydawnictwo Zysk i S-ka, Warszawa 2001.
- GHK Consultants, *The Evaluation of Socio-Economic Development. The Evalsed Guide (2nd edition)*, [www.evaled.com](http://www.evaled.com), 31.12.2006.
- Greene J.C., „Understanding social programs through evaluation”, w: N.K. Denzin, Y.S. Lincoln (red.), *Collecting and Interpreting Qualitative Materials*, s. 590–617, Sage Publications, Thousand Oaks–London 2003.
- Guba E.G., Lincoln Y.S., *Fourth Generation Evaluation*, Sage Publications, Newbury Park 1989.

- Heinrich C.J., „Measuring public sector performance”, w: G. Peters, J. Pierre (red.), *Handbook of Public Administration*, s. 25–37, Sage Publications, London 2003.
- Henry G.T., „Why not use?”, *New Directions for Evaluation* 2000, nr 88, s. 85–98.
- Henry G.T., „Realist evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 359–362, Sage Publications, Thousand Oaks–London 2005.
- Henry G.T., Julnes G., „Values and realist evaluation”, *New Directions for Evaluation* 1998, nr 78, s. 53–71.
- Henry G.T., Julnes G., Mark M.M., „Realist evaluation: an emerging theory in support of practice”, *New Directions for Evaluation* 1998, nr 78.
- Henry G.T., Mark M.M., „Beyond use: understanding evaluation’s influence on attitudes and actions”, *American Journal of Evaluation* 2003, nr 24, s. 293–314.
- Johnson B., „Towards a theoretical model of evaluation utilization”, *Evaluation and Program Planning* 1998, nr 21, s. 93–110.
- Kierzkowski T., *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2002.
- King J.A., „Participatory evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 291–294, Sage Publications, Thousand Oaks–London 2005.
- Kirkhart K.E., „Reconceptualizing evaluation use: an integrated theory of influence”, *New Directions for Evaluation* 2000, nr 88, s. 5–25.
- Langbein L., Felbinger C.L., *Public Program Evaluation. A Statistical Guide*, M.E. Shape, New York–London 2006.
- Leviton L.C., „Evaluation use: advances, challenges and applications”, *American Journal of Evaluation* 2003, nr 24, s. 525–535.
- Leviton L.C., Hughes E.F.X., „Research on the utilization of evaluations: a review and synthesis”, *Evaluation Review* 1981, nr 5, s. 525–548.
- Mark M.M., Henry G.T., „The mechanisms and outcomes of evaluation influence”, *Evaluation* 2004, nr 10, s. 35–57.
- Mark M.M., Henry G.T., Julnes G., *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*, Jossey-Bass, San Francisco 2000.
- Mathison S. (red.), *Encyclopedia of Evaluation*, Sage Publications, Thousand Oaks–London 2005.
- Morabito S.M., „Evaluator roles and strategies for expanding evaluation process influence”, *American Journal of Evaluation* 2002, nr 23, s. 321–330.
- Morris L.L., Freeman M.E., Fitz-Gibbon C.T., *How to Communicate Evaluation Findings*, Sage Publications, Newbury Park–London 1987.
- OECD, *Modernising Government: The Way Forward*, OECD, Paris 2005.
- Olejniczak K., „Teoria i praktyka ewaluacji w świetle doświadczeń europejskich”, *Studia Regionalne i Lokalne* 2005, nr 4(22).
- Olejniczak K., „Ewaluacja jako narzędzie zarządzania w sektorze publicznym”, w: G. Gorzelak, A. Tucholska (red.), *Rozwój, region, przestrzeń*, Ministerstwo Rozwoju Regionalnego, Warszawa 2007.
- Owen J., Lambert F., „Evaluation and the information needs of organizational leaders”, *American Journal of Evaluation* 1998, nr 19, s. 355–365.
- Owen J.M., *Program Evaluation: Forms and Approaches. 3rd edition*, The Guilford Press, London 2007.
- Oxford University Press, *Oxford English Dictionary*, Oxford University Press, Oxford 2002.
- Paddock S.C., (1998) „Evaluation”, w: J.M. Shafritz (red.), *International Encyclopedia of Public Policy and Administration*, s. 818–823, Westview Press, Boulder 1998.
- Patton M.Q., *Utilization-Focused Evaluation*, Sage Publications, Beverly Hills 1978.
- Patton M.Q., *Utilization-Focused Evaluation: The New Century Text*, Sage Publications, Thousand Oaks 1997.

- Patton M.Q., „Evaluation research”, w: M.S. Lewis-Beck, A. Bryman, T.F. Liao (red.), *The Sage Encyclopedia of Social Science Research Methods*, s. 337–340, Sage Publications, Thousand Oaks–2004.
- Patton M.Q., „Utilization-focused evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 429–432, Sage Publications, Thousand Oaks–London 2005.
- Pawson R., *Evidence-Based Policy: A Realist Perspective*, Sage Publications, London 2006.
- Pawson R., Tilley N., *Realistic evaluation*, Sage Publications, London 1997.
- Pawson R., Tilley N., „Realistic evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 362–367, Sage Publications, Thousand Oaks–London 2005.
- Preskill H., Torres R., „The learning dimension of evaluation use”, *New Directions for Evaluation* 2000, nr 88, s. 25–37.
- Prime Minister’s Strategy Unit, *Strategy Survival Guide*, UK Prime Minister’s Cabinet Office, London 2004, [www.strategy.gov.uk](http://www.strategy.gov.uk).
- Rich R.F., „Uses of social science information by federal bureaucrats: knowledge for action versus knowledge for understanding”, w: C.H. Weiss (red.), *Using Social Research in Public Policy Making*, s. 199–211, Lexington Books, Lexington 1977.
- Rist R.C., *Program Evaluation and the Management of Government: Patterns and Prospects across Eight Nations*, Transaction Publishers, New Brunswick–London 1990.
- Rossi P.H., Freeman H.E., Lipsey M.W., *Evaluation: A Systematic Approach. 6th edition*, Sage Publications, Thousand Oaks–London 1999.
- Rossi P.H., Lipsey M.W., Freeman H.E., *Evaluation: A Systematic Approach*, Sage Publications, Thousand Oaks–London 2004.
- Scriven M., „Pros and cons about goal-free evaluation”, w: G.V. Glass (red.), *Evaluation Studies Review Annual*, vol 1., Sage Publications, Beverly Hills 1976.
- Scriven M., *Evaluation Thesaurus*, Sage Publications, Newbury Park–London 1991.
- Shadish W.R., Cook T.D., Campbell D.T., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston 2001.
- Shadish W.R., Cook T.D., Leviton L.C., *Foundations of Program Evaluation: Theories of Practice*, Sage Publications, Newbury Park–London 1991.
- Shulha L.M., Cousins B.J., „Evaluation use: theory, research, and practice since 1986”, *Evaluation Practice* 1997, nr 18, s. 195–208.
- Smith P., „On the unintended consequences of publishing performance data in the public sector”, *International Journal of Public Administration* 1995, nr 18, s. 277–310.
- Spencer L., Ritchie J., Lewis J., Dillon L., *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*, Government Chief Social Researcher’s Office – Strategy Unit, London 2003.
- Stake R., „A modest commitment to the promotion of democracy”, *New Directions for Evaluation* 2000, nr 85, s. 97–107.
- Stake R., „Responsive evaluation”, w: S. Mathison (red.), *Encyclopedia of Evaluation*, s. 376–379, Sage Publications, Thousand Oaks–London 2005.
- Stern E., *Evaluation Research Methods*, Sage Publications, London 2005.
- Stern E., *What do we know about the utilization of evaluation?*, I konferencja ewaluacyjna – Ewaluacja programów o charakterze społeczno-gospodarczym finansowanych z funduszy strukturalnych, s. 7–29, Ministerstwo Gospodarki i Pracy, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 16 września 2005.
- Stewart J., Ranson S., „Management in the public domain”, w: D. McKevitt, A. Lawton (red.), *Public Sector Management: Theory, Critique and Practice*, s. 54–70, Open University, Sage Publications, London 1994.
- Stufflebeam D.L., „Empowerment evaluation, objectivist evaluation, and evaluation standards: where the future of evaluation should not go and where it needs to go”, *Evaluation Practice* 1994, nr 15, s. 321–338.
- Stufflebeam D.L., „Evaluation models”, *New Directions for Evaluation* 2001, nr 89, s. 7–92.



- Tavistock Institute, GHK, IRS, *The Evaluation of Socio-Economic Development. The Evalsed Guide (1st edition)*, www.evaled.com 2003, downloaded 21.08.2006.
- Torres R., Preskill H., „Evaluation and organizational learning: past, present and future”, *New Directions for Evaluation* 2001, nr 22, s. 387–395.
- Toulemonde J., „Evaluation culture(s) in Europe: differences and convergence between national practices”, *The Danish Evaluation Society Annual Conference*, 2000.
- US General Accounting Office, *Quantitative Data Analysis: An Introduction*, US General Accounting Office, Washington 1992.
- Valovirta V., „Evaluation utilization as argumentation”, *Evaluation* 2002, nr 8, s. 60–80.
- Weiss C.H., *Evaluation: Methods for Studying Programs and Policies*, Prentice Hall, Upper Saddle River 1998.
- Weiss C.H., „Have we learned anything new about the use of evaluation?”, *American Journal of Evaluation* 1998, nr 19, s. 21–33.
- Weiss C.H., „Evaluation for decisions: is anybody there? Does anybody care?”, w: E. Stern (red.), *Evaluation Research Methods*, s. 287–299, Sage Publications, London 2005.
- Weiss C.H., Bucuvalas M.J., *Social Science Research and Decision-Making*, Columbia University Press, New York–Guildford 1980.
- Weiss C.H., Bucuvalas M.J., „Truth tests and utility tests: decision-makers’ frame of reference for social science research”, *American Journal of Sociological Review* 1980, nr 45, s. 302–313.
- Weiss C.H., Murphy-Graham E., Birkeland S., „An alternate route to policy influence: how evaluations affect D.A.R.E.”, *American Journal of Evaluation* 2005, nr 26, s. 12–30.
- Wollmann H., *Evaluation in Public Sector Reform: Concepts and Practice in International Perspective*, Edward Elgar, Cheltenham 2003.



*Agnieszka Haber*

## **Proces ewaluacji *ex-post* – konceptualizacja, operacjonalizacja, realizacja badania ewaluacyjnego**

### **Wprowadzenie**

Ewaluacja jest obecnie bardzo modnym pojęciem, związanym głównie z zarządzaniem pomocą publiczną finansowaną z funduszy strukturalnych. Jej główne zadanie stanowi dostarczenie uzasadnienia dla podejmowanych decyzji strategicznych i operacyjnych<sup>1</sup>. Często postrzega się ją jako opisywanie i wartościowanie efektów danej polityki, interwencji, programu, ale także jako narzędzie definiowania problemów i ograniczeń konkretnej rzeczywistości oraz poszukiwania takich strategii i rozwiązań, które pomogłyby te problemy rozwiązać przy użyciu określonych instrumentów. Dlatego ze swej natury ewaluacja posiada wszystkie cechy badań stosowanych i, tym samym, nie różni się zasadniczo pod względem projektowania technik gromadzenia danych oraz metod analizy od innych przedsięwzięć badawczych<sup>2</sup>.

Niniejszy rozdział jest opisem ewaluacji właśnie jako procesu badawczego, mającego prowadzić do dochodzenia do obiektywnej, niezależnej oceny opartej na wiarygodnych danych empirycznych, pozyskanych zgodnie z zasadami wybranych metod – procesu, którego wykonanie wymaga rzetelnego przemyślenia, zaprojektowania i realizacji.

### **Cel ewaluacji *ex-post***

W odniesieniu do ewaluacji *ex-post* możemy mówić, iż stanowi ona systematyczną i obiektywną ocenę zakończonego projektu, programu lub polityki

---

<sup>1</sup> Zob. np. Ministerstwo Rozwoju Regionalnego, *Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, Departament Koordynacji Polityki Strukturalnej, Ministerstwo Rozwoju Regionalnego, Warszawa 2007, <http://www.fundusze-strukturalne.gov.pl/NR/rdonlyres/4E41CBAE-F849-48BE-BAC9-0614753930AA/38773/PlanewaluacjiNSROwrzesien2007180907.pdf>, s. 3.

<sup>2</sup> Por. Robson C., „Projektowanie ewaluacji”, w: L. Korporowicz (red.), *Ewaluacja w edukacji*, s. 147–152, Oficyna Naukowa, Warszawa 1997; Simons H., „Polityczne implikacje teorii ewaluacyjnych”, w: L. Korporowicz (red.), *Ewaluacja...*, op.cit., s. 65–92.

– w kontekście ich zaplanowania, wdrożenia i osiągniętych rezultatów<sup>3</sup>. Jej celem jest określenie rzeczywistych efektów oraz uzasadnienie interwencji w danej formie. Innymi słowy, ma ona za zadanie m.in.: ocenić stopień realizacji założonych celów, produktów, rezultatów; określić, czy cele i priorytety interwencji do końca zachowały swoją adekwatność w stosunku do pierwotnie postawionych problemów; ocenić oddziaływanie przedsięwzięcia i trwałość jego osiągnięć; postawić diagnozę efektów (oczekiwanych i nieoczekiwanych); zidentyfikować mocne i słabe strony interwencji; wskazać kierunki rozwoju i modyfikacji przyszłych interwencji.

Niezależnie jednak od swego szczegółowego zakresu ewaluacja *ex-post* powinna dostarczyć wiarygodnych i przydatnych informacji ze względu na swą **funkcję konkluzywną**, czyli rozliczeniową (*accountability*), oraz, w mniejszym stopniu, **funkcję poznawczą** (*knowledge*) i **funkcję formacyjną** – stymulowanie usprawnień i rozwoju organizacyjnego (*development*) w odniesieniu do projektowania i wdrażania przyszłych interwencji o podobnym charakterze<sup>4</sup>.

Określenie potrzeby przeprowadzenia ewaluacji, celu, dla którego ma być realizowana, oraz kluczowych funkcji, jakie ma spełnić, jest pierwszym krokiem zarówno w procesie poszukiwania informacji i odpowiedzi na pytania badawcze, jak i konceptualizacji oraz operacjonalizacji badania ewaluacyjnego.

## Konceptualizacja i operacjonalizacja ewaluacji *ex-post*

Proces przygotowania badania do realizacji należy podzielić na dwa etapy. Pierwszy z nich, **konceptualizacja** (planowanie ewaluacji), polega na 1) określeniu potrzeby, wartości oraz znaczenia ewaluacji danego projektu/programu, 2) możliwości jej realizacji i 3) wykorzystania wyników oraz 4) określeniu ewentualnych obszarów ryzyka i 5) politycznych implikacji<sup>5</sup> (innymi słowy, wizji sukcesu badania zawartej we wstępnym pomysle). Drugi etap, **operacjonalizacja** (projektowanie ewaluacji), to nasza wizja realizacji oraz zapewnienia jakości ewaluacji, a także oszacowanie koniecznych zasobów.

---

<sup>3</sup> Por. OECD, *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris 2002, <http://www.oecd.org/dataoecd/29/21/2754804.pdf>, s. 21.

<sup>4</sup> Więcej na ten temat w zamieszczonym w niniejszej publikacji artykule K. Olejniczaka „Teoretyczne podstawy ewaluacji *ex-post*”; Górniak J., „Funkcje ewaluacji”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, s. 16–23, Uniwersytet Ekonomiczny w Krakowie, Kraków 2007; Nevo D., „Konceptualizacja ewaluacji edukacyjnej”, w: L. Korporowicz (red.), *Ewaluacja...*, op.cit., s. 54.

<sup>5</sup> Więcej zob. Robson C., „Projektowanie ewaluacji”, op.cit., s. 158; Simons H., „Polityczne implikacje teorii ewaluacyjnych”, op.cit., s. 65–92.

## Konceptualizacja – planowanie ewaluacji

Pierwszy etap ewaluacji obejmuje przyjęcie projektu do dalszego opracowania oraz wzmocnienie celów realizacji projektu poprzez 1) regulacje instytucjonalne (wewnętrzne bądź zewnętrzne, wynikające z wymagań formalnych, np. nałożonego na instytucje obowiązku ewaluacji), 2) doprecyzowanie celów, 3) funkcji ewaluacji i 4) konsekwencji oceny (np. decyzja o usprawnieniu, reorientacji, kontynuacji lub zaprzestaniu realizacji kolejnej edycji programu; w przypadku wykrycia nieprawidłowości, konsekwencje prawne, finansowe) oraz, w tym kontekście, 5) zidentyfikowanie odbiorców ewaluacji<sup>6</sup> i 6) ich pierwszoplanowych potrzeb informacyjnych. Jest to etap, na którym powstają fundamenty zarządzania jakością projektu oraz komunikacji i dystrybucji wyników, decydujące o przyszłym ewentualnym sukcesie ewaluacji. Dopiero tak zbudowane podstawy można uzupełniać kolejnymi obszarami treści, mającymi wpływ na wstępny projekt ewaluacji.

Drugi etap to dookreślenie przedmiotu ewaluacji. Fazę tę można określić „spojrzeniem na program”, w tym m.in. na:

- typ interwencji i jej specyfikę,
- harmonogram realizacji przedsięwzięcia,
- cele interwencji,
- obszar objęty interwencją (kraj, region itd.),
- informacje/dane opisujące rzeczywisty przebieg realizacji i sposób wdrażania (m.in. sprawozdania beneficjentów, okresowe raporty monitoringu, bazy danych itd.),
- moment, w którym podejmujemy ocenę,
- efekty programu (oczekiwane i nieoczekiwane, pozytywne i negatywne, bezpośrednie i pośrednie, krótko- i długofalowe itd.),
- informacje/dane opisujące stopień osiągnięcia celów przedsięwzięcia (wyrażane poprzez wskaźniki osiągnięcia celów),
- informacje/dane konieczne do oceny makroekonomicznego oddziaływania przedsięwzięcia,
- informacje/dane dotyczące strategii (*policy*) wobec sektora/obszaru/zagadnienia stanowiącego przedmiot ewaluowanego przedsięwzięcia,
- informacje/dane opisujące zewnętrzne czynniki (wskazywane i niewskazywane w dokumentach programowych przedsięwzięcia) oddziałujące na proces realizacji i efekty przedsięwzięcia (m.in. wpływ kontekstu społeczno-gospodarczego, otoczenia instytucjonalnego i politycznego itd.).

Trzeci etap to ustalenie zakresu ewaluacji *ex-post*. Odbywa się on zazwyczaj w kilku krokach. Pierwszym z nich jest ustalenie poziomu badania (np. w przypadku ewaluacji funduszy strukturalnych, czy ewaluacja jest dokonywana z poziomu NPR/PWW lub NSS/NSRO, Programu Operacyjnego, wybranych osi priorytetowych, czy też ogranicza się do konkretnego działania, schematu

<sup>6</sup> Zob. tab. 1.

lub projektu) oraz charakteru ewaluacji, a więc czy będzie ona miała charakter pełny, odnoszący się do całej interwencji (jak np. *Ocena ex-ante Programu Współpracy Transgranicznej Litwa–Polska 2007–2013*; *Ewaluacja ex-post Planu Rozwoju Obszarów Wiejskich na lata 2004–2006*), czy też będzie odnosiła się do jakiegoś wybranego zagadnienia (jak np. *Ocena postępów wdrażania działania 2.6 ZPORR i szkolenia dla KOP w Małopolsce*; *Wpływ realizacji SPO WKP na poziom innowacyjności polskich przedsiębiorstw*), czy może obejmie zagadnienia horyzontalne, dotyczące więcej niż jednego programu (jak np. *Ocena wpływu funduszy strukturalnych UE na stan środowiska naturalnego*).

Krok drugi to ustalenie ram, czyli określenie, jak szeroko przeprowadzimy nasze badanie, np.: czy odnosimy się do jednego, czy też do kilku elementów interwencji; czy odnosimy się do jednej, czy też do kilku grup odbiorców (beneficjenci, partnerzy instytucjonalni, potencjalni odbiorcy itd.); czy bierzemy pod uwagę wyłącznie badaną interwencję, czy też uwzględniamy jej ewentualne synergie z innymi interwencjami lub programami publicznymi; czy bierzemy pod uwagę regiony, grupy osób, które nie są/nie były objęte programem, ale mogą odczuwać jego oddziaływanie; czy bierzemy pod uwagę kwestie związane z kolejnym cyklem programowania itp.<sup>7</sup>. Innymi słowy, przyjmujemy założenia dotyczące charakteru, natury badanej rzeczywistości/przedmiotu (w jakiej mierze jest ona zmienna i pod wpływem czego ulega zmianie), sposobu potraktowania przedmiotu badań z punktu widzenia jego autonomiczności (tj. czy potraktujemy go jako układ zamknięty, wyizolowany, czy też jako układ otwarty) oraz, w dalszych krokach, ewentualnych możliwości poznania tej rzeczywistości za pomocą określonej metody<sup>8</sup>.

Krok trzeci to określenie perspektywy czasowej, w jakiej będziemy dokonywać oceny po zakończeniu programu. W przypadku ewaluacji *ex-post* czas pomiaru bezpośrednio zależy od typu interwencji, specyfiki beneficjentów, możliwości pomiaru efektów, oddziaływania (krótko- lub długoterminowego). I tak na przykład projekty nakierowane na wsparcie inwestycyjne przedsiębiorstw (np. współfinansowanie zakupu środków trwałych, takich jak maszyny lub urządzenia) nie mogą być badane wcześniej niż w okresie 12–24 miesięcy po zakończeniu inwestycji, gdyż efekty i ich wpływ na funkcjonowanie i rozwój beneficjenta (przedsiębiorstwa) wymagają czasu m.in. na modyfikację procesu produkcyjnego, przeszkolenie personelu, wdrożenie nowych produktów i wprowadzenie ich na rynek, czy wreszcie zwiększenie produkcji i obrotów, spadek zadłużenia, zatrudnienie nowych osób i wzrost nakładów inwestycyjnych ze środków własnych i zobowiązań (z pominięciem dotacji). Z kolei projekty krótkie, takie jak nakierowanie na podnoszenie kwalifikacji (np. szkolenie dla pracowników przedsiębiorstw, osób bezrobotnych itp.), w tym samym

<sup>7</sup> Zob. inny przykład budowania zakresu projektu opartego na *Modelu 4 kapitalów*: <http://www.konferencja-ewaluacja.pl/2007/materialy.php> (prezentacja prof. G. Gorzelaka, slajd 4).

<sup>8</sup> Por. Słoma L., *Metody i techniki badań socjologicznych. Wybrane zagadnienia*, Uniwersytet Warmińsko-Mazurski, Olsztyn 2002, s. 29.

czasie (1–2 lata po ich zakończeniu) byłyby dla wielu beneficjentów trudne do odtworzenia w pamięci. Dlatego należy badać je zdecydowanie wcześniej: od trzech miesięcy do roku po zakończeniu udziału w programie.

Krok czwarty polega na pojęciowej strukturyzacji przedmiotu, czyli określeniu znaczenia wszystkich pojęć, jakie będziemy badać w celu systematyzacji komunikacji (z odbiorcami wyników ewaluacji i respondentami) oraz uniknięcia ewentualnych nieporozumień co do badanej rzeczywistości. Eksplikacja pojęć jest procesem ciągłym, rozpoczynającym się na poziomie konceptualizacji ewaluacji, a doprecyzowywanym już w toku gromadzenia i interpretacji danych. E. Babbie, za C. Marshalllem i G. Rossmanem, wskazuje na tak zwany „lejek pojęciowy”, dzięki któremu zainteresowania badacza stają się coraz bardziej skoncentrowane<sup>9</sup>. I tak np. wzrost konkurencyjności czy innowacyjności przedsiębiorstw w skali gospodarki może zawęzić się nawet do poziomu indywidualnej sytuacji grup przedsiębiorstw, ze względu na takie ich cechy, jak: wielkość przedsiębiorstwa, liczbę lat działalności, sektor, skalę działalności (lokalną, regionalną, ogólnokrajową, międzynarodową itp.), wielkość podejmowanych inwestycji w innowacje, nakłady na działalność B+R itp. Należy pamiętać, iż w przypadku ewaluacji proces ten jest silnie związany nie tylko z językiem badanych obiektów, szerszej rzeczywistości społeczno-gospodarczej, ale również z „językiem” samej interwencji, programu.

Krok piąty to sformułowanie problematyki badań, czyli szczegółowych pytań badawczych<sup>10</sup> (jakimi informacjami dysponujemy w chwili projekto-

<sup>9</sup> Zob. Babbie E., *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2003, s. 147.

<sup>10</sup> Przykłady pytań zastosowanych w ewaluacji *ex-post* projektu z obszaru rynku pracy:

1. Jaki był stopień osiągnięcia celów programu w zakresie zaplanowanych produktów, rezultatów i wpływu?
2. Czy nie wystąpił efekt creamingu – czy do programu rekrutowane były osoby, które faktycznie potrzebowały takiej pomocy?
3. Jak zmieniła się sytuacja zawodowa beneficjentów po zakończeniu uczestnictwa w projekcie?
4. Czy uczestnictwo w projekcie wpłynęło na wzrost zdolności adaptacyjnych na rynku pracy jego uczestników?
5. Czym spowodowane były problemy w rekrutacji beneficjentów zaliczonych do danej kategorii (np. absolwenci, długotrwale bezrobotni)? Czy wpływ na to miała specyfika regionalna, czy nierealne, nieprzystosowane do potrzeb regionu zdefiniowanie celów operacyjnych, czy też inne czynniki?
6. Czy zmianę sytuacji zawodowej beneficjentów ostatecznych możemy uznać za trwałą, czy tymczasową?
7. Na ile uzyskane efekty utrzymują się po zaprzestaniu finansowania?
8. Czy istnieją znaczące różnice w trwałości programu pomiędzy województwami, a jeśli tak, to jakie są tego przyczyny?
9. Czy istnieją znaczące różnice w trwałości programu pomiędzy podprojektami, a jeśli tak, to jakie?

Zob. też np.: PARP, *SIWZ Ewaluacji ex-post Phare 2003 Spójność Społeczna i Gospodarcza komponent Rozwój Zasobów Ludzkich*, PARP, Warszawa 2007, [http://bip.parp.gov.pl/\\_gAllery/10/72/1072.pdf](http://bip.parp.gov.pl/_gAllery/10/72/1072.pdf), s. 5–9; PARP, *SIWZ Ewaluacji sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego*, PARP, Warszawa 2007, [http://bip.parp.gov.pl/\\_gAllery/17/24/1724.pdf](http://bip.parp.gov.pl/_gAllery/17/24/1724.pdf), s. 1–4.

wania ewaluacji; jakie są potrzeby informacyjne ewaluatorów oraz zamawiających i czy te potrzeby są zbieżne; czy odnosimy się do całej logiki interwencji – celów, procesów, efektów, oddziaływania itp.), oraz postawienie hipotez (poprzedzonych ewentualnym rozpoznaniem badawczym), które poddamy weryfikacji. Jest to jeden z bardziej newralgicznych momentów konceptualizacji. Z jednej strony stanowi identyfikację potrzeb informacyjnych różnych grup odbiorców ewaluacji, z drugiej zaś – ich systematyzację w odniesieniu do kluczowych kryteriów oceny<sup>11</sup>.

### Operacjonalizacja – projektowanie ewaluacji

Po doprecyzowaniu celu badania, przedmiotu oraz ogólnego zakresu ewaluacji przechodzimy do fazy operacjonalizacji<sup>12</sup>, czyli do czynności, jakie należy wykonać, by poznać badaną interwencję/program.

Stworzenie konkretnych procedur badawczych obejmuje następujące obszary:

- Dobór i zdefiniowanie wskaźników i zmiennych, których wartości pozwalają na zmierzenie występowania zjawisk będących obiektem zainteresowania (np. rezultatów, oddziaływania danej interwencji). Dobór wskaźników powinien być raczej, jak pisze J. Górniak, konsekwencją wstępnego zidentyfikowania i konceptualizacji problemu lub wskazań niż sposobem na ich odkrycie<sup>13</sup>. W konstruowaniu polityk publicznych wskaźniki traktowane są jednak często nazbyt formalistycznie i zdają się pełnić jedynie funkcje dekoracyjną – dodawaną jako odrębną część programu, bez wyraźnego związku logicznego z celami programu<sup>14</sup>. Taka sytuacja często uniemożliwia sprawny pomiar *ex-post* całego procesu interwencji (nakładów → wyników → rezultatów → oddziaływania) i zmusza ewaluatorów niejako do odtwarzania i ponownego doboru wskaźników na poziomie celów ogólnych (wskaźnik wpływu globalnego) oraz celów szczegółowych (wskaźnik wpływu bezpośredniego). Należy jednak pamiętać, iż wskaźniki nie są sposobem (ani bezpośrednim, ani pośrednim) ustalenia stopnia wykonania poszczególnych wymiarów realizowanego programu, mają one jedynie umożliwić konstrukcję narzędzi pomiarowych o wyższej rzetelności i sprawdzalnej trafności<sup>15</sup>.

<sup>11</sup> Więcej na ten temat w zamieszczonym w tej publikacji artykule M. Szałaja „Mechanizmy doboru kryteriów oceny w ewaluacji *ex-post*”.

<sup>12</sup> Zob. przykład operacjonalizacji metodologii badań *foresight*: Kopyciński P., Mamica Ł., *Operacjonalizacja metodologii badań foresight*, Małopolska Szkoła Administracji Publicznej Akademii Ekonomicznej w Krakowie, Kraków 2006, <http://www.foresight.msap.pl/download/operacjonalizacja.pdf>.

<sup>13</sup> Górniak J., Mazur S., *Wskaźniki – instrument pomiaru skuteczności i efektywności polityki spójności*, <http://www.konferencja-ewaluacja.pl/2007/materialy.php>, s. 3.

<sup>14</sup> Por. Górniak J., „Ewaluacja cyklu polityk publicznych”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych...*, op.cit., s. 18.

<sup>15</sup> Por. ibidem, s. 19. Więcej na ten temat w artykule K. Keler i J. Górniaka „Wskaźniki w ewaluacji *ex-post* programów publicznych”.



- Wskazanie zbiorowości, w której będą realizowane badania – tzw. *study population* (kogo lub co chcemy badać). Podstawową populacją w badaniach ewaluacyjnych jest zazwyczaj grupa beneficjentów bądź podmiotów zaangażowanych w projektowanie, zarządzanie lub/i wdrażanie danej interwencji/programu. Niestety, nieco rzadziej ewaluatorzy w badaniach *ex-post* obejmują grupy nie powiązane bezpośrednio z daną interwencją, jak np. potencjalni odbiorcy programu, podmioty, które ubiegały się o wsparcie w danym obszarze, ale ich wnioski zostały odrzucone, grupy kontrolne, inne instytucje realizujące interwencje o podobnym charakterze itp. Ustalenie ostatecznej populacji nie zawsze jest zadaniem łatwym. Badacz w tym procesie może zetknąć się z szeregiem ograniczeń, np. ze względu na niekompletność listy podmiotów i/lub ich charakterystyki (np. cech indywidualnych, takich jak wielkość przedsiębiorstwa, sektor działalności, czy informacji dotyczących wsparcia, wielkość wnioskowanego/uzyskanego wsparcia, kwota pomocy publicznej uzyskana w ciągu trzech ostatnich lat), ich dostępności (np. zaprzestanie działalności podmiotów, rotacja kadry w instytucjach państwowych, jakość systemów informatycznych itd.). Może również zdecydować się na celowe ograniczenie badanej populacji ze względu na pewne zmienne, np. wybrać tylko tych beneficjentów, którzy korzystają ze wsparcia po raz pierwszy lub, wręcz przeciwnie, korzystali z niego wielokrotnie (np. przedsiębiorcy, którzy pozyskiwanie środków publicznych niejako wbudowują w strategię rozwoju firmy), albo beneficjentów, którzy skorzystali ze wszystkich elementów programu (np. osoby bezrobotne, które skorzystały z doradztwa, szkolenia, staży). Dopiero po dokonaniu takich zabiegów korekty badanego zbioru można podejmować decyzje co do objęcia badaniem całej populacji bądź też realizacji badania na określonej próbie<sup>16</sup> oraz doborze metod i technik badania.
- Wybór metod i technik badawczych, baz źródłowych oraz decyzji co do zasady analizy uzyskanego materiału empirycznego. Jak wskazuje C. Robson, ewaluacja nie jest jakąś nową czy też inną strategią badawczą, lecz badaniem, które służy realizacji określonego celu (ocenie rezultatów i efektów)<sup>17</sup>. Korzysta zatem z metod i technik badawczych wypracowanych w naukach społecznych, czyli metod ilościowych i jakościowych. Z wyborem metody wiąże się szereg uwarunkowań, które trzeba uwzględnić przy konstrukcji określonego badania ewaluacyjnego. Do kryteriów takiego wyboru należą m.in.: 1) specyfika interwencji publicznej, 2) cel badania ewaluacyjnego, 3) zakres i charakter poszukiwanych informacji, 4) moment oceny, 5) podmiot zlecający oraz 6) realizujący badanie, 7) perspektywa podmiotowo-celowa (czyli potencjalni odbiorcy, beneficjenci

<sup>16</sup> Więcej na temat doboru próby zob. Babbie E., *Badania społeczne...*, op.cit., s. 200–244.

<sup>17</sup> Robson C., „Projektowanie ewaluacji”, op.cit., s. 147.

interwencji, inni uczestnicy), 8) źródła, jakość i dostępność danych, a także 9) zasoby czasowe, finansowe, ludzkie, które wprowadzicie same w sobie bezpośrednio nie decydują o jakości gromadzonych danych, ale pośrednio decydują o możliwości realizacji badania. Wszystkie podjęte decyzje w obszarze wyboru metod i technik badawczych będą wprost rzutować na jakość wyników końcowych oceny<sup>18</sup>.

- Wybór metod oceny to etap konstrukcji projektu stanowiący pewien koncept, który na etapie realizacji badania, w fazie analiz i oceny, jest zawsze weryfikowany i uzupełniany o dodatkowe metody, niezbędne do przeprowadzenia pełnej analizy i weryfikacji postawionych hipotez oraz oszacowania efektów i określenia relacji przyczynowo-skutkowych. Wybór metody na tym etapie jest szczególnie istotny, przede wszystkim ze względu na takie metody, jak panel ekspertów, *benchmarking*, modele mikro/makro ekonomiczne, gdyż wymagają one przygotowania i zapewnienia odpowiednio wcześniej zaplecza (np. w przypadku panelu ekspertów konieczny jest wybór dobrych ekspertów, opracowanie zasad generowania produktów; realizowanie *benchmarking* wymaga dokonania wyboru oraz zaproszenia jednostek do porównania, a także opracowania adekwatnych metod zebrania danych; w modelach mikro/makro ekonomicznych konieczne jest zaś opracowanie modelu oraz ustalenie potrzebnych danych wchodzących w jego skład itd.)<sup>19</sup>.

Ostatnie dwa obszary operacjonalizacji to:

- Oszacowanie zasobów potrzebnych do realizacji projektu. Oszacowanie musi uwzględniać 1) zasoby czasowe – należy ustalić m.in. jaki będzie harmonogram prac: czy badania nie przypadną na okres urlopowy, jak długo będzie trwała terenowa realizacja badania, ile czasu potrzeba na analizy, czy wyniki kolejnych faz badania będą mogły się uzupełniać (np. wpływ badania jakościowego na konstrukcję narzędzi w badaniach kwestionariuszowych), 2) zasoby ludzkie, zarówno po stronie wykonawcy (skład zespołu), jak i po stronie zamawiającego (kierownik projektu oraz ewentualnie eksperci pomocniczy, skład grupy ds. ewaluacji itd.), 3) zasoby finansowe – budżet badania uwarunkowany skalą badania, doбором metod badawczych, ilością ekspertów itd., bądź też, zapewne dużo częściej, warunkujący ramy projektu.
- Decyzja o podmiocie dokonującym oceny oraz, w przypadku zamówień publicznych, trybie wyboru wykonawcy (zgodnie z Prawem Zamówień Publicznych<sup>20</sup>). Niezależnie od tego trybu, przy wyborze należy kiero-

<sup>18</sup> Zagadnieniu temu poświęcona została część II niniejszej publikacji.

<sup>19</sup> Więcej na ten temat zob.: European Commission, *Evaluating Socio-Economic Development*, Sourcebook 2, [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/evalsed/sourcebooks/method\\_techniques/index\\_en.htm](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/index_en.htm).

<sup>20</sup> Zob. Ustawa Prawo zamówień publicznych z dnia 29 stycznia 2004 r. (Dz. U. z 2006 r. Nr 164, poz. 1163, Nr 170, poz. 1217 i Nr 227, poz. 1658 oraz z 2007 r. Nr 64, poz. 427 i Nr 82, poz. 560).

wać się nie tylko 1) technicznymi możliwościami realizacyjnymi projektu, 2) doświadczeniem w realizacji projektów o podobnym charakterze oraz 3) poszczególnych technik badawczych, lecz także 4) doświadczeniem w zakresie badania danego obszaru tematycznego ewaluacji oraz 5) składem zespołu ekspertów. Warto zaznaczyć, że właśnie od składu zespołu zależy jakość i rzetelność ewaluacji oraz trafność, a także użyteczność rekomendacji. Dlatego też dla wielu zamawiających o jakości oferty przesądza zaproponowany przez wykonawcę zespół ekspertów. I tak na przykład, według J. Szlachty, w obszarze ewaluacji funduszy strukturalnych dobry ewaluator powinien posiadać kompetencje w trzech obszarach: 1) metod, technik i narzędzi ewaluacji, 2) różnych polityk Unii Europejskiej, w tym, przede wszystkim, polityki spójności, 3) wiedzy profesjonalnej dotyczącej poszczególnych obszarów interwencji polityki ekonomicznej państwa, takich jak: kapitał ludzki, przedsiębiorczość, w tym wspieranie małych i średnich przedsiębiorstw, gospodarka oparta na wiedzy i społeczeństwo informacyjne, infrastruktura techniczna i społeczna, polityka regionalna itp.<sup>21</sup>

Na koniec warto zaznaczyć, że każdy projekt ewaluacji powinien być skonstruowany w oparciu o kilka kluczowych zasad: przydatności, wykonalności oraz poprawności metodologicznej ewaluacji, a także w oparciu o zasady etyki badań<sup>22</sup>. Dopiero spełnienie tych warunków otwiera drogę do jego realizacji.

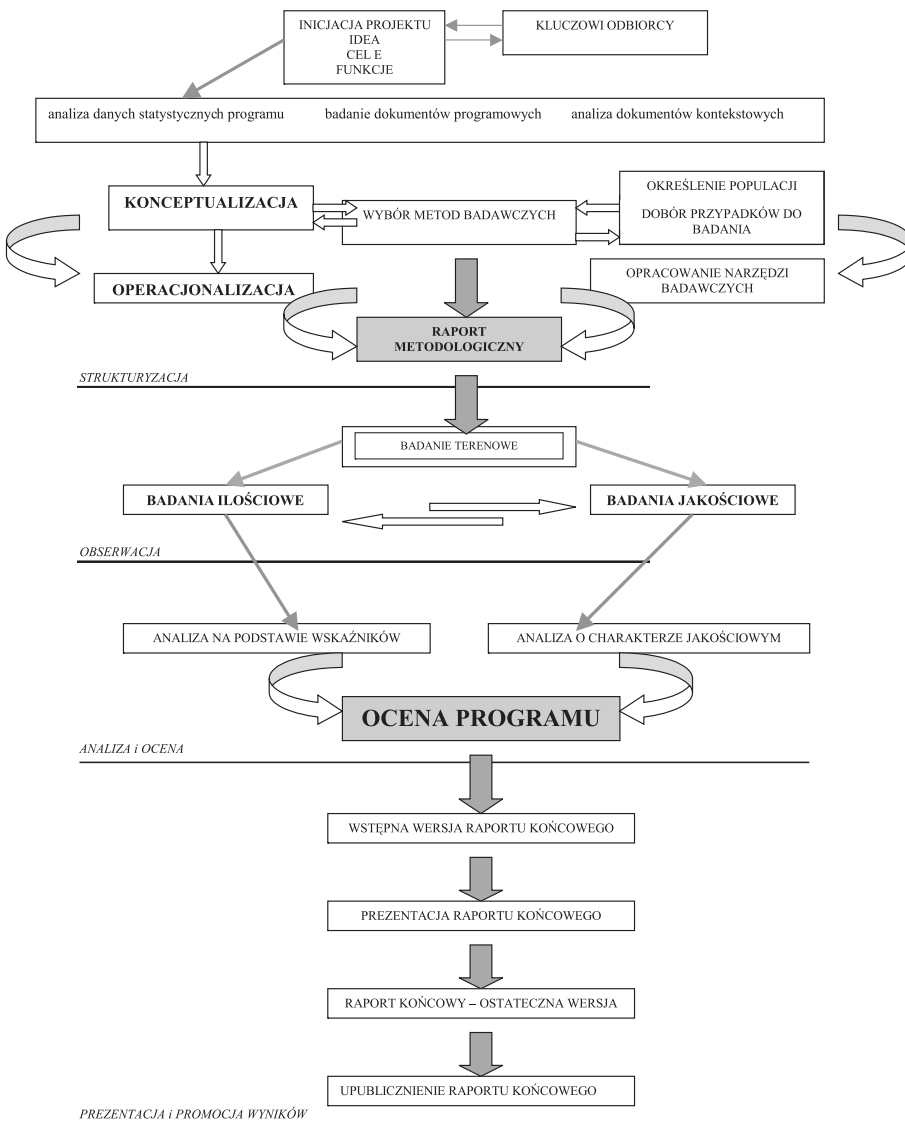
## **Realizacja badania ewaluacyjnego**

Realizację badania ewaluacyjnego można rozpatrywać z perspektywy zamawiającego, z perspektywy wykonawcy oraz z poziomu poszczególnych faz realizacji: 1) strukturyzacji, 2) obserwacji, 3) analizy, 4) oceny, 5) raportowania wyników (w tym wniosków i rekomendacji z oceny) oraz ich upublicznienia.

---

<sup>21</sup> Szlachta J., „System ewaluacji funduszy strukturalnych w Polsce”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych...*, op.cit., s. 86.

<sup>22</sup> Por. Nevo D., „Konceptualizacja...”, op.cit., s. 63.



Rysunek 1. Proces Ewaluacji

Źródło: Opracowanie własne na podstawie: Guba E., Lincoln Y., *Effective Evaluation*, Jossey-Bass, San Francisco 1981; Babbie E., *Badania społeczne...*, op.cit.

**Faza strukturyzacji** obejmuje dopracowanie projektu ewaluacji w obszarze: wyboru oraz dookreślenia kryteriów i elementów, które mają być poddane ewaluacji; wyboru wskaźników, za pomocą których będzie badane dane zjawisko; metodologii badania ewaluacyjnego; wyboru i opracowania narzędzi obserwacji; określenia sposobu dotarcia do respondentów badania. Na tym etapie stosuje się szereg metod projektowania ewaluacji, np. analizę SWOT,

matrycę logiczną, mapy oddziaływania, metaplan, studium wykonalności, konsultacje z interesariuszami, modele logiczne<sup>23</sup>. Faza ta jest fundamentem dla dalszych działań w projekcie ewaluacyjnym. Na tym etapie zarówno zamawiający, jak i ewaluatorzy popełniają najczęściej błędów. Do najczęściej spotykanych należą:

- schematyzm w doborze metod badawczych,
- zbyt szeroki projekt ewaluacji – długa lista zagadnień badawczych oraz brak priorytetyzacji obszarów badawczych,
- niedoszacowanie zasobów finansowych lub ludzkich po stronie zamawiającego lub wykonawcy,
- zbyt krótki czas przewidziany na rzetelne przygotowanie projektu, zapoznanie się z przedmiotem przez zespół, opracowanie narzędzi badawczych,
- brak elastyczności ze strony zamawiającego (w przypadku zamówień publicznych – ograniczenia wynikające z SIWZ i oferty wykonawcy),
- ograniczenia w komunikacji lub brak współpracy pomiędzy wykonawcą/zamawiającym/grupą sterującą ds. ewaluacji,
- brak jasnej strategii ewaluacji (dla kogo przeprowadza się ewaluację? jaki jest jej cel? jak wykorzystać wyniki? itd.),
- błędna bądź niepełna identyfikacja potrzebnych źródeł danych zastanych – głównie baz danych, ale również dokumentacji programowej, danych z monitoringu, wcześniejszych wyników ewaluacji, urzędowych danych statystycznych, ekspertyz itp.,
- brak jasnego podziału kompetencji i zadań pomiędzy członkami zespołu.

**Faza obserwacji** obejmuje zbieranie danych potrzebnych do analizy – danych zastanych (dokumentacji programowej, danych z monitoringu, wyników wcześniejszych ewaluacji, badań w obrębie tematu, danych statystycznych, ekspertyz itd.) oraz danych gromadzonych w oparciu o badanie terenowe. W fazie tej wykorzystywane są metody ilościowe oraz jakościowe<sup>24</sup>. Faza obserwacji to etap szczególnej dbałości o jakość i wiarygodność danych. Do najczęstszych błędów, obok błędów formalnych, wynikających z niepoprawnego zastosowania technik bądź braku rzetelności realizacji, należą:

- brak zróżnicowania narzędzi obserwacji (nieuwzględnienie zasady triangulacji) lub też ograniczenie się tylko do analizy *desk research*, wyboru metod ilościowych bądź jakościowych,
- zalewanie danymi (zbyt rozbudowane narzędzia badawcze, gromadzenie danych, które nie zostaną uwzględnione w analizach, czy danych kontekstowych, które są mało przydatne z punktu widzenia badanego przedmiotu).

<sup>23</sup> Więcej na ten temat zob.: European Commission, *Evaluating...*, op.cit.

<sup>24</sup> Zob. część II niniejszej publikacji.

**Faza analiz** obejmuje: 1) interpretację zebranych danych, ich porównanie i uchwycenie różnic, odkrywanie prawidłowości zachodzących procesów oraz podstawowych znaczeń i struktur, 2) weryfikację hipotez, 3) analizę przyczynowo-skutkową oraz 4) oszacowanie efektów programu przy użyciu metod statystycznych, analizy jakościowej, modeli ekonomicznych, grup porównawczych, paneli ekspertów i in. W praktyce badawczej na tę fazę na ogół rezerwuje się bardzo mało czasu bądź czas ten, w przypadku występowania opóźnień, jest konsumowany przez wcześniejsze etapy. Powoduje to często, iż analizy są pobieżne, jednowymiarowe, pozostające na poziomie tzw. „stop klatki” – opisu stanu zastanego, bez pogłębionej interpretacji. Obok powyższych, do często spotykanych błędów zaliczyć można:

- zbyt dużą koncentrację ewaluatorów na szczegółach, bez uchwycenia ogólnego obrazu wyników,
- brak weryfikacji hipotez,
- brak analizy przyczynowo-skutkowej,
- brak oszacowania efektów programu,
- pozostawanie w analizach na poziomie efektu brutto.

**Faza oceny** obejmuje ocenę efektów programu w odniesieniu do uprzednio sformułowanych pytań ewaluacyjnych i w oparciu o przyjęte kryteria oceny. Konstrukcja oceny powinna być budowana w oparciu o schemat oceny, który, z jednej strony, ułatwia zespołowi jej rzetelne przeprowadzenie, z drugiej zaś – ułatwia odbiorcom zrozumienie jej przebiegu i podstaw wydanych osądów. Do metod wykorzystywanych na tym etapie zalicza się m.in.: analizę kosztów i korzyści, analizę kosztów i efektywności, analizę wielokryteriową, panel ekspertów, *benchmarking*.

Fundamentem oceny zawsze powinny być rzetelne i wiarygodne dane oraz szczegółowa analiza. Ponadto, ocena powinna być uzasadniona, czyli wynikająca z faktów, trafna. Zarówno ewaluatorzy, jak i zamawiający powinni zadbać o jej bezstronność, obiektywizm i, przede wszystkim, przejrzystość. Obok pominięcia powyższych zasad do najczęściej spotykanych błędów na tym etapie należą:

- ocena efektów programu bez pełnego odniesienia się do uprzednio sformułowanych pytań ewaluacyjnych lub przyjętych kryteriów oceny,
- brak mocnych podstaw oceny – uzasadnienia, oparcia na faktach, wynikach badań, wiarygodnych danych,
- brak przeniesienia wyników oceny na rzetelne oraz użyteczne wnioski i rekomendacje,
- brak dyskusji/interpretacji w gronie ekspertów/odbiorców, przez co wynik oceny mogą być ograniczone jedynie do subiektywnej interpretacji ewaluatorów,
- brak zastosowania zróżnicowanych narzędzi prezentacji wyników oceny (dostosowania do odbiorców wyników).

Tabela 1. Cykl życia ewaluacji i jego elementy kluczowe

	Strukturyzacja		Obserwacja Analiza Ocena	Prezentacja wyników oceny
	KONCEPTUALIZACJA	OPERACJONALIZACJA		
Fazy życia projektu ewaluacyjnego	inicjowanie definiowanie powołanie kierownika projektu/ jednostki odpowiedzialnej za wykonanie*/zlecenie ewaluacji	określenie struktury projektu planowanie przebiegu projektu planowanie zasobów	realizacja kontrola koordynacja	zamknięcie projektu
Charakter działań	konceptyjny	planistyczny	wykonawczy kontrolny koordynacyjny	sprawozdawczy odbiorczy promocyjny wdrożeńowy
Produkty	projekt ewaluacji ekspertyzy, analizy, opracowania pomocnicze	SIWZ lub zakres prac wykonawcy	raport metodologiczny schemat oceny raporty cząstkowe z badań raport wstępny i końcowy	raport końcowy wraz z rekomendacjami prezentacje wyników tabela zaleceń publikacje wyników/artykuly prasowe
Bezpośredni uczestnicy / kluczowi odbiorcy	zleceniodawca kierownik projektu kierownictwo instytucji zaangażowanych w zarządzanie/wdrażanie inicjatywy/programu/projektu grupa ds. ewaluacji ekspertci	kierownik projektu zespół projektowy ekspertci	kierownik projektu wraz z zespołem projektowym ze strony zamawiającego ewaluację wykonawca wraz z zespołem ewaluacyjnym kluczowi aktorzy/ respondenci–beneficjenci/ aplikujący o wsparcie/projekt	wykonawca zleceniodawca instytucje zaangażowane w realizację oceny interwencji ( <i>management</i> , pracownicy szeregowi) kluczowi aktorzy/ respondenci–beneficjenci/aplikujący o wsparcie/projekt

\* w przypadku ewaluacji realizowanych wewnętrznie

Źródło: Opracowanie własne, por.: Griffin R.W., *Podstawy zarządzania organizacjami*, przekł. M. Rusiński, Wydawnictwo Naukowe PWN, Warszawa 1996; Trocki M., Grucza B., Ogonek K., *Zarządzanie projektami*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2003.

Ostatnia faza, często pomijana w wielu opracowaniach, obejmuje końcowy produkt procesu ewaluacji – **raport końcowy, jego upublicznienie i promocję**. Bardzo często zdarza się, że etap ten jest nie dość dobrze dopracowany. A przecież od formy, zakresu, sposobu, dobrania sposobu komunikacji i dystrybucji raportu zależy wykorzystanie wyników ewaluacji. Raport końcowy jest produktem, który, tak samo jak inne, wymaga oprawy i promocji, dlatego też na tym etapie można posługiwać się szeregiem metod wypracowanych w takich dziedzinach jak *Public Relations* czy marketing. Bez wątpienia, napisanie dobrego raportu jest sztuką, ale jeszcze większą sztuką jest napisanie go i zaprezentowanie w przystępnej formie oraz dotarcie z wynikami do jak największej grupy odbiorców. Dlatego zachodzi konieczność współpracy wykonawcy z zamawiającym oraz poszukiwania jak najlepszych form prezentacji oraz kanałów komunikacji. Dziś już nie wystarczy organizacja jednego spotkania, na którym wykonawca zaprezentuje najważniejsze wnioski i rekomendacje, umieszczenie raportu na stronach internetowych instytucji, rozesłanie raportu e-mailem. Codziennie jesteśmy zalewani tysiącami informacji i musimy dokonywać ich selekcji, toteż zadaniem ewaluatora nie może być tylko zrobienie dobrego badania; równie ważną pracą jest skondensowanie i zaprezentowanie informacji (wniosków, rekomendacji) tak, by były one ciekawe, zrozumiałe, przejrzyste, a nade wszystko – użyteczne dla odbiorców. Niestety, w przeciwieństwie do projektów badawczych realizowanych dla firm komercyjnych, sfera ta w ewaluacji wciąż jest bardzo niedopracowana, przez co wiele wartościowych projektów pozostaje niezauważonych, i to nie tylko przez otoczenie zewnętrzne wobec danej interwencji, lecz także – może nazbyt często – przez środowisko wewnątrz danej instytucji czy projektu. W efekcie, często to właśnie końcowy etap decyduje o spełnieniu faktycznych funkcji ewaluacji, czyli ewentualnym sukcesie bądź porażce naszego działania.

Powyższe spojrzenie na ewaluację w ujęciu procesowym oraz jego ewentualne newralgiczne punkty można uzupełnić o dodatkową optykę – jego etapy, charakter działań, produkty, poszczególne typy uczestników i odbiorców zaangażowanych na różnych etapach realizacji projektu (zob. tab. 1).

## Podsumowanie

Zapewne wielu tak teoretyków, jak i praktyków zgodzi się ze mną, iż za jakość ewaluacji, jej produkt końcowy oraz wykorzystanie wyników odpowiedzialni są w równej mierze ewaluatorzy, jak i sami zamawiający ewaluację. Od współpracy, zaangażowania oraz świadomości wagi poszczególnych kroków w procesie ewaluacji zależy jakość produktu końcowego oraz ewentualny sukces bądź porażka projektu. Oczywiście, owo zaangażowanie nie będzie na wszystkich etapach takie samo. W poszczególnych fazach procesu role i zadania głównych aktorów będą inne. I tak na początku, w fazie inicjo-



wania projektu, określania celu ewaluacji oraz konceptualizacji, główna rola przypada zamawiającemu, samemu bądź w asyście ekspertów, grupy ds. ewaluacji. Dopiero później na scenę wkracza ewaluator, który musi dopracować, zoperacjonalizować pomysł oraz w „kolejnych odsłonach” wykonać prace badawczo-analityczne. Zamawiający tylko pozornie na tym etapie jest „niepotrzebny”. Jego zaangażowanie powinno mieć charakter wspierający, koordynacyjny i kontrolny. Dopiero pojawienie się raportu końcowego uaktywnia go w pełni, aż do zakończenia projektu. Warto jednak zaznaczyć, iż zarówno role, jak i poszczególne zadania powinny być jasno określone, gdyż może się zdarzyć sytuacja (zamierzona, w wyniku nacisków ze strony zamawiającego, lub zupełnie nieplanowana, np. w wyniku sympatii ewaluatora dla zespołu po stronie zamawiającego ewaluację), w której ewaluator zatraci obiektywizm oceny bądź zamawiający przejmie rolę czy też zadania ewaluatora, co zawsze będzie prowadziło do obciążeń oceny końcowej.

Mam nadzieję, że powyższe przybliżenie ewaluacji w kontekście wielu perspektyw ułatwi czytelnikowi zarówno płynne poruszanie się po jej skomplikowanym procesie, jak i implementację najlepszych rozwiązań. Na pewno artykuł ten nie wyczerpuje wszystkich kwestii i może być jedynie wstępem do budowania najbardziej odpowiednich, indywidualnych strategii badawczych oraz dalszej dyskusji, która będzie przyczyniać się do rozwoju zarówno teorii, jak i praktyki ewaluacyjnej.

---

**Agnieszka Haber** – absolwentka Wydziału 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 Zasobów Ludzkich, Rozwoju sektora MSP oraz projektów infrastrukturalnych wdrażanych przez Agencję. Od 2005 roku organizator międzynarodowej cyklicznej Konferencji Ewaluacyjnej, odbywającej się pod auspicjami MRR oraz PARP. Od 2005 roku uczestnik grup ds. ewaluacji kolejnych programów: EFS, PO IG, PO KL, PO PRPW. Autorka prowadzi wykłady i warsztaty z zakresu ewaluacji interwencji publicznych w Wyższej Szkole Przedsiębiorczości i Zarządzania im. L. Koźmińskiego, Krajowej Szkole Administracji Publicznej oraz Uniwersytecie Warszawskim.

## Bibliografia

- Babbie E., *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2003.
- European Commission, *Evaluating Socio-Economic Development*, Sourcebook 2, [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/evalsed/sourcebooks/method\\_techniques/index\\_en.htm](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/index_en.htm).

- Górniak J., „Funkcje ewaluacji”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, s. 16–23, Uniwersytet Ekonomiczny w Krakowie, Kraków 2007.
- Górniak J., Mazur S., *Wskaźniki – instrument pomiaru skuteczności i efektywności polityki spójności*, <http://www.konferencja-ewaluacja.pl/2007/materialy.php>.
- Griffin R.W., *Podstawy zarządzania organizacjami*, przekł. M. Rusiński, Wydawnictwo Naukowe PWN, Warszawa 1996.
- Guba E., Lincoln Y., *Effective Evaluation*, Jossey-Bass, San Francisco 1981, <http://www.konferencja-ewaluacja.pl/2007/materialy.php>.
- Kopyciński P., Mamica Ł., *Operacjonalizacja metodologii badań foresight*, Małopolska Szkoła Administracji Publicznej Akademii Ekonomicznej w Krakowie, Kraków 2006, <http://www.foresight.msap.pl/download/operacjonalizacja.pdf>.
- Ministerstwo Rozwoju Regionalnego, *Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, Departament Koordynacji Polityki Strukturalnej, Ministerstwo Rozwoju Regionalnego, Warszawa 2007.
- Nevo D., „Konceptualizacja ewaluacji edukacyjnej”, w: L. Korporowicz (red.), *Ewaluacja w edukacji*, Oficyna Naukowa, Warszawa 1997.
- OECD, *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris 2002, <http://www.oecd.org/dataoecd/29/21/2754804>.
- PARP, *SIWZ Ewaluacji ex-post Phare 2003 Spójność Społeczna i Gospodarcza komponent Rozwój Zasobów Ludzkich*, PARP, Warszawa 2007, [http://bip.parp.gov.pl/\\_gAllery/10/72/1072.pdf](http://bip.parp.gov.pl/_gAllery/10/72/1072.pdf).
- PARP, *SIWZ Ewaluacji sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego*, PARP, Warszawa 2007, [http://bip.parp.gov.pl/\\_gAllery/17/24/1724.pdf](http://bip.parp.gov.pl/_gAllery/17/24/1724.pdf).
- Robson C., „Projektowanie ewaluacji”, w: L. Korporowicz (red.), *Ewaluacja w edukacji*, s. 147–152, Oficyna Naukowa, Warszawa 1997.
- Simons H., „Polityczne implikacje teorii ewaluacyjnych”, w: L. Korporowicz (red.), *Ewaluacja w edukacji*, s. 65–92, Oficyna Naukowa, Warszawa 1997.
- Słoma L., *Metody i techniki badań socjologicznych. Wybrane zagadnienia*, Uniwersytet Warmińsko-Mazurski, Olsztyn 2002.
- Szlachta J., „System ewaluacji funduszy strukturalnych w Polsce”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Uniwersytet Ekonomiczny w Krakowie, Kraków 2007.
- Trocki M., Grucza B., Ogonek K., *Zarządzanie projektami*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2003.

*Maciej Szalaj*

## **Mechanizmy doboru kryteriów oceny w ewaluacji *ex-post***

### **Wprowadzenie**

Ewaluacja w rozumieniu przyjętym w Polsce utożsamiana jest z oceną interwencji publicznych. Interpretacja ta spotyka się z licznymi głosami krytyki ze strony praktyków i teoretyków. Wskazuje się zazwyczaj na silne konotacje przypisane w polszczyźnie słowu „ocena”. W istocie, nie powinno się postrzegać ewaluacji wyłącznie poprzez pryzmat pojęć związanych z tym terminem co najmniej z dwóch powodów. Po pierwsze, jest to podejście chybione poznawczo, gdyż rola ewaluacji wykracza dalece poza funkcje oceniające czy też kontrolne. Po drugie, obranie takiej perspektywy utrudnia kulturowe zakorzenienie ewaluacji w lokalnym kontekście. Obecność ewaluatorów w instytucjach zaangażowanych w badane projekty odbierana jest częstokroć jako potencjalne zagrożenie. Samych ewaluatorów identyfikuje się natomiast jako przedstawicieli instytucji całkowicie zewnętrznych wobec badanej rzeczywistości.

Cechą odróżniającą ewaluację od badań akademickich czy też działań typowo kontrolnych jest właśnie jej organiczne powiązanie ze społeczeństwem poprzez sieć specyficznych zależności. Komponent oceny wpisuje się naturalnie w każdą ewaluację, jednakże w żadnym przypadku nie obejmuje całości procesu ewaluacyjnego. Ewaluacja składa się z określonych elementów, które wieńczy dopiero ocena i rekomendacje. Jednym z newralgicznych elementów procesu ewaluacji są kryteria oceny, będące tematem tego artykułu.

W części pierwszej omówiona zostanie lokalizacja kryteriów oceny w ogólnej logice procesu ewaluacyjnego. Pozwoli to na wskazanie swoistego „rdzenia tożsamości”, wspólnego każdej ewaluacji, niezależnie od specyfiki badanego przedmiotu czy też obieranej metody badawczej, bez względu na specyficzne dla danego przypadku podejście i obszar interwencji publicznej. W drugiej części tekstu przedstawione zostaną klasyfikacja kryteriów oceny oraz wymiary warunkujące ich dobór. Część ostatnia stanowi natomiast analizę specyfiki doboru kryteriów na potrzeby ewaluacji *ex-post*.

Wykorzystując zarysowane w artykule podłoże teoretyczne, spróbujemy wreszcie rozstrzygnąć, czy możliwym jest wskazanie jednolitego standardu doboru właściwych kryteriów oraz jakie cechy powinny posiadać owe kryteria.

## Kryteria oceny w logice procesu ewaluacyjnego

Istotą każdego rodzaju badań jest dążenie ku poznaniu, podporządkowane określonym kanonom postępowania, które gromadzi się zazwyczaj w opasłych podręcznikach metodologii lub, przybierających nieco mniej okazałe rozmiary, kodeksach dobrych praktyk badawczych. W zakresie refleksji metodologicznej ewaluacja osiągnęła stadium samoświadomości, dorównując dyscyplinom o pozycji ugruntowanej wielowiekową tradycją. Ewaluatorzy prowadzący badania w bardzo różnych obszarach, hołdujący różnorodnym orientacjom badawczym oraz, często kolidującym ze sobą, wartościom, postępują więc w swojej praktyce wedle jednolitego wzorca, na który składają się, oprócz etapów wspólnych wielu dziedzinom badań, także etapy całkowicie specyficzne dla badań ewaluacyjnych. Do tej drugiej kategorii zalicza się ustalanie kryteriów oceny.

Wspomniany powyżej wzorzec procesu badawczego Fourier określa jako model ogólnej logiki ewaluacji<sup>1</sup>. Oprócz ustalania kryteriów, składa się on z następujących elementów:

- wyznaczanie standardów jakości,
- pomiar efektów i zestawienie ich ze standardami,
- synteza i ocena uzyskanych wyników.

Logika ogólna obecna jest we wszystkich badaniach ewaluacyjnych. Wyjątku nie stanowią tu nawet projekty realizowane w ramach nurtów postrzeganych jako kontestujące powszechnie dotychczas przyjmowane założenia badawcze. Ramy wyznaczone przez tę logikę określają granicę ewaluacyjnego „rdzenia tożsamości”. Cytowany powyżej autor w obrazowy sposób ujął rolę ogólnej logiki ewaluacji, pisząc: „[ogólna logika dowodzenia] określa reguły gry, którym podporządkować muszą się wszyscy ewaluatorzy prowadzący badania w jakimkolwiek obszarze. Decyduje ona także o osobliwościach gry zwanej ewaluacją. Tak więc każdy, kto twierdzi, iż zajmuje się ewaluacją, musi tym samym wyznaczać kryteria i standardy jakości, a następnie zgodnie z nimi identyfikować i mierzyć efekty oraz syntetyzować uzyskane dane w celu dokonania ocen”<sup>2</sup>.

---

<sup>1</sup> Fourier D.M., „Establishing evaluative conclusions: a distinction between general and working logic”, w: E. Stern (red.), *Evaluation Research Methods*, vol. 1, Sage Publications, London 2005.

<sup>2</sup> Ibidem.

Bazująca na jednolitym zestawie reguł gry tożsamość dyscypliny przekłada się bezpośrednio na rozwój tożsamości zawodowej ewaluatorów, wywodzących się z bardzo różnorodnych środowisk. Ugruntowana tożsamość metodologiczna dostarcza także racji bytu stowarzyszeniom i sieciom profesjonalnym, które zyskują dzięki niej podstawy do tworzenia prawomocnych kanonów praktyki ewaluacyjnej. Wyznaczanie kryteriów oceny stanowi pierwszy etap realizacji projektów, które podlegają weryfikacji w świetle ogólnie przyjętych standardów. Pozwalają one odpowiedzieć na pytanie, czy ewaluator nie popełnia błędów w sztuce. Kanony praktyki nie rozstrzygają, jak dokładnie powinien zostać zaprojektowany i zrealizowany dany projekt. Nie jest bowiem możliwe wskazanie jednego odpowiedniego zestawu pytań i takich metod badawczych, kryteriów oceny czy też standardów jakości.

Ograniczony zakres wytycznych wpływających z kanonu metodologicznego obrazuje różnice pomiędzy ogólną logiką ewaluacji a logiką operacyjną (*working logic*) poszczególnych projektów. Logika operacyjna, w odróżnieniu od ewaluacyjnej, jest zróżnicowana w zależności od przedmiotu ewaluacji i zastosowanego wobec niego podejścia. Można przyjąć, iż pomiędzy logiką ogólną a operacyjną zachodzi podobna relacja jak pomiędzy strategią a taktyką. Element strategiczny, jakim są kryteria oceny, posiada zatem swój taktyczny odpowiednik w postaci konkretnych zestawów kryteriów, wyznaczonych w toku praktyki ewaluacyjnej. Literatura przedmiotu rozróżnia dwa rozumienia logiki operacyjnej<sup>3</sup>.

Wedle pierwszego z nich, logikę operacyjną definiuje się jako zestaw parametrów dotyczących problemu badawczego, fenomenu, pytań ewaluacyjnych, twierdzeń. Wypadkowa tych parametrów określa specyfikę poszczególnych paradygmatów badawczych. Na ich podstawie wyznaczyć można również wymiary, wedle których teoretycy budują swoje podejście do praktyki ewaluacyjnej. Dobór kryteriów oceny na potrzeby danego projektu warunkowany jest poprzez charakterystyczne dla określonego podejścia rozumienie fenomenu badanego obszaru. Fenomen należy w tym kontekście postrzegać jako roboczą koncepcję natury przedmiotu badania, ustaloną na podstawie przyjętej definicji głównego problemu badawczego.

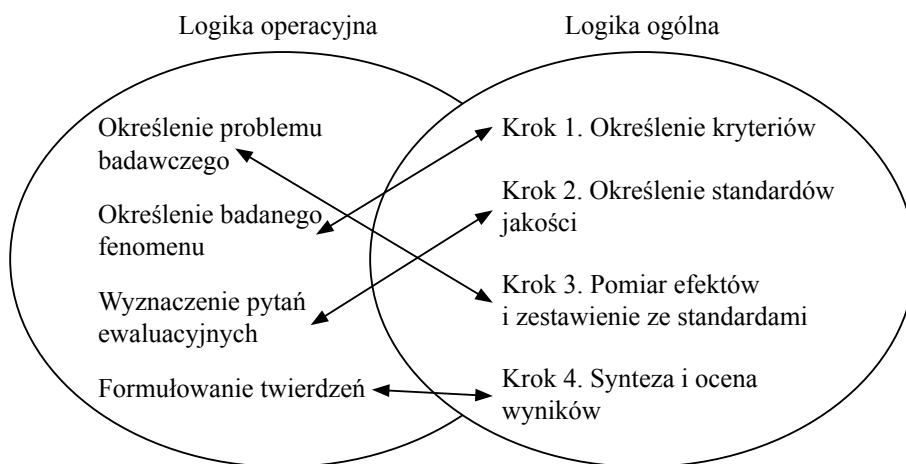
Relacja pomiędzy fenomenem a kryteriami oceny jest kluczowym elementem każdej ewaluacji, determinującym charakter kolejnych etapów realizacji projektu. Przyjęte kryteria wyznaczają bowiem ramy schematu procesu badawczego aż do momentu syntezy danych i formułowania oceny.

Drugie rozumienie logiki operacyjnej odnosi się do struktury argumentacji, określającej schemat dowodzenia, prowadzonego z uwzględnieniem określonych parametrów.

Miejsce kryteriów oceny w tak rozumianej logice operacyjnej nie jest precyzyjnie określone. Ich rola polega raczej na współwyznaczaniu kontek-

<sup>3</sup> Por. *ibidem*, s. 56–64.

stu realizacji projektu. Oddziaływanie doboru kryteriów widać chociażby w metodach badawczych i sposobie gromadzenia dowodów, gdyż określa on rodzaje poszukiwanych danych (np. ocena wedle kryterium skuteczności osiągania zakładanych celów implikuje poszukiwanie danych dotyczących efektów interwencji, mieszczących się w wyznaczonych celami kategoriach). Określana przez kryteria oceny perspektywa może stanowić źródło legitymizacji dla badań, których poprawność metodologiczna nie poddaje się prostej ocenie z wykorzystaniem zwyczajowych standardów naukowości. Obecność kryteriów oceny w kontekście badawczym jest, ogólnie 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.



Rysunek 1. Relacje między logiką ogólną a logiką operacyjną

Źródło: Fourier D.M., „Establishing evaluative conclusions...”, op.cit., s. 58.

## Rodzaje kryteriów

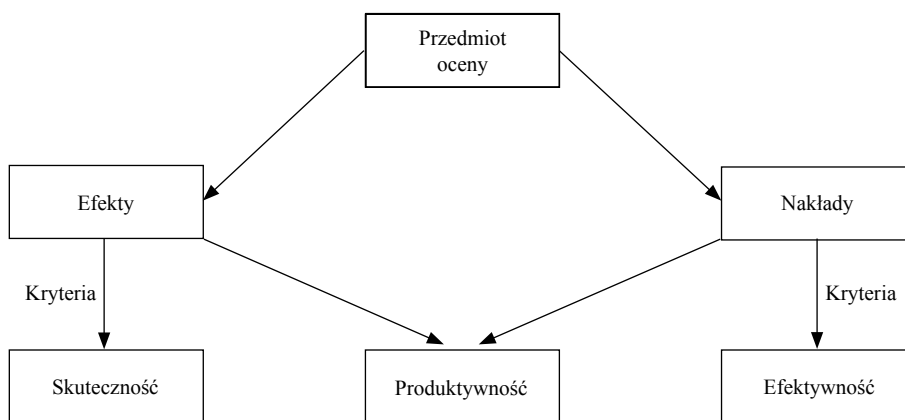
Kwestia rodzajów kryteriów stosowanych w ewaluacji może wydawać się zagadnieniem błahym. Wedle niektórych opinii, dobór kryteriów dyktowany jest wyłącznie specyfiką projektów, przez co tworzenie ogólnych typologii i jednolitych definicji mija się z celem. Sprawę komplikuje jednak fakt, iż w źródłach polskojęzycznych, wymieniających zestaw najczęściej występujących w praktyce kryteriów, nie opisuje się wyczerpująco uwarunkowań decydujących o ich powszechności<sup>4</sup>. Kolejny problem wiąże się z recepcją

<sup>4</sup> Przykładowe źródła: strony internetowe Władzy Wdrażającej Programy Europejskie: [http://www.wvpw.gov.pl/index.php?params%5Bsection\\_id%5D=4](http://www.wvpw.gov.pl/index.php?params%5Bsection_id%5D=4); Polskie Towarzystwo Ewaluacyjne, *Ewaluacja funduszy strukturalnych*, Opracowanie na zlecenie Wydziału Monitoringu i Raportowania

terminów anglojęzycznych (*efficiency, effectiveness, utility, relevance* itp.). W bezpośrednim tłumaczeniu brzmią one niezbyt naturalnie, co wytwarza swego rodzaju chaos pojęciowy. Dodatkowym powodem nieporozumień jest rozpowszechnione wśród ewaluatorów mniemanie, iż nazwy kryteriów nie posiadają jednoznacznej wykładni, a ich znaczenie może być interpretowane z bardzo daleko idącą dowolnością.

Podjęcie próby uporządkowania terminologii towarzyszącej kryteriom oceny wpłynie zatem pozytywnie na przejrzystość dalszej części wywodu. Zaprezentowana poniżej klasyfikacja opiera się na podziale według dwóch najbardziej elementarnych aspektów wszystkich interwencji publicznych – nakładów i efektów.

Przyjmując te aspekty za oś podziału pomiędzy rodzajami kryteriów oceny, uzyskamy dwie kategorie główne oraz jedną kategorię pośrednią



Rysunek 2. Główne kategorie kryteriów

Źródło: Opracowanie własne na podstawie: Vedung E., „Merit criteria and performance standards”, w: E. Stern (red.), *Evaluation Research Methods*, op.cit., s. 183–200.

Pierwsza główna kategoria to efektywność, czyli szeroko rozumiane kryterium uwzględniające relację nakładów do osiągniętych efektów. Kategoria ta nie stanowi źródła istotnych nieporozumień definicyjnych, gdyż jej rozumienie jest jednorodne i raczej nie spotyka się tu rozbieżnych względem siebie podejść. Ocena na podstawie tego kryterium prowadzona jest z wykorzystaniem dwóch głównych metod:

- analizy kosztów i korzyści (*cost-benefit*), która opiera się na kalkulacji wyrażanych w jednostkach pieniężnych efektów w odniesieniu do wyrażanych w jednostkach pieniężnych nakładów,

- analizy kosztów i skuteczności (*cost-effectiveness*), będącej kalkulacją efektów wyrażanych w wartościach fizycznych i wyrażanych w kategoriach pieniężnych kosztów.

Pośrednią kategorią kryteriów według omawianej klasyfikacji jest produktywność. Kryterium to bierze pod uwagę zarówno efekty, jak i koszty. Jego adekwatność wobec specyfiki badań ewaluacyjnych nie jest jednak pełna, nie definiuje się go więc jako pełnowartościowego kryterium oceny. Produktywność uwzględnia bowiem analizę poniesionych nakładów, jednakże zakłada ich bardzo wąskie rozumienie jako kosztów finansowych.

Drugą główną kategorię stanowić będzie skuteczność, rozumiana jednak znacznie szerzej niż zazwyczaj, kiedy to utożsamiana jest z oceną stopnia realizacji bezpośrednich celów ilościowych interwencji. Kategoria ta zawiera w sobie wiele podkryteriów, których cechą wspólną stanowi spektrum oceny skoncentrowane na różnie rozumianych efektach. Zauważmy, iż założenie to spełnione jest zarówno w przypadku kryteriów użyteczności, wpływu, trwałości czy wreszcie skuteczności, rozumianej w węższym, opisanym powyżej sensie.

Kryteria uwzględniające skuteczność osiągania efektów obecne są w każdej chyba ewaluacji. Każda z podkategorii tego kryterium posiada jednak swe unikalne właściwości. Zakres ich stosowalności podlega w związku z tym dającym się wytyczyć ograniczeniom. Zależnościom pomiędzy doбором kryteriów oceny a różnymi wymiarami, wyznaczającymi specyfikę projektów, poświęcona będzie dalsza część artykułu.

## Dobór kryteriów

Wyznaczenie pakietów kryteriów stanowi bardzo ważną część etapu strukturalizacji każdego badania ewaluacyjnego. Pominięcie tego etapu uniemożliwiłoby bowiem ewaluatorowi opracowanie szczegółowych założeń metodologicznych oraz narzędzi badawczych. Wbrew obiegu opinii wybór kryteriów nie jest nigdy bezpośrednim zastosowaniem wytycznych albo wskazań podręczników. Nie istnieją żadne kanoniczne rozwiązania, obowiązujące w odniesieniu do określonych instytucji zamawiających czy też obszarów interwencji. Za każdym razem dobór kryteriów warunkowany jest natomiast przez charakterystykę problemu badawczego, określoną przez ewaluatora na pierwszym etapie pracy.

Zakres oceny warunkują najsilniej następujące aspekty:

- funkcja ewaluacji,
- moment badania,
- zakres badania,
- paradygmat badawczy.

Każdy z wymienionych aspektów we właściwy sobie sposób wpływa na dobór kryteriów. Ostateczny kształt zestawu kryteriów jest wypadkową tych oddziaływań i stanowi o osobliwościach poszczególnych ewaluacji.



**Funkcje ewaluacji** uległy daleko idącemu zróżnicowaniu wraz ze zmieniającym się kontekstem kulturowym. Ewaluacja doświadczała tym samym, podobnie jak nauki społeczne, rewolucji rewidującej fundamentalne założenia metodologiczne. Obecnie wymienia się cztery główne funkcje ewaluacji: poznawczą, konkluzyjną, społeczno-polityczną oraz administracyjną<sup>5</sup>.

W przypadku pierwszej z wymienionych funkcji nacisk położony jest na kryteria oceny, z których wynikają pytania przyczynowo-skutkowe (pytania skoncentrowane na relacji pomiędzy działaniami a uzyskanymi efektami rzeczowymi) lub pytania predyktywne (przewidywania efektów rzeczowych z uwzględnieniem systemu i kontekstu wdrażania). Z potrzebami badawczymi określanymi przez poznawczą funkcję ewaluacji koresponduje kryterium trafności<sup>6</sup>. W najpopularniejszej postaci stosuje się je w ewaluacjach *ex-ante* w celu oceny systemu wdrażania i projektu interwencji. Wówczas kryterium trafności przypisywane są pytania predyktywne, dotyczące najbardziej prawdopodobnych oczekiwanych efektów ocenianego programu. Funkcja poznawcza może być jednak realizowana również poprzez ocenę trafności, prowadzoną w późniejszych etapach cyklu interwencji. Przypisanie ocenie funkcji poznawczych ukierunkowuje wtedy proces badawczy na empiryczną weryfikację przyjętych w trakcie programowania założeń.

Funkcje konkluzyjna oraz formatywna mają charakter instrumentalny, a więc zakładają decyzyjne wykorzystanie wyników. Ewaluacje, w które są wpisane, koncentrują się na efektach ilościowych. Implikują zatem stosowanie kryterium skuteczności, rozumianego jako ocena uwzględniająca stopień realizacji zakładanych celów rzeczowych.

Podobny charakter ma funkcja administracyjna, która stanowi odpowiedź na potrzeby instytucji związane z zarządzaniem publicznym. Wpisanie tej funkcji w proces oceny wiąże się z oczekiwaniem, iż ewaluacja przełoży się bezpośrednio na praktykę administracyjną. Wykorzystywane są wówczas kryteria skuteczności i efektywności. Zastosowanie pierwszego z nich uzasadnione jest podobnymi względami, co w ewaluacjach pełniących funkcje formatywne i konkluzyjne. Kryterium efektywności daje natomiast podstawę do oceny zarządzania środkami publicznymi. Konsekwencją takiej oceny mogą być przesunięcia funduszy w ramach budżetu badanej interwencji. Realokacje stanowią nieraz także konsekwencję ewaluacji pełniących funkcje konkluzyjne, dlatego również w ich przypadku kryterium efektywności znajduje zastosowanie. Jednak tu efektywność rozumiana jest zazwyczaj

<sup>5</sup> Por. Polska Agencja Rozwoju Przedsiębiorczości, *Ocena (ewaluacja) programów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, wydana przez Ministerstwo Finansów w ramach grantu Banku Światowego na wsparcie systemu monitoringu i oceny użycia funduszy pomocowych Unii Europejskiej i innych funduszy pomocowych, Ministerstwo Finansów, Warszawa 2002, <http://archiwum.parp.gov.pl/raport11.html>, s. 18–21.

<sup>6</sup> Trafność rozumiana jest w tym przypadku jako kryterium oceny, uwzględniające dopasowanie celów przypisanych danej interwencji do kontekstu realizacji, a w szczególności do problemów zidentyfikowanych na etapie programowania.

w kategoriach kosztów–korzyści, podczas gdy funkcji administracyjnej lepiej odpowiada efektywność rozumiana jako kalkulacja kosztów–skuteczności. Zastosowanie kryterium skuteczności w takim kształcie stanowi przeważnie odpowiedź na postulat odpowiedzialności społecznej (*accountability*), będącej główną przesłanką konkluzywnych funkcji ewaluacji.

Ostatnia z wymienionych funkcji, czyli funkcja społeczno-polityczna, jest rzadziej wykorzystywana przez instytucje publiczne, częściej natomiast przez organizacje pozarządowe lub niezależne ośrodki analityczno-badawcze (*think tanks*). Funkcja ta wyklucza pełen obiektywizm badawczy, gdyż jej realizacja zakłada promocję określonej idei lub rozwiązania systemowego. Ewaluacje społeczno-polityczne obejmują oczywiście jednakowe rygory metodologiczne, co w przypadku wszystkich badań ewaluacyjnych. Koncepcja badawcza, a co za tym idzie, również dobór kryteriów, podporządkowane są jednak w tym przypadku określonym celom, mają bowiem za zadanie dostarczyć wyników i rekomendacji wpływających na procesy decyzyjne w przyjęty z góry sposób.

**Moment badania** jest wymiarem będącym podstawą chyba najbardziej popularnej klasyfikacji ewaluacji. Podział ten ugruntował się w dokumentach określających warunki wdrażania funduszy strukturalnych. W okresie programowania 2004–2006 zdefiniowano trzy rodzaje ewaluacji: *ex-ante*, *mid-term* oraz *ex-post*<sup>7</sup>.

Przyjmuje się, iż w ewaluacjach wstępnych (*ex-ante*) najbardziej odpowiednim jest **kryterium trafności**. Podstawy takiego twierdzenia są oczywiste, gdyż trafność jako jedyne spośród kryteriów ogólnych nie odnosi się do efektów rzeczywistych, lecz do efektów oczekiwanych. Ocenie w tym przypadku poddaje się założenia, a sama ocena opiera się wyjątkowo silnie na wiedzy eksperckiej ewaluatorów lub ich konsultantów. W konsekwencji, ewaluacje wstępne polegają wyłącznie na formułowaniu prognoz. Należy tu jednak podjąć zasygnalizowaną już myśl o możliwości prowadzenia oceny trafności na różnych etapach cyklu interwencji.

Ocena przyjętych założeń często stanowi element ewaluacji śródkresowej i końcowej. Weryfikacji poddana może zostać wówczas aktualność przyjętej na początku logiki projektu oraz jej adekwatność w obliczu zmieniającego się kontekstu. Ocenie można poddać także zakładane cele bezpośrednie lub ogólne, co nieraz istotnie wzbogaca użyteczność rekomendacji, a zarazem całej ewaluacji.

W ewaluacjach prowadzonych po rozpoczęciu fazy operacyjnej uzupełnieniem oceny efektów rzeczowych może być także **kryterium użyteczności**. W jego przypadku ocena bazuje na opiniach beneficjentów końcowych odnośnie satysfakcji z otrzymanego wsparcia. Kryterium użyteczności nazywa się niekiedy „zwierciadlanym odbiciem kryterium trafności”. Porównanie to,

<sup>7</sup> Rozporządzenie Rady Unii Europejskiej (WE) 1260/1999, art. 40–43.

w moim przekonaniu, jest tylko częściowo uprawnione. Sugeruje bowiem, iż mamy do czynienia z tym samym w istocie kryterium, stosowanym na różnych etapach cyklu interwencji. Tymczasem nie są to kryteria tożsame, pomimo że posiadają wspólną cechę wyróżniającą, jaką jest brak skierowania spektrum oceny na bezpośrednie efekty. Oceny użyteczności i trafności mogą zatem bez wątplenia stanowić odrębne elementy ewaluacji *mid-term* i *ex-post* oraz wspólnie uzupełniać ocenę opartą na efektach rzeczowych, pełniąc zarazem odmienne, komplementarne funkcje. Zastosowanie kryterium trafności byłoby w takich sytuacjach na przykład kontynuacją oceny eksperckiej, zapoczątkowanej w trakcie ewaluacji wstępnej. Kryterium to miałoby przypisane do siebie funkcje poznawcze, których realizacja wymaga zaangażowania obiektywnej wiedzy fachowej, mogącej znaleźć podwójne zastosowanie poprzez jednoczesne wzmacnianie funkcji formatywnych. Ocena użyteczności natomiast ze swej natury wzmacnia funkcje konkluzywne, otwierając drogę dla demokratyzacji kontroli procesu zarządzania publicznego. Wzbogacenie katalogu kryteriów o użyteczność stanowi także instrument przebudowy całościowego paradygmatu administracji według modelu *New Public Management*, który postuluje urynkowanie usług publicznych w celu maksymalizacji satysfakcji ich odbiorców. Dygresja ta jest o tyle istotna, że powszechność ewaluacji w Europie należy uznać za pokłosie reformatorskiego zapału lat osiemdziesiątych, kiedy to postulaty *New Public Management* zyskały szeroko podzielaną akceptację na znacznej części tego kontynentu.

Pomimo silnego wpływu na kształt wielu ewaluacji, jak też roli odgrywanej w kształtowaniu całościowej kultury ewaluacyjnej, kryteria oceny bazujące na założeniach programowych albo opiniach pełnią zazwyczaj rolę uzupełniającą. Ocena prowadzona po rozpoczęciu realizacji interwencji zdominowana jest bowiem przez kryteria odwołujące się do efektów rzeczowych. Zarówno w ewaluacji *mid-term*, jak i *ex-post* wykorzystywane są przeważnie **kryteria skuteczności i efektywności** (oceny efektów na poziomie produktu i rezultatu). Ewaluację końcową można poszerzyć o ocenę efektów na poziomie wpływu. Dlatego też w opracowaniach podręcznikowych podaje się, iż najbardziej typowym **kryterium** dla tego rodzaju ewaluacji jest **trwałość**<sup>8</sup>. Zastosowanie go daje możliwość rozstrzygnięcia tej newralgicznej kwestii, której nie można podjąć na wcześniejszych etapach cyklu. Ocena trwałości pozwala odpowiedzieć na pytanie, czy polityka publiczna w danym obszarze sprzyja rzeczywistemu rozwiązywaniu problemów. Niska trwałość może wskazywać, na przykład, na występowanie zjawiska wyuczonej bezradności (polityka społeczna) lub na brak oczekiwanego rozwoju instytucjonalnego (programy przedakcesyjne, pomoc techniczna). Potrzeba uzyskania tego rodzaju wiedzy była jedną z kluczowych przesłanek, decydujących o systemowym wpisaniu ewaluacji w mechanizmy wdrażania funduszy strukturalnych.

<sup>8</sup> Trwałość rozumiana jest w tym przypadku jako zdolność efektów interwencji do przetrwania po zakończeniu finansowania zewnętrznego.

**Zakres badania** jest następnym czynnikiem silnie determinującym dobór kryteriów oceny. Ewaluacje prowadzone z poziomu instytucji centralnych lub szczebla ponadnarodowego skoncentrowane są przeważnie na bardzo szerokim spektrum zagadnień i poddają ocenie całokształt efektów wygenerowanych przez wiele interwencji. W ewaluacjach programów sektorowych, prowadzonych przez kraje członkowskie UE, zastosowanie znajduje zatem kryterium oddziaływania. Ocena oparta na tym kryterium daje podstawy do wnioskowania o zakresie synergii, a przez to – o sposobie zarządzania polityką strukturalną. Rzetelne i systematyczne wykorzystywanie kryterium oddziaływania w ewaluacjach o dużej skali jest szczególnie wskazane w kulturach administracyjnych naznaczonych „syndromem sektorowości”, czyli brakiem zdolności koordynacyjnych.

W nieco innej formie ocenę oddziaływania stosuje się również w ewaluacjach o mniejszej skali, prowadzonych, na przykład, na szczeblu poszczególnych instytucji wdrażających lub instytucji władzy regionalnej. Wówczas ocena poświęcona jest relacjom pomiędzy wpływem efektów, wygenerowanych na poziomie poszczególnych projektów, a celami ogólnymi danego obszaru polityki publicznej. Zastosowanie kryterium oddziaływania w tego rodzaju ewaluacjach weryfikuje trafność systemowego ulokowania instytucji prowadzących ewaluację. Ocenie może być ponadto poddawana logika interwencji w jej węzłowym punkcie, łączącym sąsiadujące ze sobą poziomy wdrażania.

Kryteriami właściwymi dla ewaluacji o dużej skali są także specyficzne kryteria **horyzontalne**. Ich wykorzystanie ma na celu dokonanie oceny realizacji określonych celów, którym przypisuje się znaczenie wspólne dla całej polityki publicznej. W przypadku dużych ewaluacji polityki strukturalnej UE ocenia się na przykład zakres wartości dodanej, wpływ na wyrównanie szans kobiet i mężczyzn lub wpływ interwencji na środowisko naturalne.

Zakres ewaluacji warunkuje również wykorzystywanie innego rodzaju **kryteriów specyficznych**. W tym przypadku różnice wynikają jednak nie z poziomu ulokowania instytucji realizującej, lecz z zasięgu spektrum badawczego. Wspomniana klasyfikacja ewaluacji, bazująca na momencie badania, wskazuje na ewaluacje o charakterze przekrojowym. Zakres tematyczny oceny bywa jednak ograniczony do ściśle wyznaczonych zagadnień, stanowiących zazwyczaj obszary zidentyfikowane w trakcie wdrażania jako szczególnie problematyczne. Ewaluacje tego rodzaju mają charakter bieżący (*on-going*), a stosowane w nich kryteria odnoszą się do ściśle sprecyzowanych aspektów lub stanowią uszczegółowioną postać kryteriów ogólnych.

Wybór określonego **paradygmatu badawczego** jest w praktyce badań ewaluacyjnych tożsamy z określeniem badanego fenomenu. Specyfikę kryteriów oceny warunkuje zatem wynikające bezpośrednio z przyjętego paradygmatu założenie. Dla przykładu, zastosowanie paradygmatu nieokreślonych celów (*goal-free evaluation*) lub paradygmatu pluralistycznego stanowi bar-

dzo silną przesłankę do skoncentrowania oceny na kryterium użyteczności. Paradygmat przyczynowo-skutkowy natomiast zakłada ocenę opartą przede wszystkim o kryteria skuteczności realizacji celów oraz trafności.

Dotychczas omówiono merytoryczne uwarunkowania doboru kryterium. Wskazać należy także na **uwarunkowania organizacyjne**, które w praktyce ewaluacji bardzo wyraźnie oddziałują na kształt badań. Dobór kryteriów musi bowiem przebiegać zgodnie z profesjonalnymi standardami, wymagającymi, aby każda ewaluacja spełniała między innymi warunki użyteczności i wykonalności. Zestaw kryteriów trzeba zatem dopasować do ograniczeń czasowych i finansowych. Brak uwzględnienia uwarunkowań organizacyjnych rzutuje bezpośrednio na kształt produktu końcowego ewaluacji. W sytuacji występowania istotnych ograniczeń zachodzi konieczność kompromisu w obszarze merytorycznym. Wtedy ostateczny, wytyczony kryteriami zakres oceny powinien być ustanowiony dopiero po zawarciu owego kompromisu. W takim przypadku tylko spełnienie tego założenia pozwoli mówić o „dobrej ewaluacji”, która nie będzie stanowiła dla zamawiających źródła rozczarowań, a dla wykonawców – przyczynku do zaniżania zawodowych standardów.

### **Dobór kryteriów w ewaluacji *ex-post***

Osobliwości ewaluacji *ex-post* w każdym aspekcie, a więc również w zakresie doboru kryteriów, są oczywiście pochodną przede wszystkim specyficznego momentu realizacji. Jak wskazałem w poprzedniej części artykułu, ewaluacje prowadzone po zakończeniu wdrażania interwencji koncentrują się ze swej natury na ocenie efektów rzeczywistych wszystkich poziomów, w tym wpływu.

Pomimo że wspomniana powyżej cecha decyduje o specyfice doboru kryteriów oceny w ewaluacji *ex-post*, nie stanowi ona bynajmniej jedyne wymiaru ukierunkowującego selekcję kryteriów na potrzeby poszczególnych projektów. Dobór kryteriów jest warunkowany przez wszystkie spośród wymienionych uprzednio wymiarów (funkcje ewaluacji, zakres, paradygmat, uwarunkowania organizacyjne).

W europejskiej kulturze badań ewaluacyjnych bardzo silnie zakorzenione jest skojarzenie ewaluacji *ex-post* z kompleksową oceną, prowadzoną przez Komisję Europejską po zakończeniu okresów programowania. Jest to przejaw dominującej roli, jaką w dotychczasowej historii ewaluacji w Europie odgrywa polityka strukturalna Wspólnoty, z wpisanym w system wdrażania obowiązkiem przeprowadzania oceny. O ile jednak znaczenie polityki strukturalnej dla rozwoju ewaluacji i modernizacji zarządzania sprawami publicznymi w krajach członkowskich jest niepodważalne, o tyle „organiczne” zespolenie ewaluacji *ex-post* ze ściśle określonym przez przepisy działaniem należy uznać za niepożądane. Na dalszy plan schodzą bowiem wtedy raczej

praktyczne, które powinny być najważniejszą przesłanką określającą celowość prowadzenia badań ewaluacyjnych oraz ich charakter.

Tematyka i zakres ewaluacji *ex-post* mogą więc przybierać zgoła inną postać niż przewidują zapisy rozporządzeń Rady UE. Uwaga ta odnosi się do wszystkich elementów i etapów procesu oceny, a więc także do doboru kryteriów. Chcąc zilustrować różnice wynikające z odmiennego zakresu i funkcji, przeanalizuję sposób doboru kryteriów w ewaluacjach realizowanych na różnych poziomach systemu wdrażania środków wspólnotowych.

Wspomniałem już, że ewaluacja *ex-post* bywa w kontynentalnej tradycji badawczej utożsamiana z ocenami prowadzonymi przez Komisję Europejską. Specyfika tego rodzaju ewaluacji jest określona przez regulacyjne ramy wdrażania funduszy strukturalnych. Regulacje rozstrzygają także o zakresie ewaluacji i wskazują podmioty odpowiedzialne za przygotowanie i przeprowadzenie oceny. Rozporządzenie Rady Unii Europejskiej (WE) 1260/1999 wskazuje również kryteria, które powinny zostać uwzględnione w ocenie, zaliczając do tej grupy skuteczność realizacji efektów na wszystkich poziomach (produkty, „wykorzystanie środków”, rezultaty, oddziaływanie), efektywność („wydajność pomocy”) oraz kryteria horyzontalne, służące ocenie wpływu polityki strukturalnej UE na wzmocnienie spójności społeczno-gospodarczej w obrębie Wspólnoty<sup>9</sup>.

Zapisy Rozporządzenia definiują charakter „komisyjnej” oceny *ex-post* nie tylko poprzez wskazanie listy kryteriów, lecz także poprzez określenie zakresu i momentu badania. Przedmiot oceny jest w tym przypadku zdefiniowany bardzo szeroko jako całokształt działań objętych finansowaniem funduszy strukturalnych.

Nie zostały natomiast sprecyzowane funkcje specyficzne tej oceny. Wynikają one jednak z pozostałych zapisów i kontekstu prowadzenia ewaluacji. Oczekiwanie, iż ewaluacja dostarczy wiedzy o „czynnikach przyczyniających się do sukcesu lub niepowodzeniu wykonania”<sup>10</sup> wskazuje na funkcję poznawczą. Jeśli w jej realizację wkomponowany ma być element oceny, wówczas katalog kryteriów oceny będzie należało poszerzyć o trafność. Powracamy w tym miejscu do wspomnianego już zagadnienia stosowalności tego kryterium w ewaluacjach prowadzonych po zakończeniu realizacji. Jego zastosowanie w przypadku omawianej oceny *ex-post* jest nie tylko możliwe, ale stanowi wręcz konieczny warunek spełnienia wszystkich oczekiwanych funkcji ewaluacji.

Oprócz funkcji poznawczej, ewaluacji „komisyjnej” przypisane są przede wszystkim funkcje instrumentalne, zwłaszcza funkcja konkluzyjna. Z uwagi na moment badania, funkcja formatywna jest, siłą rzeczy, ograniczona. W obszarze kryteriów dominująca rola funkcji instrumentalnych przejawia

<sup>9</sup> Rozporządzenie Rady Unii Europejskiej (WE) 1260/1999, art. 43.

<sup>10</sup> Ibidem.

się w typowy dla nich sposób, czyli poprzez wyeksponowanie kryterium skuteczności, które wskazywane jest zarówno w zapisach szczegółowych dotyczących ewaluacji *ex-post*<sup>11</sup>, jak i w zapisach ogólnych poświęconych procesowi oceny<sup>12</sup>.

Opisane powyżej uwarunkowania specyficzne dla ewaluacji *ex-post* wpisanej w regulacje, określające ramy wdrażania polityki strukturalnej, implikują także konkretne rozumienie badanego fenomenu, a więc wyraźnie rzutują na logikę operacyjną ewaluacji. W rezultacie, kryteria oceny zostaną na etapie realizacji badania zastosowane w sposób, który wynika z operacyjnej specyfiki „komisyjnej” ewaluacji *ex-post*. Poza obszarem badawczym, wytyczonym przez wspólnotowe regulacje, pozostaje bardzo rozległa przestrzeń, stanowiąca pole dla projektów ewaluacji *ex-post*, komplementarnych wobec działań Komisji Europejskiej. W przypadku tych projektów dobór kryteriów będzie przynajmniej częściowo zbliżony do opisanego powyżej, co wynika bezpośrednio z cech wspólnych dla każdej ewaluacji *ex-post*. Odmierna logika operacyjna sprawi jednak, że kryteria te będą stanowić punkt wyjścia dla odmiennych pytań ewaluacyjnych i problemów badawczych. Zjawisko to ilustrują przypadki z polskich badań ewaluacyjnych. Rodzime instytucje publiczne rozpoczęły prowadzenie oceny *ex-post* na kilka lat przed akcesją, w odpowiedzi na bieżące potrzeby informacyjne, powstałe w trakcie procesu wdrażania. W rezultacie, charakter polskich ewaluacji *ex-post* miał przeważnie mniej przekrojowy charakter niż w przypadku oceny prowadzonej na poziomie ogólnoeuropejskim. Zakres ewaluacji ograniczony był do określonych programów, przez co rzadko wykorzystywano kryterium oddziaływania. Ocena prowadzona według tego kryterium koncentrowała się natomiast na szacowaniu oddziaływania programów na społeczności lokalne lub regionalne, nie opierając się na analizach makrostrukturalnych.

Jedną z instytucji prowadzących w Polsce ewaluację *ex-post* w okresie przedakcesyjnym była Polska Agencja Rozwoju Przedsiębiorczości (instytucja wdrażająca program Phare Spójność Społeczno-Gospodarcza)<sup>13</sup>. Dobór kryteriów w ewaluacjach *ex-post* realizowanych przez PARP był warunkowany przez funkcję konkluzywną<sup>14</sup>. Oceny zdominowały zatem kryteria szeroko rozumianej skuteczności, w tym, typowa dla ewaluacji *ex-post*, trwałość. Ze względu na duże znaczenie przypisywane funkcji konkluzywnej, ewaluacje PARP wykorzystywały w przeważającej części kryterium użyteczności. Podejście badawcze odzwierciedlało elementy paradygmatu pluralistycznego oraz nosiło cechy *New Public Management*. W tym przypadku obranie takiej

<sup>11</sup> Ibidem.

<sup>12</sup> Ibidem, art. 40.

<sup>13</sup> Raporty ewaluacyjne PARP są dostępne na stronach internetowych Polskiej Agencji Rozwoju Przedsiębiorczości: <http://www.parp.gov.pl/index/index/110>.

<sup>14</sup> Zawiera się w niej orientacja na rezultaty, uwzględnienie głównie celu instrumentalnego/kontrolnego i perspektywy punktowej (badanie beneficjentów rok po zakończeniu interwencji).

perspektywy wynikało w dużej mierze z uwarunkowań organizacyjnych. Rola instytucji utrzymującej bezpośrednie kontakty z końcowymi odbiorcami nadawała bowiem dużą wagę kwestiom jakości produktów dostarczanych beneficjentom i zadowolenia beneficjentów z tych produktów.

PARP w ewaluacjach końcowych wykorzystywała również kryterium trafności. Jego zastosowanie podyktowane było jednak nieco innymi potrzebami poznawczymi niż w ewaluacjach na poziomie europejskim. Trafność nie odnosiła się bowiem przede wszystkim do szerokiego kontekstu wdrażania, lecz do założeń programowych, co stanowi zazwyczaj obszar oceny w ramach ewaluacji *ex-ante*.

Odmienne uwarunkowania odnajdziemy w przypadku ewaluacji *ex-post* pomocy przedakcesyjnej, realizowanych przez Urząd Komitetu Integracji Europejskiej<sup>15</sup>. Ich zakres był bowiem określony przez regulacje dotyczące systemu wdrażania Phare. Zastosowano tu podejście podobne do opisanego wcześniej podejścia „komisyjnego”. Przybrało ono zatem charakter przekrojowy, wyrażony w akcentach położonych na pogłębioną ocenę skuteczności realizacji celów operacyjnych i bezpośrednich oraz oddziaływanie. Na poziomie kryteriów ujawniało się to chociażby dzięki zastosowaniu kryteriów tematycznych, odnoszących się do realizacji głównych celów programu.

Jak pokazują omówione powyżej przypadki, dobór kryteriów na potrzeby ewaluacji *ex-post* posiada swoją różnorodnie warunkowaną dynamikę. Sam moment przeprowadzania oceny, będący cechą definicyjną ewaluacji *ex-post*, wywiera w tym obszarze bardzo silny wpływ, stanowiąc o specyfice doboru kryteriów w ewaluacjach końcowych. Jest to jednak tylko jeden z wielu czynników decydujących o wyborze określonych kryteriów oceny *ex-post*. W przypadku tego rodzaju ewaluacji zastosowanie znajdują bowiem wszystkie mechanizmy selekcji, oparte o wymienione wcześniej kluczowe wymiary, natomiast twierdzenia o ściśle wyznaczonych granicach oceny końcowej są chybione.

## Zakończenie

Kryteria stanowią swoisty ewaluacyjny „rdzeń tożsamości”. Określają bowiem najważniejsze charakterystyki każdej oceny. Sposoby ich selekcji decydują więc w bardzo dużym stopniu o powodzeniu całego procesu badawczego. Dlatego też można założyć, że trafny dobór kryteriów jest koniecznym warunkiem „dobrej ewaluacji”. Problem praktyczny polega jednak na tym, iż zarówno trafny schemat doboru kryteriów, jak i dobra ewaluacja nie poddają się precyzyjnym definicjom.

---

<sup>15</sup> Raporty ewaluacyjne dostępne są na stronach internetowych: <http://www1.ukie.gov.pl/www/serce.nsf/DocByWykazZmian/5142FF4BF2348620C1256ED9004B0299?open>.



Znaczna część praktyków ewaluacji przyjmuje stanowisko, iż poszukiwanie uniwersalnych przepisów stanowi w dziedzinie ewaluacji przedsięwzięcie równie jałowe, co próby wypracowania ogólnej definicji dobrego dzieła sztuki. Specyfika ewaluacji jest bowiem nie mniej zróżnicowana niż trendy estetyczne czy nurty literackie. Sprowadzenie wszystkich ewaluacji do wspólnego mianownika wymagałoby więc zastosowania pewnego **meta-kryterium**, abstrahującego od szczegółowych założeń przyjmowanych w poszczególnych przypadkach. Nawiązując do nakreślonych już ram ogólnej logiki ewaluacji, można uznać, iż rolę meta-kryterium mogłaby pełnić **zgodność z celem** (*purpose-fit*).

Ocena zgodności z celem znajduje swoje pierwsze zastosowanie bezpośrednio po zakończeniu doboru kryteriów. Trafność tego doboru może wówczas zostać zweryfikowana poprzez odniesienie do celów, wyznaczonych po zakończeniu etapu określania problemu. Jest to zarazem drugi z krytycznych punktów, wyznaczających miejsce kryteriów w procesie ewaluacyjnym. Wypełnienie tych newralgicznych punktów właściwą „tkanką” wymaga zaniechania aplikowania schematycznych rozwiązań i zastąpienia tego podejścia intensywną pracą ewaluatora, bazującą na przemyślanej i uświadomionej koncepcji produktu końcowego.

Poruszając kwestię świadomości ewaluatora odnośnie koncepcji końcowej produktu ewaluacji, należy pamiętać, iż właściwie funkcjonujący system ewaluacji opiera się na umiejętnej koordynacji działań prowadzonych przez poszczególne komórki ewaluacyjne. Wizja oceny w każdym z przypadków powinna zatem uwzględniać plany badawcze partnerów. Powstaje w związku z tym potrzeba intensywnego, niesformalizowanego dialogu, w ramach którego dyskusja nad doбором kryteriów musi stanowić wyjściowy etap wspólnego planowania strategii ewaluacyjnej. Wkładem w tę dyskusję jest niniejszy artykuł.

---

**Maciej Szalaj** – absolwent Instytutu Socjologii Uniwersytetu Warszawskiego. Pracuje w Sekcji Ewaluacji Polskiej Agencji Rozwoju Przedsiębiorczości. Problematyką badań ewaluacyjnych zajmuje się od czterech lat. Specjalizuje się w ewaluacji programów ukierunkowanych na wspieranie rozwoju kapitału ludzkiego. Kierownik projektów badawczych oraz współautor ich koncepcji i metodologii. Posiada doświadczenie w zakresie ewaluacji programów przedakcesyjnych i funduszy strukturalnych. Członek Grupy Sterującej Ewaluacją SPO RZL. Brał udział w pracach nad systemem wskaźników efektów dla Programu Operacyjnego „Kapitał ludzki”. Obecnie zaangażowany w tworzenie systemu ewaluacji na poziomie Priorytetu II PO KL oraz koordynację procesu wieloetapowej ewaluacji sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego. Autor opracowań poświęconych efektom programów współfinansowanych ze środków wspólnotowych, m.in. studium przypadku poświęconego roli środków przedakcesyjnych w polityce regionalnej Podkarpacia.

**Bibliografia**

- Fourier D.M., „Establishing evaluative conclusions: a distinction between general and working logic”, w: E. Stern (red.), *Evaluation Research Methods*, vol. 1, Sage Publications, London 2005.
- Polska Agencja Rozwoju Przedsiębiorczości, *Ocena (ewaluacja) programów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, wydana przez Ministerstwo Finansów w ramach grantu Banku Światowego na wsparcie systemu monitoringu i oceny użycia funduszy pomocowych Unii Europejskiej i innych funduszy pomocowych, Ministerstwo Finansów, Warszawa 2002, <http://archiwum.parp.gov.pl/raport11.html>.
- Polskie Towarzystwo Ewaluacyjne, *Ewaluacja funduszy strukturalnych*, Opracowanie na zlecenie Wydziału Monitoringu i Raportowania ZPORR, Departament Wdrażania Programów Rozwoju Regionalnego, Ministerstwo Gospodarki, Pracy i Polityki Społecznej, Warszawa 2003.

*Agnieszka Sternik*

## **Funkcja autoedukacyjna ewaluacji w rozwoju organizacji. Analiza rozwiązań międzynarodowych**

### **Wstęp**

Monitoring, a w szczególności ewaluacja dostarczają wiedzy, która, o ile zostanie zaakceptowana i następnie zinternalizowana przez bezpośrednich uczestników badania, może wspomóc wewnętrzne procesy uczenia się, a w konsekwencji wdrażanie programów publicznych w organizacjach i zarządzanie nimi. Procesy te powinny być jednakże zintegrowane z całym cyklem programowym poprzez włączenie do niego efektywnego systemu informacji zwrotnej. Dodatkowym warunkiem, niezbędnym do tego, aby wyniki ewaluacji i płynące z nich wnioski oraz rekomendacje mogły być w pełni wykorzystywane, jest udostępnienie tej informacji w odpowiedniej formie właściwej grupie odbiorców. Upowszechnianie wyników ewaluacji odnosi się zatem do działań, dzięki którym uzyskana w procesie oceny wiedza zostaje udostępniona wszystkim zainteresowanym podmiotom.

Analiza i wykorzystanie rekomendacji wynikających z przeprowadzonej ewaluacji powinny stanowić element każdej oceny, zwłaszcza wówczas, kiedy wskaźniki realizacji programu nie są w pełni zadowalające. Otwartość na wyniki badań sprzyja bowiem doskonaleniu w organizacji wewnętrznych procesów przygotowywania i zarządzania podobnymi programami w przyszłości. Jednym z największych wyzwań ewaluacji pozostaje zatem, oprócz analizy danych oraz sformułowania na ich podstawie wniosków i rekomendacji, także praktyczne wykorzystanie ich przez zamawiających badania.

Celem tego artykułu jest wskazanie rozwiązań, które umożliwią bardziej efektywne wykorzystanie wyników ewaluacji w procesach podejmowania decyzji i uczenia się organizacji. W tym kontekście należy nawiązać do podstawowych zasad prowadzenia ewaluacji, sporządzania raportów oraz formułowania wniosków i rekomendacji. Poniższa analiza będzie odwoływać się do rozwiązań stosowanych przez organizacje międzynarodowe w tym zakresie.

## Definicja pojęć

Od początków ewaluacji w latach siedemdziesiątych ubiegłego stulecia ten typ badań zyskał sobie szereg definicji, teoretycy zaś podejmowali liczne próby typologizacji różnorodnych podejść ewaluacji. Nie zagłębiając się w tę dyskusję, należy wspomnieć o jednej z klasyfikacji, która rozróżnia dwa rodzaje badań: ewaluację formatywną i podsumowującą. Pierwsza, prowadzona w trakcie realizacji programu, ma na celu głównie ulepszenie procesu wdrażania i zarządzania interwencją; druga pełni niejako rolę rozliczeniową zarówno w aspekcie finansowym, jak i rzeczowym, analizując rezultaty interwencji. Rozróżnienie to jest o tyle ważne, że wiele z zagadnień poruszonych w tym artykule odwoływać się będzie do specyfiki tych ewaluacji.

Badanie ewaluacyjne ma wspierać instytucje lub organizacje w poprawianiu prowadzonych przez nie polityk, programów i projektów. Ewaluacje zarówno wspomagają proces definiowania założeń interwencji (zwłaszcza określanie jej priorytetów i celów) oraz ułatwiają alokację zasobów, jak i relacjonują rezultaty przeprowadzonych działań<sup>1</sup>. Z tego względu kluczową sprawą pozostaje optymalizowanie wykorzystania wniosków prezentowanych w raportach ewaluacyjnych oraz wykreowanie pozytywnego środowiska dla odbioru także tych badań, które wskazują na ewidentne problemy w zarządzaniu programami. Nieodłącznym elementem ewaluacji jest też opracowanie, na podstawie przeprowadzonej analizy danych, rekomendacji dla instytucji uczestniczących w realizacji interwencji.

Próbując zdefiniować rekomendacje można stwierdzić, iż stanowią one propozycję dotyczącą sposobów doskonalenia opracowywania ocenianych polityk, programów lub projektów. Zadaniem rekomendacji jest zatem wskazanie konkretnych działań, które mogą prowadzić do modyfikowania dotychczasowych sposobów prowadzenia interwencji i przygotowania nowych, bardziej efektywnych programów w przyszłości. Dobrze opracowane rekomendacje mają charakter operacyjny, a także przewidują, jak i przez kogo będą wdrażane. Ponadto mogą zawierać w sobie pewną hierarchię ważności lub przyjmować formę zestawu opcji, określając jednocześnie konsekwencje wyboru każdej z nich<sup>2</sup>. Proste rekomendacje, zwłaszcza w przypadku złożonych i angażujących wiele podmiotów programów, nieraz okazują się niewystarczające, a wtedy staje się konieczne sięgnięcie przez ewaluatorów po jedno ze wspomnianych wyżej rozwiązań.

<sup>1</sup> European Commission, „Introduction”, w: *Evaluating EU Activities*, DG Budget 2005.

<sup>2</sup> European Commission, *Guidelines for Dissemination and Feedback of Evaluations*, DG AIDCO, Brussels 2003.

## Wykorzystywanie informacji zwrotnych i rekomendacji w praktyce

Kwestię wykorzystywania wyników ewaluacji, w szczególności zaś wdrażania formułowanych na jej podstawie rekomendacji, poruszają coraz częściej zarówno akademicy, jak i praktycy ewaluacji. Tymczasem jeszcze nie tak dawno, bo w 2001 roku, w brytyjskim kwartalniku *Evaluation* ukazał się artykuł Basil Cracknell, w którym autorka, powołując się na wyniki swoich badań, wskazała, iż w latach 1995–2001 tylko jeden artykuł zamieszczony w *Evaluation* dotyczył wykorzystania informacji zwrotnej płynącej z ewaluacji<sup>3</sup>.

W środowisku organizacji międzynarodowych zagadnienie to nader często rozpatruje się w kontekście zarządzania przez wiedzę (*knowledge management*) i rozwijania organizacji uczących się (*learning organizations*)<sup>4</sup>. Analizując te procesy Bank Światowy proponuje rozróżnienie dwóch podejść: *learning from doing* oraz *learning from evaluation*, zwracając uwagę na potrzebę wprowadzenia mechanizmów umożliwiających doskonalenie organizacji na podstawie wiedzy dostarczonej zarówno ze źródeł zewnętrznych, jak i wewnętrznych<sup>5</sup>. Zjawiskiem, które z pewnością sprzyja wzmocnieniu roli ewaluacji, jest stopniowe odchodzenie od traktowania jej wyłącznie jako narzędzia rozliczeniowego (*accountability*) i przyjmowanie przez organizacje opartego na wiedzy podejścia do zarządzania. W praktyce oznacza to, że organizacje, używające dotychczas ewaluacji do legitymizacji potrzeb i dla zgodności z procedurami prowadzonych przez siebie programów, mogłyby zacząć postrzegać raporty jako źródło dodatkowej wiedzy, która może usprawnić zarządzanie tymi inicjatywami. Jednocześnie istnieje wiele czynników skutecznie uniemożliwiających uczenie się organizacji. Wśród nich wymienić należy kulturę organizacji, nacisk na osiąganie celów, brak miejsca na refleksję i wyciąganie wniosków z dotychczasowych doświadczeń, brak zachęty do nauki, rotacje pracowników, brak pamięci organizacji<sup>6</sup>.

Przeprowadzony w latach dziewięćdziesiątych przez OECD systematyczny przegląd działań ewaluacyjnych w różnych organizacjach międzynarodowych prowadzących politykę rozwojową pokazał, że tylko nieliczne z nich posiadały jakiegokolwiek procedury umożliwiające pełne wykorzystanie informacji zwrotnej płynącej z badań<sup>7</sup>. Podobną sytuację stwierdzono w wielu jednostkach administracji publicznej w państwach członkowskich Unii Europejskiej,

<sup>3</sup> Cracknell B., „The role of aid-evaluation feedback as an input into the learning organization”, *Evaluation* 2001, nr 7(1), s. 133.

<sup>4</sup> Ibidem.

<sup>5</sup> Davies H., McDermott R., *How Knowledge Management Affects the Bank's. Change Management Programme*, ICD 1996.

<sup>6</sup> OECD, „Evaluation feedback for effective learning and accountability”, *Evaluation and Aid Effectiveness* 5, OECD 2001.

<sup>7</sup> Ibidem.

z których zaledwie kilka wprowadziło konkretne zasady postępowania w momencie przyjmowania raportów ewaluacyjnych. Przegląd ujawnił, że raporty czytało zazwyczaj tylko niewielkie grono pracowników komórek zlecających badania w poszczególnych instytucjach. Dominującą praktyką w zespołach zamawiających ewaluacje było natomiast przekazywanie całych raportów jednostkom zajmujących się analizowanym programem oraz informowanie pozostałych aktorów jedynie za pomocą skrótego zestawienia wniosków i rekomendacji. Co więcej, nawet jeśli z raportem zapoznawała się szersza grupa odbiorców, brakowało procedur, które ułatwiałyby przełożenie przedstawionych w nim rekomendacji na konkretne działania. Toteż pytanie, czy owe rekomendacje były wdrażane, czy też nie, odgrywało zdecydowanie drugorzędną rolę.

Z drugiej jednak strony, wspomniany przegląd praktyk europejskich umożliwił również zidentyfikowanie państw, w których wykorzystywano raporty ewaluacyjne w sposób bardziej systematyczny. Wśród nich znalazły się Wielka Brytania, Holandia i Niemcy, kraje o rozwiniętej kulturze ewaluacyjnej. Autorzy publikacji zwrócili uwagę na czynniki wyróżniające organizacje działające w wymienionych krajach; jednym z podstawowych czynników był fakt, że zespoły odpowiadające za ewaluację miały ściśle określone zadania. Zobowiązywano je nie tylko do przekazywania wszystkich raportów ewaluacyjnych instytucjom zarządzającym programami, lecz także do włączania tych instytucji w proces podejmowania decyzji o tym, które z rekomendacji mają być wdrożone, w jakim zakresie, terminie i przez kogo. I tak na przykład w programach pomocowych w Niemczech i Holandii minister odpowiedzialny za politykę rozwojową zobowiązywał się wobec parlamentu do podjęcia konkretnych decyzji w oparciu o wyniki ewaluacji oraz raportowania postępów ich wdrażania w ściśle określonym czasie. W Wielkiej Brytanii, w ramach ministerstwa zajmującego się pomocą rozwojową (Departament for International Development, czyli DfID) powołany został specjalny komitet (Project and Evaluation Committee), zajmujący się zatwierdzaniem rekomendacji do wdrożenia. Na podstawie otrzymanego raportu ewaluacyjnego oraz krótkiej notatki zawierającej listę własnych rekomendacji, sformułowanych przez podmiot zlecający badanie na podstawie rekomendacji przedstawionych przez ewaluatorów, komitet podejmował decyzję o przyjęciu rekomendacji. Ponadto jednostki i ministerstwa wskazane przez Komitet miały co sześć miesięcy raportować postęp we wdrażaniu rekomendacji.

### **Wpływ kontekstu instytucjonalnego, podmiotów uczestniczących w ocenie oraz typu badania na wykorzystanie wyników ewaluacji**

Andreas Balthasar analizował na podstawie dziesięciu studiów przypadków badań ewaluacyjnych przeprowadzonych w Szwajcarii, w jaki sposób

zamawiający korzystają z informacji zwrotnej płynącej z ewaluacji<sup>8</sup>. Autor wybrał badania, które osadzone były w odmiennych kontekstach instytucjonalnych i charakteryzowały się zróżnicowanym dystansem pomiędzy ewaluatorami a ewaluowanymi. Wykorzystując metodę *Qualitative Comparative Analysis*, autor badał, czy dystans pomiędzy ewaluatorem a podmiotami ewaluowanymi ma wpływ na to, w jaki sposób wykorzystywane są wyniki ewaluacji. Analiza ta skłoniła do wyprowadzenia wniosku, iż w niektórych sytuacjach zarówno kontekst, jak i dystans oraz typ badania mają wpływ na wykorzystanie ewaluacji. Na tej podstawie Balthasar zaproponował cztery możliwe aspekty procesu ewaluacji.

Po pierwsze, ewaluację wykorzystuje się w sposób **instrumentalny**, kiedy zachodzi konieczność podjęcia konkretnej decyzji lub rozwiązania jakiegoś problemu przez organizację. Dzięki bezpośredniemu stosowaniu się badaczy do rekomendacji, ewaluacja staje się wyznacznikiem kierunku, w którym powinno podążać zarządzanie lub wdrażanie programu. Dlatego instrumentalne wykorzystanie ewaluacji ma miejsce zwłaszcza w przypadku ewaluacji formatywnych i niezależnie od wielkości dystansu pomiędzy wykonawcami a podmiotami ewaluowanymi. W ewaluacjach podsumowujących działa ono pod warunkiem zachowania dużego dystansu pomiędzy zamawiającym a wykonawcą.

Po drugie, zamawiający ewaluację często definiują ją jako czynnik lepszego zrozumienia procesów zachodzących w organizacji – czynnik pogłębiający ich analizę. Takie wykorzystanie zdefiniowane zostało jako podejście **konceptyjne**. Mamy z nim do czynienia w przypadku badań łączących w sobie charakterystyczne cechy ewaluacji podsumowujących i ewaluacji formatywnych oraz wtedy, kiedy ewaluacje mają charakter formatywny i prowadzone są w bliskiej współpracy z osobami zarządzającymi ocenianymi programami.

Po trzecie, ewaluacje wykorzystywane są w sposób **symboliczny**, kiedy instytucje zlecające przeprowadzenie ewaluacji szukają przede wszystkim potwierdzenia słuszności podjętych decyzji i kierunków działania, mniejsze zaś znaczenie przywiązują do możliwości wykorzystania dostarczonej wiedzy do ulepszania dotychczasowej polityki i prowadzonych przez siebie projektów. Niestety, badanie nie pozwoliło Balthasarowi na przeprowadzenie pogłębionej analizy tego aspektu, gdyż liczba badanych przypadków nie była wystarczająca. Ostatnia kategoria to wykorzystanie ewaluacji **zorientowane na proces**, w którym szczególną rolę odgrywa rozwijanie zdolności organizacji do uczenia się.

Na podstawie przeprowadzonej analizy autor wnioskuje, że – niezależnie od dystansu pomiędzy ewaluatorem a osobami zarządzającymi ocenianymi programami – w analizowanych przypadkach najczęściej wykorzystywano

---

<sup>8</sup> Balthasar A., Rieder S., „Learning from evaluation. Effects of the evaluation of the Swiss Energy 2000 Programme”, *Evaluation* 2000, nr 6(3), s. 245–260.

wyniki ewaluacji o charakterze formatywnym. Wyniki ewaluacji podsumowujących wykorzystywano tylko wtedy, kiedy instytucje miały już doświadczenie w postępowaniu z ewaluacjami.

### **Czynniki wpływające na skuteczność wykorzystywania informacji zwrotnej i rekomendacji formułowanych w procesie ewaluacji**

Wielu autorów podejmowało w swoich publikacjach temat wykorzystywania informacji zwrotnej i rekomendacji, formułowanych w procesie ewaluacji polityk, programów i projektów. Andreas Balthasar analizował typy praktycznego wykorzystywania ewaluacji ze względu na zakres i sposób prowadzenia badań oraz ich kontekst instytucjonalny<sup>9</sup>. Patton tymczasem podkreślał, że opracowanie nawet najbardziej skutecznych metod dostarczania informacji zwrotnej nie będzie miało większego wpływu na ich faktyczne wykorzystanie, jeśli jakość samych ewaluacji okaże się niezadowolająca<sup>10</sup>. W jednej ze swoich publikacji wymienił on trzy warunki, których spełnienie jest niezbędne, aby można było uznać ewaluację za udaną, w tym: skrupulatne planowanie badania (również w aspekcie czasowym), identyfikacja odbiorców wyników badania oraz efektywne raportowanie<sup>11</sup>.

Głębiej analizując te zagadnienia, można nawiązać do pracy Balthasara, według którego szczególne znaczenie ma kontekst instytucjonalny, w jakim prowadzona jest ewaluacja. Każda instytucja publiczna realizująca programy i projekty oraz zamawiająca ich ewaluacje posiada własne, wypróbowane metody działania, precyzujące rolę oraz sposób wykorzystania ewaluacji w procesie zarządzania. Praktyki te określają mandat zlecających badanie, a tym samym definiują jego kontekst instytucjonalny. Co więcej, ważne jest również to, na ile organizacja zlecająca badanie była wewnętrznie przekonana o potrzebie ewaluacji i płynących z niej korzyściach, a zwłaszcza, czy miała wsparcie ze strony wyższych urzędników i managerów. Niemałe znaczenie ma także zakres prowadzonej ewaluacji, czyli określenie, jakie aspekty interwencji zostaną poddane analizie. Wykorzystywanie wyników ewaluacji w dużym stopniu zależy bowiem od obszaru i rodzaju analizowanych problemów oraz od tego, czy dotyczą one jedynie wąskiej grupy osób/organizacji, czy też większej liczby odbiorców.

W literaturze podkreśla się wagę analizy potencjalnych odbiorców wyników badania jeszcze przed rozpoczęciem samego procesu badawczego<sup>12</sup>.

<sup>9</sup> Balthasar A., „The effects of institutional design on the utilization of evaluation, evidence using qualitative comparative analysis (CQA)”, *Evaluation* 2006, nr 12(3), s. 355.

<sup>10</sup> Patton M.Q., *Utilisation-Focused Evaluation: The New Century Text*, Sage Publications, Beverly Hills 1997.

<sup>11</sup> Ibidem.

<sup>12</sup> Ibidem.



Ustalenie grupy odbiorców, zdaniem Pattona, pozwala na dostosowanie przekazu, działań oraz kanału komunikacji do ich potrzeb. Dodatkowe metody powinny być brane pod uwagę w przypadku grup, które są trudno dostępne<sup>13</sup>. Szczególne znaczenie ma również informowanie o prowadzonych badaniach osób odpowiedzialnych za podejmowanie decyzji w organizacjach. Wysokiej jakości raport ewaluacyjny to produkt zorientowany na klientów. Tymczasem dość często zarzuca się rekomendacjom zbytnią ogólność lub, przeciwnie, nadmierną szczegółowość, co sprawia, że nie spełniają stawianych przed nimi wymagań. W niektórych organizacjach do oceny jakości raportu i rekomendacji angażowana jest specjalna grupa lub komitet złożony z wyższych urzędników; otrzymują oni raport i decydują, które z zawartych w nim rekomendacji powinny zostać wdrożone, a następnie sprawdzają, czy rzeczywiście zostały one wykorzystane<sup>14</sup>.

W wielu publikacjach autorzy podkreślają, że nie mniej ważny jest sposób organizacji całości procesu ewaluacji niż metoda prowadzenia poszczególnych analiz. Należy się tu odwołać do standardów jakości, które jednoznacznie wskazują, iż dobra ewaluacja powinna jednocześnie wypełniać kryteria użyteczności, staranności oraz prawidłowości metodologicznej<sup>15</sup>. Zasadnicze znaczenie ma zatem projekt i realizacja badania w praktyce oraz sposób organizacji dyskusji nad wnioskami i rekomendacjami z raportu, natomiast decyzje dotyczące zakresu warunków zamówienia, metodologii i ram czasowych badania należy podejmować z myślą o jego późniejszym wykorzystaniu. W wytycznych wielu instytucji (m.in. Komisji Europejskiej) podkreśla się, iż sfinalizowanie raportu końcowego to dopiero połowa sukcesu w procesie ewaluacyjnym, a kluczowa jego część zaczyna się faktycznie po opublikowaniu raportu<sup>16</sup>. Zachowanie wysokiej jakości raportu procesu ewaluacji wymaga także dopilnowania, by formułowane rekomendacje umocowane były w przeprowadzonej analizie i odnosiły się bezpośrednio do wniosków postawionych przez ewaluatorów. Zapewnienie rzetelności danych, wiarygodności analizy i uprawnionych wniosków stanowi absolutną podstawę do korzystania z rekomendacji płynących z badań.

Mając to na względzie, najważniejszym wydaje się właściwe określenie czasu, w którym ewaluacja jest przeprowadzana, a jej wyniki dostarczane zainteresowanym odbiorcom. Z uwagi na – najczęściej długotrwały – proces badawczy, dobrze byłoby tak zsynchronizować go z programem, aby wnioski ewaluacyjne i rekomendacje wciąż mogły zostać użyte jako dodatkowe źródło informacji przy podejmowaniu decyzji w sprawie kształtu obecnych lub

<sup>13</sup> European Commission, *Guidelines...*, op.cit.

<sup>14</sup> UKIE, *Podręcznik ewaluacji okresowej Programu PHARE i Transition Facility*, UKIE 2006.

<sup>15</sup> Patton M.Q., „Discovering process use”, *Evaluation* 1998, nr 4(2), s. 225–233.

<sup>16</sup> European Commission, *Responding to Strategic Needs: Reinforcing the Use of Evaluation*, Communication to the Commission from Ms Grybauskaitė in agreement with the President, Brussels 2006.

przyszłych interwencji. Stąd potrzeba stosowania wieloletnich planów ewaluacyjnych wszędzie tam, gdzie jest to możliwe. W praktyce w wielu przypadkach ewaluację przeprowadza się o wiele za późno i rekomendacje przestają być użyteczne. Jednocześnie należy pamiętać, iż ewaluacje podsumowujące (*ex-post*) przeprowadzane są po całkowitym zakończeniu realizacji programu, a w przypadku projektów inwestycyjnych nawet po dziesięciu latach od formalnego zamknięcia pomocy. Często dopiero tak odległa perspektywa czasowa umożliwi analizę faktycznego znaczenia projektu.

Nie mniej istotną rolę odgrywają ewaluatorzy i to, w jaki sposób przygotowują oni raport ewaluacyjny, gdyż jego forma w dużej mierze wpływa na ogólną użyteczność badania dla managerów<sup>17</sup>. Praktyka pokazuje, iż koniecznym jest wyjście poza tradycyjne metody prezentacji wyników ewaluacji w postaci wielostronicowego raportu. Raporty nie stanowią już wyłącznego lub jedyne dokumentu finalizującego proces badawczy, ponieważ aby zwiększyć oddziaływanie badania i konkretnych rekomendacji często są uzupełniane dodatkowymi materiałami szkoleniowymi, warsztatami lub prezentacjami audiowizualnymi. Należy dbać także o to, by język raportów był przejrzysty i łatwy w odbiorze, wolny od żargonu specyficznego chociażby dla Funduszy Strukturalnych UE.

## Adresaci ewaluacji

Konieczność wypracowania wieloaspektowego podejścia do przekazywania informacji zwrotnej oznacza w praktyce poszukiwanie takich rozwiązań, które ułatwiają przekazywanie ważnej dla beneficjentów i potrzebnej im informacji<sup>18</sup>. Komisja Europejska w swoich materiałach podkreśla, że pierwszym krokiem na drodze do skutecznego upowszechniania wyników ewaluacji jest zdefiniowanie, już na etapie przygotowywania specyfikacji badania, jego adresatów<sup>19</sup>. Także Polskie Towarzystwo Ewaluacyjne zaleca, aby cały proces badawczy zaczynać od sporządzenia listy potencjalnych odbiorców raportu<sup>20</sup>. Przed rozpoczęciem ewaluacji należy odpowiedzieć na pytanie, kto jest adresatem raportu, jakie są jego potrzeby informacyjne, jaką posiada wiedzę na temat ewaluacji, czy był zaangażowany w projektowanie badania, jakie może odnieść korzyści z badania, czy pytania i zagadnienia ewaluacyjne są dla niego interesujące. Świadomość tego pozwala różnicować sposób przekazywania informacji, dostosowując go do charakterystyk odbiorców.

<sup>17</sup> OECD, „Evaluation feedback for effective learning and accountability...”, op.cit.

<sup>18</sup> Cracknell B., „The role of aid-evaluation feedback...”, op.cit.

<sup>19</sup> European Commission, „Introduction”, op.cit.

<sup>20</sup> Polskie Towarzystwo Ewaluacyjne, Ewaluacja Funduszy Strukturalnych, opracowanie na zlecenie Wydziału Monitoringu i Raportowania ZPORR, Departament Wdrażania Programów Rozwoju Regionalnego, Ministerstwo Gospodarki, Pracy i Polityki Społecznej 2003.

W zależności od typu ocenianej interwencji wyniki ewaluacji mogą być interesujące dla osób decydujących o głównych kierunkach polityki, osób odpowiedzialnych za operacyjną stronę interwencji, innymi słowy, za zarządzanie i wdrażanie programów, beneficjentów programów oraz wszystkich innych uczestników życia publicznego, którzy zainteresowani są realizacją konkretnych programów. Pierwsza grupa odbiorców to parlament, urzędnicy ministerstw, zarządy agencji. Holandia jest przykładem kraju, w którym wyniki ewaluacji rządowych programów pomocowych prezentowane są właściwemu ministrowi, a następnie, wraz z notatką będącą niejako formalną odpowiedzią właściwego ministerstwa na wnioski z badania, przesyłane do dyskusji do parlamentu.

Kolejną grupą odbiorców są media. Praktyka europejska pokazuje jednakże, iż wciąż nieliczne jednostki zlecające ewaluacje blisko współpracują z mediami. Dominuje poczucie braku kontroli nad informacjami przekazywanymi przez media i obawa przed selekcją tego, co zaprezentowano w raporcie końcowym. Dlatego dobrą praktyką wydaje się stosowane w niektórych krajach opracowywanie informacji prasowych oraz, w przypadku kontrowersyjnych projektów, uzupełnienie ich konferencją prasową, która umożliwia wyjaśnienie wszelkich wątpliwości dotyczących wyników ewaluacji. Niejako za wzór takiego działania może posłużyć sposób przedstawienia wyników ewaluacji programów finansowanych z Funduszy Strukturalnych Unii Europejskiej w Walii w 2005 roku. Na początku stycznia 2006 roku w BBC1 odbyła się debata z udziałem przedstawicieli administracji oraz ewaluatorów; jej celem było omówienie wyników ewaluacji i dyskusja na temat wykorzystania proponowanych rekomendacji dla ulepszenia dotychczasowej polityki spójności w Walii.

Jednymi z głównych aktorów są organizacje pozarządowe i instytucje dialogu publicznego. Zazwyczaj zarówno prowadzą one ewaluacje realizowanych przez siebie projektów, jak i są ewaluacją objęte. Zdaniem specjalistów tworzenie dogodnych warunków współpracy pomiędzy nimi a podmiotami zamawiającymi ewaluacje wspomaga tak funkcję rozliczeniową, jak i rozwojową ewaluacji. Ostatecznie, także szeroko rozumiana opinia publiczna postrzegana jest jako typowa grupa docelowa, którą informuje się o wynikach ewaluacji.

## **Wybór kanałów upowszechniania rekomendacji**

Właściwe zidentyfikowanie grupy odbiorców ułatwia dostosowanie przekazu oraz umożliwia osiągnięcie lepszego efektu w upowszechnianiu wyników badań ewaluacyjnych. Pamiętać należy, że rzadko można dotrzeć do wszystkich potencjalnych grup odbiorców badania, a zatem za każdym razem trzeba skupiać się na tych najważniejszych i wybierać właściwe kanały komu-

nikacji. Tabela 1 pokazuje, w jaki sposób z tym zadaniem poradziło sobie Ministerstwo Spraw Zagranicznych w Danii – DANIDA.

Tabela 1. DANIDA – Podsumowanie strategii przekazywania wyników ewaluacji konkretnym odbiorcom

Podstawowy cel	Rodzaje strategii		
	Rozliczeniowa	Uczenia się	Partnerzy
	Rodzaje odbiorców		
	Parlament, liderzy polityczni, opinia publiczna	Akademicy, studenci, badacze, konsultanci, NGO	Agencje w krajach rozwijających się, inne agencje
Pełny raport ewaluacyjny		X	X
4-stronicowe podsumowanie	X	X	X
25-stronicowy raport – ekstrakt	X		
Publikacja prasowa	X		
Film	X		X
Roczny Raport dla zarządu	X		
Roczny Raport DANIDA	X		
DANIDA Newsletter	X		
Spotkania dla szerokiej publiczności i profesjonalistów		X	
Wykłady uniwersyteckie		X	
Informacja w Internecie	X	X	X
Seminaria, warsztaty			X
Udział własny w ewaluacji			X

Źródło: DANIDA, *External Feedback – Danida's Dissemination Vehicles*, Paper presented to DAC Tokyo Workshop on Evaluation Feedback for Effective Learning and Accountability, September 2000.

Powinno się zatem przetransponować wyniki ewaluacji w formie, która najskuteczniej umożliwi ich bezpośrednie wykorzystanie przez potencjalnych odbiorców. Wydawać się może, że lista możliwych sposobów komunikowania wyników ewaluacji jest długa, w praktyce jednak używa się zaledwie kilku sprawdzonych kanałów komunikacji zidentyfikowanych przez Dyрекcję Generalną EuropeAid<sup>21</sup>. Powszechnie praktykuje się publikowanie raportów w wersji elektronicznej w Internecie. Jednak, mimo że Internet jest niezwykle przydatnym instrumentem upowszechniania ewaluacji, nie umożliwia on dotarcia do ściśle określonej grupy odbiorców. Dlatego też należy uważać, co

<sup>21</sup> European Commission, *Guidelines...*, op.cit.

i w jaki sposób się publikuje, często uzupełniając raport właściwy raportem rozbieżności (o ile taki został sporządzony) pomiędzy wykonawcą a zamawiającym badania. Raporty w całości bądź krótkie artykuły, przedstawiające główne wnioski płynące z ewaluacji, umieszczane są również w intranecie organizacji, co ułatwia upowszechnianie wiedzy na temat rekomendacji wewnątrz instytucji. Niektóre organizacje zamieszczają informacje o prowadzonych ewaluacjach w różnego rodzaju *newsletters*. W przypadku DG EuropeAid jest to *Evaluation Feedback Newsletter*, a także, przygotowywany co sześć miesięcy przez zespół ewaluacyjny, specjalny *Feedback Newsletter*, w którym przedstawiane są, obok informacji na temat ewaluacji *on-going*, wyniki ukończonych ewaluacji. Nie mniej ważnym elementem wspomagającym upowszechnianie wyników ewaluacji jest organizowanie wkrótce po opublikowaniu raportu końcowego seminarium lub warsztatów tematycznych.

W tym kontekście niezmiennie pojawia się pytanie o to, w jaki sposób mówić o negatywnych wynikach ewaluacji. W teorii istnieje powszechna zgoda co do tego, że wszystkie wyniki ewaluacji powinny być upowszechniane, chociaż, jak przyznają przedstawiciele organizacji międzynarodowych, nie zawsze praktyka podąża za teorią<sup>22</sup>. Nie można się ludzić, iż ujawnienie nieprawidłowości albo braku postępu w realizacji konkretnych programów nie wywoła negatywnej reakcji wśród decydentów. Wydaje się jednak, że w dłuższej perspektywie czasowej taka szczerość opłaca się również samej organizacji, wymusza dokładniejszy *follow-up* za rekomendacjami oraz wyciągnięcie wniosków na przyszłość.

## **Przykłady rozwiązań instytucjonalnych stosowanych przez organizacje międzynarodowe**

Celem tej części artykułu jest prezentacja przykładów rozwiązań instytucjonalnych w zakresie ewaluacji stosowanych w różnych organizacjach międzynarodowych, wpływających na wzmocnienie roli rekomendacji, wspomaganie rozwoju organizacji i doskonalenia prowadzonych przez nią interwencji.

### **Komisja Europejska**

Praktyka ewaluacji w Komisji Europejskiej sięga lat osiemdziesiątych. Jej prawdziwy rozkwit nastąpił jednak wraz z wprowadzeniem zasady systematycznej ewaluacji programów finansowanych przez UE w latach dziewięćdziesiątych i został ugruntowany reformą Komisji Europejskiej z 2000 roku. W związku z tym, że zdecydowana większość Dyrekcji Generalnych w Komisji Europejskiej zajmuje się realizacją bądź koordynacją projektów

---

<sup>22</sup> OECD, „Evaluation feedback...”, op.cit., s. 26.

finansowanych ze środków wspólnotowych, duże znaczenie przywiązuje się do jakości systematycznie prowadzonych ocen oraz wykorzystywania płynących z nich wniosków. Nie powinno zatem zaskakiwać, że procedury określające sposób prowadzenia badań, raportowania wyników oraz, przede wszystkim, ich upowszechniania, podlegają jasnej definicji na poziomie całej instytucji. Podkreśla się jednocześnie, iż aby raport trafił do właściwych adresatów konieczne jest opracowanie odrębnych strategii, które zidentyfikują i scharakteryzują potencjalnych odbiorców ewaluacji oraz dostosują, w przypadku poszczególnych ewaluacji, kanały komunikacji.

W dokumentach Komisji podkreśla się obowiązek komunikowania wyników ewaluacji w taki sposób, aby zapewnić ich maksymalne wykorzystanie<sup>23</sup>, opisując standardy mechanizmu informacji zwrotnej, obejmującego raportowanie, upowszechnianie rezultatów oraz wdrażanie rekomendacji w procesie decyzyjnym. Wszystkie jednostki organizacyjne są zobowiązane do systematycznego monitorowania wykorzystania rekomendacji proponowanych przez ewaluatorów. Zakłada się również, że wszystkie raporty przekazuje się do Dyrekcji Generalnej ds. Budżetu, która stanowi centralny punkt w systemie ewaluacji Komisji Europejskiej. Proces ten przewiduje zarówno informowanie osób odpowiedzialnych za podejmowanie decyzji w poszczególnych jednostkach organizacyjnych o wynikach ewaluacji, jak i upowszechnianie tych wyników wśród obywateli. System informacji zwrotnej ma umożliwić wykorzystanie wyników ewaluacji do opracowywania nowych rozwiązań oraz przeglądu prowadzonych programów europejskich.

Niektóre z serwisów Komisji Europejskiej stosują specyficzne metody umożliwiające kontrolę nad wdrażaniem rekomendacji z ewaluacji. Na przykład w 2003 roku Dyrekcja Generalna EuropeAid, która odpowiada za ewaluację projektów pomocowych realizowanych w krajach trzecich, opublikowała wytyczne dotyczące upowszechniania wyników ewaluacji oraz przekazywania informacji zwrotnych na temat zawartych w nich rekomendacji<sup>24</sup>. Oprócz tego, że ewaluacje pełnią zasadniczą funkcję rozliczeniową (*accountability*), dostarczają także wielu rekomendacji oraz wiadomości o sukcesach i porażkach projektów, które wykorzystuje się przy projektowaniu kolejnych interwencji. W przypadku wszystkich tych programów upowszechnianie wyników ewaluacji i przekazywanie informacji zwrotnej stanowią integralną część procesu ewaluacji i są wykorzystywane w praktyce przez osoby przygotowujące kolejne programy. EuropeAid stosuje tzw. *fiche contradictoire*, za pomocą którego rekomendacje przedstawione przez ewaluatorów przekazuje się konkretnym jednostkom; one zaś akceptują je bądź, w uzasadnionych przypadkach, odrzucają<sup>25</sup>. Odmienne podejście stosuje DG Agri, prowadzące obowiązkowy

<sup>23</sup> European Commission, *Responding to Strategic Needs...*, op.cit.

<sup>24</sup> European Commission, *Guidelines...*, op.cit.

<sup>25</sup> Ibidem.

*follow-up* stopnia wdrożenia rekomendacji przez właściwe jednostki w pół roku po przekazaniu wyników ewaluacji<sup>26</sup>.

### **UNDP (Program Narodów Zjednoczonych ds. Rozwoju)<sup>27</sup>**

Zastosowanie metod wykorzystujących wiedzę płynącą z systemu monitoringu i ewaluacji stanowi powszechną praktykę w UNDP. Proces uczenia się w oparciu o doświadczenia z realizacji poprzednich programów jest w tym przypadku szczególnie istotny, jako że większość projektów realizowanych przez UNDP ma charakter innowacyjny i polega na testowaniu nowych rozwiązań, które same w sobie zakładają duży poziom niepewności. UNDP, prowadząc ewaluacje, współpracuje blisko z partnerami realizowanych przez siebie programów, starając się poprzez jasne określenie potrzeb informacyjnych ułatwiać im proces zbierania danych. W procesie ewaluacji UNDP skupia się przede wszystkim na poszukiwaniu tych informacji, które umożliwiłyby wprowadzenie istotnych zmian w kolejnych generacjach programów i projektów.

UNDP zakłada, iż rekomendacje płynące z monitoringu i ewaluacji powinny być analizowane oddzielnie. Zgodnie z wytycznymi UNDP pierwszym krokiem w analizie jest stworzenie listy pytań dotyczących rodzaju oraz konsekwencji wprowadzenia tych rekomendacji w krótko-, średnio- i długookresowej perspektywie. Analiza tych pytań pozwala zidentyfikować konkretne działania, które należy podjąć w zależności od perspektywy czasowej. Kolejnym krokiem jest organizacja warsztatów dla pracowników organizacji, w znacznym stopniu przyczyniających się do upowszechnienia wyników zakończonych ewaluacji. Szkolenia te skupiają się na tym, jak w efektywny sposób ulepszać programy realizowane przez UNDP, korzystając w tym zakresie z przedstawionych przez ewaluatorów rekomendacji.

### **DANIDA (Royal Danish Ministry of Foreign Affairs)<sup>28</sup>**

Raporty ewaluacyjne przygotowywane dla Duńskiego Ministerstwa Spraw Zagranicznych znajdują duże grono odbiorców wśród polityków, w mediach i zewnętrznych instytucjach badawczych. Ze względu na rodzaj programów prowadzonych przez DANIDA ewaluacja odgrywa bardzo ważną rolę jako źródło informacji o mocnych i słabych stronach dotychczasowego systemu i potencjalnych rozwiązań. Po otrzymaniu raportu z ewaluacji zespół ds. ewaluacji przygotowuje krótką notatkę skierowaną do wyższych urzędników, zawierającą główne wnioski i rekomendacje, a także identyfikuje konkretne instytucje, które są odpowiedzialne za ich wdrożenie. Streszczenie raportów

<sup>26</sup> European Commission, *Responding to Strategic Needs...*, op.cit.

<sup>27</sup> Opracowane na podstawie: UNDP, „Knowledge and learning: use of evaluative evidence”, w: *Handbook on Monitoring and Evaluation for Results*, Evaluation Office UNDP 2002.

<sup>28</sup> Opracowane na podstawie: DANIDA, *Danida's Evaluation Policy 2006*, <http://www.netpublikationer.dk/um/7481/index.htm>.

ewaluacyjnych przekazuje się również pracownikom DANIDA, co umożliwia im włączenie tego doświadczenia w zakres przyszłej działalności. Co roku raporty ewaluacyjne przedstawiane są także Dyrekcji Generalnej DANIDA i Parlamentarnemu Komitetowi ds. Zagranicznych. Tak zwane *follow-up memo* dyskutowane jest na wysokim szczeblu sekretarza stanu. Wydział zajmujący się ewaluacją prowadzi aktywną politykę upowszechniania wyników ewaluacji. Obok publikacji raportów ewaluacyjnych Wydział upowszechnia wyniki badań poprzez opracowywanie specjalnych podsumowań ewaluacji, raportów rocznych, raportów tematycznych, seminariów i warsztatów, informacji prasowych i debat publicznych.

### **UNHCR (Wysoki Komisarz Organizacji Narodów Zjednoczonych ds. Uchodźców)<sup>29</sup>**

UNHCR, podobnie jak wiele innych organizacji humanitarnych, przez dłuże lata skupiał się wyłącznie na przygotowywaniu ewaluacji podsumowujących po zakończeniu realizacji programów i projektów. Pomimo iż takie ewaluacje w większości przypadków przynosiły oczekiwane rezultaty w postaci zmian we wdrażaniu kolejnych generacji programów, zauważono, że często rekomendacje płynące z ewaluacji w ogóle nie były wdrażane. W odpowiedzi na te niedoskonałości systemu UNHCR wprowadził nową politykę, która miała na celu wzmocnienie roli ewaluacji. Odwołuje się ona do wspomnianych czterech filarów zaproponowanych przez Pattona<sup>30</sup>. Ponadto powołano Komitet Sterujący, kontrolujący poprzez zaangażowanie w cały proces ewaluacji, na ile projekt i samo badanie spełniają warunki umożliwiające późniejsze wdrażanie rekomendacji.

UNHCR kładzie szczególny nacisk na proces konsultacji przed zakończeniem ewaluacji i sformułowaniem końcowych wniosków i rekomendacji. Raporty zarówno publikuje się w wersji papierowej, jak i umieszcza na stronach UNHCR, natomiast rekomendacje z ewaluacji prezentowane są na specjalnym komitecie *Senior Steering Committee*. Managerowie z krajowych organizacji objętych badaniem są formalnie zobowiązani do wdrażania rekomendacji płynących z ewaluacji. W ciągu trzech miesięcy od otrzymania raportu ewaluacyjnego mają oni obowiązek przesłania do biura Komisarza (*High Commissioner*) sprawozdania z tego, jak poszczególne rekomendacje zostały wprowadzone w życie oraz, w przypadku odrzucenia którejś z nich, wyjaśnienia, dlaczego tak się stało. Na tej podstawie Zarząd UNHCR (*Executive Office*) ma prawo objąć każdego z managerów szczegółową kontrolą wdrażania tych rekomendacji.

<sup>29</sup> Opracowane na podstawie: UNHCR, *UNHCR's evaluation policy*, Evaluation and Policy Analysis Unit, UNHCR 2002, <http://www.unhcr.org/research/RESEARCH/3d99a0f74.pdf>.

<sup>30</sup> Patton M.Q., *Utilisation-Focused Evaluation...*, op.cit.



### **United Nations (Organizacja Narodów Zjednoczonych)<sup>31</sup>**

Organizacja Narodów Zjednoczonych stworzyła odrębny kodeks dotyczący prowadzenia ewaluacji. Spośród wielu punktów w nim zawartych, kilka odnosi się bezpośrednio do kwestii związanych z upowszechnianiem wyników i wdrażaniem rekomendacji płynących z ewaluacji. Przykładowo jedna z reguł (Standard 1.4) głosi, iż wszystkie organizacje Narodów Zjednoczonych powinny wprowadzić standardy, które umożliwiłyby prowadzenie *follow-up* raportów ewaluacyjnych. Mechanizmy te muszą w szczególności ułatwiać wdrażanie rekomendacji formułowanych w raportach ewaluacyjnych do przygotowywanych nowych programów i projektów. Ponadto wszystkie organizacje powinny posiadać przejrzyste procedury prowadzące do upowszechnienia raportów ewaluacyjnych tak, aby stały się one dostępne zarówno dla managerów organizacji, jak i dla szeroko rozumianej opinii publicznej. Przewiduje się jednakże możliwość zastrzeżenia treści raportów, jeśli wymaga tego poufność zawartych w raporcie informacji na temat któregośkolwiek z badanych podmiotów<sup>32</sup>.

Inna reguła (Standard 3.17) zakłada, że każdy raport ewaluacyjny wymaga formalnej odpowiedzi ze strony instytucji zajmującej się analizowanym programem, wobec której ewaluator sformułował konkretne rekomendacje. W tym przypadku wytyczne Narodów Zjednoczonych przewidują możliwość sporządzenia Planu Działania lub innego dokumentu, określającego w sposób nie pozostawiający żadnych wątpliwości obowiązki związane z wdrażaniem owych rekomendacji. Spełnienie powyższych zobowiązań monitorowane jest cyklicznie przez Dyrektora organizacji. Kolejna reguła (Standard 4.16) odnosi się bezpośrednio do ewaluatorów, podkreślając, że rekomendacje zawarte w raportach powinny, w miarę możliwości, definiować, kto i kiedy powinien je wdrożyć.

### **USAID (United States Agency for International Development)<sup>33</sup>**

USAID jako jedna z największych agencji, udzielających finansowego wsparcia w krajach rozwijających się, prowadzi systematyczną analizę prowadzonych przez siebie programów i projektów. Przygotowane przez USAID wytyczne w tym zakresie dokładnie określają, kiedy przeprowadza się ewaluację, jaki jest jej zakres, metodologia, sposób dokumentacji rezultatów oraz proces wdrażania płynących z niej rekomendacji. Dokument zakłada przynajmniej jedną ewaluację każdego z finansowanych przez organizację programów. Ewaluacja ta obok standardowych pytań ewaluacyjnych, dotyczących

<sup>31</sup> Opracowane na podstawie: UN, *Standards for Evaluation in the UN System*, United Nations Evaluation Group (UNEG) 2005.

<sup>32</sup> UN, *Standards for Evaluation...*, op.cit., s. 4.

<sup>33</sup> Opracowane na podstawie: USAID, *Assessing and Learning. Functional Series 200 – Programming ADS 203*, USAID 2004.

efektów prowadzonych działań i ich trwałości, jakości wdrażania oraz analizy kontekstu krajowego, dostarczałyby konkretnych przykładów rozwiązań, które można zastosować w kolejnych programach przygotowywanych przez USAID. Charakterystyczne jest zachęcanie do angażowania w proces ewaluacji organizacji z krajów, gdzie prowadzone są projekty USAID, włączając w to administrację rządową. Z tej perspektywy USAID wyróżnia autoewaluacje, badania prowadzone przez zewnętrznych ekspertów, ewaluacje opierające się na współpracy z jedną lub kilkoma agencjami oraz ewaluacje uczestniczące, angażujące wiele podmiotów zajmujących się wdrażaniem konkretnego programu. Każdy raport obok wniosków opartych na analizie oraz konkretnych sugestii zmian w postaci rekomendacji musi zawierać identyfikację dobrych praktyk, które są następnie wykorzystywane przy projektowaniu programów o podobnym charakterze lub wdrażanych w podobnym środowisku instytucjonalnym. Komórki USAID mają obowiązek podjęcia konkretnych kroków w celu wdrożenia rekomendacji; należą do nich m.in. spotkanie się z ewaluatorami i przedyskutowanie wniosków, systematyczna analiza rezultatów badania i przedstawionych rekomendacji oraz ich akceptacja bądź odrzucenie. Na podstawie tej analizy określa się dokładnie, jakie działania powinny być wdrożone, kto ma za nie odpowiadać i kiedy mają być realizowane. Wytyczne zakładają również przeprowadzenie otwartej dyskusji na temat wyników ewaluacji i płynących z niej rekomendacji z instytucjami korzystającymi z pomocy. Obowiązkowe jest również przekazanie raportów ewaluacyjnych do Development Experience Clearinghouse (DEC), swoistej bazy danych, w której archiwizowane są programy USAID, a także dotyczące ich analizy, dzięki czemu raporty pozostają dostępne dla całej organizacji i mogą być wykorzystywane przy planowaniu innych projektów w przyszłości<sup>34</sup>.

## Zakończenie

Wykorzystanie wyników ewaluacji, a w szczególności zarządzanie rekomendacjami płynącymi z badań to jeden z najczęściej poruszanych tematów w publikacjach dotyczących ewaluacji. Ewaluacja nie tylko legitymizuje tworzenie i realizację interwencji publicznych, lecz także odgrywa zasadniczą rolę, dostarczając racjonalnych i systematycznych informacji, potrzebnych w procesie decyzyjnym. Tym samym ewaluacja wspomaga opracowywanie nowych programów poprzez analizowanie specyficznych pytań, na które odpowiedzi poszukują ich projektodawcy.

Należy podkreślić, że ewaluacja jest użyteczna tylko wtedy, kiedy jej wyniki wykorzystuje się do ulepszania i rozwoju programów. Żeby zwiększyć efektywność zastosowania i wykorzystywania rekomendacji płynących

<sup>34</sup> Wszystkie dokumenty dostępne są na stronie Development Experience Clearinghouse: [www.dec.org](http://www.dec.org).

z ewaluacji oraz, przede wszystkim, tworzyć lepsze programy, trzeba spełnić kilka warunków. Po pierwsze, zadania, które stawia się przed ewaluacją, powinny zostać dobrze określone, tak aby stało się oczywiste, że ewaluacja jest instrumentem zorientowanym na rezultaty programu i system wdrażania. Jednocześnie ewaluację należy wyraźnie oddzielić od innych narzędzi dostarczających informacji managerom programów, takich jak monitoring albo kontrola finansowa. Dużo zależy również od tego, w jaki sposób ewaluacja była zaplanowana i przeprowadzona. Ważnym czynnikiem sukcesu jest ściśle współdziałanie wykonawców ze wszystkimi podmiotami zaangażowanymi w realizację programu oraz aktorami potencjalnie zainteresowanymi wynikami badania. W celu wzmocnienia znaczenia i zwiększenia wykorzystania ewaluacji można również wprowadzić przepisy prawne, nakazujące stosowanie ewaluacji oraz wykorzystywanie jej wyników, tak jak przewiduje to system Funduszy Strukturalnych. Regulacje Unijne zakładają obowiązek prowadzenia ewaluacji w przypadku konieczności przesunięcia środków pomiędzy poszczególnymi priorytetami programu (EC Regulation 1083/2006). Inne programy wiążą ewaluację z alokacją budżetową, czyli np. z przekazywaniem dodatkowych funduszy dla programów odnoszących sukcesy.

Ewaluacja wymaga wspólnych definicji, a zwłaszcza jasnego zdefiniowania pojęć sukcesu i porażki w odniesieniu do programów, tak aby różni uczestnicy procesu ewaluacji podobnie te pojęcia rozumieli. Ewaluacja i płynące z niej rekomendacje powinny odpowiadać w bezpośredni sposób na potrzeby uczestników tego procesu. Wyniki badań ewaluacyjnych mają docierać do właściwych odbiorców w odpowiednim czasie, tak by można było oczekiwać, że zostaną one wzięte pod uwagę. Praktycznie wszystkie instytucje, które wdrażają programy publiczne, a następnie zlecają ich ewaluację, muszą także podjąć decyzję o tym, jakie są najskuteczniejsze metody upowszechniania wniosków i rekomendacji z ewaluacji.

---

**Agnieszka Sternik** – od 2004 roku zajmuje się ewaluacją funduszy europejskich, w szczególności Europejskiego Funduszu Społecznego. Absolwentka Wydziału Socjologii na Uniwersytecie Warszawskim (2001) oraz Polsko-Francuskich Studiów Europejskich w Szkole Głównej Handlowej (2002). Laureatka Chevening Scholarship przyznawanego przez rząd brytyjski, dzięki któremu ukończyła roczne studia z zakresu badań społecznych i ewaluacji na University of Bath. W 2004 roku jako pracownik Departamentu Zarządzania EFS stworzyła jednostkę ds. ewaluacji EFS oraz, współpracując z Krajową Jednostką Ewaluacyjną, wspierała budowę potencjału ewaluacyjnego w Polsce. Od 2006 roku pracuje w Komisji Europejskiej, w Dyrekcji Generalnej ds. zatrudnienia, polityki społecznej i równych szans. Zajmuje się ewaluacją inicjatyw KE w zakresie polityki zatrudnienia i integracji społecznej, a także współpracuje z krajami członkowskimi przy ewaluacji EFS.

## Bibliografia

- Balthasar A., „The effects of institutional design on the utilization of evaluation, evidence using qualitative comparative analysis (CQA)”, *Evaluation* 2006, nr 12(3), s. 353–371.
- Balthasar A., Rieder S., „Learning from evaluation. Effects of the evaluation of the Swiss Energy 2000 Programme”, *Evaluation* 2000, nr 6(3), s. 245–260.
- Cracknell B., „The role of aid-evaluation feedback as an input into the learning organization”, *Evaluation* 2001, nr 7(1), s. 132–145.
- DANIDA, *Danida's Evaluation Policy 2006*, <http://www.netpublikationer.dk/um/7481/index.htm>.
- DANIDA, *External Feedback – Danida's Dissemination Vehicles*, Paper presented to DAC Tokyo Workshop on Evaluation Feedback for Effective Learning and Accountability, September 2000.
- Davies H., McDermott R., *How Knowledge Management Affects the Bank's. Change Management Programme*, ICD 1996.
- Eason C., Denman S., *Lessons for Objective 1 from the Previous Generation of Programmes*, WEFO 2002.
- European Commission, *Guidelines for Dissemination and Feedback of Evaluations*, DG AIDCO, Brussels 2003.
- European Commission, „Final reporting, dissemination and use of evaluation results”, w: *Evaluating EU Activities*, DG Budget 2004, [http://ec.europa.eu/budget/evaluation/pdf/pub\\_eval\\_activities\\_full\\_en.PDF](http://ec.europa.eu/budget/evaluation/pdf/pub_eval_activities_full_en.PDF).
- European Commission, „Introduction”, w: *Evaluating EU Activities*, DG Budget 2005.
- European Commission, *Responding to Strategic Needs: Reinforcing the Use of Evaluation*, Communication to the Commission from Ms Grybauskaitė in agreement with the President, Brussels 2006.
- Gray D., *Doing Research in the Real World*, Sage Publications, London 2005.
- Lang J., „Improving structural policy under conditions of hybrid governance: multi-actor constellations, integrated feedback instruments and the use of evaluation results”, *Evaluation* 2001, nr 7(1), s. 7–23.
- OECD, „Evaluation feedback for effective learning and accountability”, *Evaluation and Aid Effectiveness 5*, OECD 2001.
- Patton M.Q., *Utilisation-Focused Evaluation: The New Century Text*, Sage Publications, Beverly Hills 1997.
- Patton M.Q., „Discovering process use”, *Evaluation* 1998, nr 4(2), s. 225–233.
- Polskie Towarzystwo Ewaluacyjne, *Ewaluacja Funduszy Strukturalnych*, Opracowanie na zlecenie Wydziału Monitoringu i Raportowania ZPORR, Departament Wdrażania Programów Rozwoju Regionalnego, Ministerstwo Gospodarki, Pracy i Polityki Społecznej 2003.
- Schaumburg-Müller H., „Use of aid evaluation from an organizational perspective”, *Evaluation* 2005, nr 11(2), s. 207–222.
- Taut S., Brauns D., „Resistance to evaluation. A psychological perspective”, *Evaluation* 2003, nr 9(3), s. 247–264.
- UKIE, *Podręcznik ewaluacji okresowej Programu PHARE i Transition Facility*, UKIE 2006.
- UN, *Standards for Evaluation in the UN System*, United Nations Evaluation Group (UNEG) 2005.
- UNDP, „Knowledge and learning: use of evaluative evidence”, w: *Handbook on Monitoring and Evaluation for Results*, Evaluation Office UNDP 2002.
- UNFPA, „Evaluation standards”, w: *Programme Manager's Planning, Monitoring and Evaluation Toolkit*, UNFPA 2004.
- UNHCR, *UNHCR's evaluation policy*, Evaluation and Policy Analysis Unit, UNHCR 2002, <http://www.unhcr.org/research/RESEARCH/3d99a0f74.pdf>.

USAID, *Assessing and Learning. Functional Series 200 – Programming ADS 203*, USAID 2004.

WEFO, *European Structural Funds in Wales – Research Monitoring and Evaluation Strategy*, Welsh Assembly Government 2005.

Weiss C., „The interface between evaluation and public policy”, *Evaluation* 1999, nr 5(4), s. 468–486.



Część II

**Wybrane zagadnienia metodologiczne**





*Mikołaj Jasiński, Michał Kowalski*

## **Falszywa sprzeczność: metodologia jakościowa czy ilościowa?**

Umiejętność dokonania właściwych wyborów metodologicznych jest kwestią kluczową dla pomyślnego zrealizowania każdego przedsięwzięcia badawczego. Wybór właściwej metodologii warunkuje natura stojącego przed nami problemu badawczego. Trudno więc przedstawić w tym miejscu ogólne wskazówki dotyczące projektowania metodologii badań. Zawsze będzie ona zależać od charakteru problemu, z którym przyjdzie nam się zmierzyć. Jak zauważa Silverman<sup>1</sup>, wybór metodologii powinien odzwierciedlać całościową strategię badawczą, obrana metodologia zaś decydować będzie o zastosowaniu konkretnych technik gromadzenia danych, sposobów ich analizy i interpretacji. Wybór metodologii jest zawsze silnie związany z założeniami teoretycznymi. Musi on odzwierciedlać poczynione przez badacza założenia ontologiczne i epistemologiczne. W przypadku ewaluacji *ex-post* charakter badań, założenia teoretyczne i dostępna metodologia wraz z założeniami określone są przez czasowe usytuowanie badania w procesie realizacji ewaluowanego projektu, a także przez oczekiwania stawiane ewaluatorom, czyli dostarczanie informacji o efektach interwencji.

Ewaluacja *ex-post* podsumowuje oraz ocenia program jako całość wtedy, gdy wyniki programu są już skonkretyzowane<sup>2</sup>; zwraca zatem uwagę na dwa główne elementy – moment realizacji badań i efekty programu. Oba te czynniki, niejako definicyjne dla ewaluacji *ex-post*, nakładają znaczące ograniczenia na metodologię badań prowadzonych w ramach tego typu ewaluacji.

Po pierwsze, w takich badaniach mamy do czynienia przede wszystkim z danymi zastanymi, zazwyczaj wytworzonymi intencjonalnie podczas działań związanych z realizacją określonego programu interwencji społecznej. Tego typu dane dokumentują realizację działań podejmowanych przez osoby odpowiedzialne za wdrożenie innowacji (zmiany) społecznej. Toteż sposób

---

<sup>1</sup> Silverman D., *Doing Qualitative Research. A Practical Handbook*, Sage Publications, London 2000.

<sup>2</sup> Zob.: Ekiert K., *Ewaluacja w administracji publicznej, funkcje, standardy i warunki stosowania*, Departament Spraw Społecznych i Ustrojowych 2006; Brouwer A., *Wdrażanie Europejskiego Funduszu Rozwoju Regionalnego*, Projekt Phare PI/IB/2002/OT/03C 2002.

tworzenia takich dokumentów można określić jako intencjonalny. Stąd też wynikają znaczące ograniczenia dotyczące jakości i wiarygodności danych. Dokumentacja tworzona intencjonalnie przez organizatorów programu interwencji społecznej jest najmniej wiarygodna. Jak zauważa metodolog A. Sułek, na dane urzędowe wpływają intencje osób je tworzących i pozostających w rozmaitych relacjach strukturalnych zależności: „Podwładni mogą dążyć do uzyskania od przełożonych deficytowych dóbr, a wówczas w swych sprawozdaniach będą wyolbrzymiać własne obowiązki lub ukrywać zasoby. Jeśli dążą do uzyskania pozytywnych ocen swej działalności, to mogą zawyżać wskaźnik efektywności”<sup>3</sup>. Przedstawiciele szczebli nadrzędnych również biorą udział w „poprawianiu” danych urzędowych – „uczają” niższe szczeble dostarczania tylko tych danych, których oczekują. Dlatego w badaniach społecznych rzadziej korzysta się z danych urzędowych, niż mogłyby tego oczekiwać osoby niezwiązane z problematyką badań. Dane te bowiem w niewielkim stopniu mówią o rzeczywistości; pokazują raczej intencje osób je tworzących i okoliczności, w jakich powstają. Tak więc ostateczne wnioskowanie o efektach danego programu na podstawie dokumentów wytworzonych przez jego realizatorów obniża rzetelność analiz.

Co więcej, ten typ ewaluacji można określić jako ewaluację konkluzywną: przeprowadzaną, gdy realizacja programu jest już zamknięta i możemy obserwować jedynie efekty działań podejmowanych w ramach programu; nie mamy zaś w ogóle kontaktu z programem jako takim. Z tego właśnie powodu dokumenty wytworzone w trakcie realizowanych działań są głównym źródłem informacji wykorzystywanych przez badacza w ramach ewaluacji *ex-post*.

Drugim znaczącym ograniczeniem dla realizacji badań ewaluacyjnych typu *ex-post* jest skoncentrowanie uwagi badaczy na mierzalnych efektach programu. Takie analizy we wszelkich badaniach społecznych należą do najtrudniejszych i najbardziej kosztownych. Zasadnicze trudności sprawiają:

- rzetelne opisanie stanu początkowego, poprzedzającego planowaną interwencję,
- objęcie pełnej kontroli nad warunkami realizacji danej interwencji społecznej,
- konieczność wyodrębnienia wszystkich (pozytywnych i negatywnych) czynników związanych bezpośrednio z realizacją danego programu oraz wszystkich czynników (pozytywnych i negatywnych), które, choć niezwiązane bezpośrednio z danym programem (z daną interwencją), mogły mieć wpływ (osłabiający bądź wzmacniający) na jej efekty<sup>4</sup>.

<sup>3</sup> Sułek A., *Ogród metodologii socjologicznej*, Wydawnictwo Naukowe Scholar, Warszawa 2002, s. 125.

<sup>4</sup> Najpowszechniejszym typem czynników zewnętrznych dla realizowanego programu mogą być np. zmiany regulacji prawnych lub też zmiany kursu walut, zachodzące w tym samym czasie, w którym realizowany jest program interwencji. Chociaż takie zmiany zewnętrznego kontekstu realizacji programu dają się najłatwiej kontrolować, jednoznaczne oszacowanie efektów wpływu

Jak stwierdził E. Stern na I Konferencji Ewaluacyjnej<sup>5</sup>, w praktyce badania tego typu są nierealizowalne, badacz zaś, któremu udałoby się je zrealizować, „zasługiwałby na nagrodę Nobla”.

Taka opinia wiąże się z koniecznością odejścia od technokratyczno-urzędowego modelu ewaluacji. We współczesnej praktyce ewaluacyjnej coraz częściej zaznacza się potrzeba rozszerzenia zestawu technik, metod i podejść badawczych.

Z naszych doświadczeń wynika, że wielu badaczy nadal patrzy na problem ewaluacji przez pryzmat preferowanej metody, nie starając się dostosować metodologii do charakteru problemu badawczego oraz specyfiki badanej społeczności. Został wypracowany pewien schemat prowadzenia badań ewaluacyjnych, który, niezależnie od podejmowanej problematyki badań, realizowany jest przy wykorzystaniu wąskiego wachlarza technik badawczych i analitycznych. Schemat ten w przypadku ewaluacji *ex-post* opiera się na analizie zgromadzonych dokumentów, realizacji badań ilościowych wśród beneficjentów programu (techniką wywiadu CATI i/lub ankiety pocztowej), wywiadów – z reguły trzech lub czterech – z kluczowymi informatorami (są to przede wszystkim osoby odpowiedzialne za realizację ewaluowanego programu). Naszym zdaniem taka praktyka badawcza w zasadniczy sposób redukuje jakość wniosków i rekomendacji dających się sformułować po przeprowadzonych badaniach, a jedynym uzasadnieniem jej stosowania są stosunkowo niskie koszty takich badań oraz łatwość ich realizacji. Pewien niepokój budzą także sformułowania zawarte w wielu raportach z przeprowadzonych ewaluacji *ex-post*, w których autorzy – naszym zdaniem w sposób nieuprawniony – podkreślają fakt wykorzystania metod jakościowych i ilościowych jako swoistej „triangulacji”<sup>6</sup>, służącej „weryfikacji wniosków”. Przykład takiego podejścia pokazuje następujący cytat z przeprowadzonych badań: „Jednowymiarowy obraz rzeczywistości prezentowany w dokumentach został zweryfikowany metodą pogłębionych wywiadów indywidualnych (IDI) i badań kwestionariuszowych. Dane uzyskane z wywiadów CATI miały charakter zarówno ilościowy, jak i jakościowy, przy czym przeważały dane ilościowe”. Pozornie wszystko jest w porządku. Wydawać się może, iż w tym przypadku mamy do czynienia z integracją ilościowych i jakościowych

---

tych czynników na działania podejmowane w ramach programu może być trudne. Nieraz jeszcze większy problem stanowi oszacowanie wpływu trudno mierzalnych czynników zewnętrznych. Na przykład efekty zmian regulacji prawnych są trudne do oszacowania w chwili ich wprowadzenia. Efekty podjętych decyzji mogą być widoczne w dłuższym okresie czasu; np. o efektach zmiany regulacji, dotyczących rejestracji używanych samochodów sprowadzanych z zagranicy, dla ochrony środowiska naturalnego można wnioskować parę lat po wprowadzeniu w życie tej decyzji, chociaż prawdopodobnie od razu po wprowadzeniu danej regulacji zacznie ona bezpośrednio wpływać na indywidualne decyzje osób sprowadzających samochody.

<sup>5</sup> <http://www.konferencja-ewaluacja.pl/2005/materialy.php>.

<sup>6</sup> Zagadnieniem „triangulacji”, kluczowym dla integracji metod, zajmujemy się w dalszej części artykułu.

metod badawczych (triangulacją), służącą weryfikacji informacji pochodzących z różnych źródeł. Jednakże, jak można zorientować się z lektury raportu, wywiady indywidualne zostały przeprowadzone z osobami tworzącymi analizowane dokumenty. Trudno w takim razie mówić o triangulacji i weryfikacji źródeł informacji. Trudno także stwierdzić, w jakim stopniu przeprowadzone wywiady służyły poznaniu opinii badanych, w jakim zaś były sposobem gromadzenia informacji o faktach. Wątpliwości te związane są z klasycznym problemem metodologicznym dotyczącym celu wywiadu: czy dostarcza on wiedzy o faktach, czy też służy zdobyciu wiedzy o sposobie interpretacji faktów przez informatora? Pierwsze założenie charakterystyczne jest dla podejścia ilościowego, które, odwołując się do częstości wskazań wśród respondentów, stawia sobie za cel oszacowanie natężenia opisywanego zjawiska. Drugie natomiast, poprzez pogłębioną analizę wypowiedzi każdego z rozmówców, prowadzi do tworzenia typologii sposobów interpretacji zjawiska społecznego. Istotą rzeczy jest tu więc właściwe umiejscowienie każdej z metod. Niezachowanie tej reguły prowadzi do nadinterpretacji wyników badań, a nawet do całkowitego zafałszowania wniosków.

W tym kontekście często pojawiające się pytanie, czy należy wybrać jakościowe, czy ilościowe metody badawcze, jest problemem fikcyjnym, cały zaś konflikt między „ilościowcami” a „jakościowcami” – swoistym nieporozumieniem. O zastosowaniu konkretnych metod decyduje cel badań, problemy badawcze, charakterystyka analizowanych zjawisk społecznych, wreszcie, wielowymiarowość rzeczywistości społecznej, a nie preferencje poszczególnych badaczy, wynikające często z braku kompetencji metodologicznych w podejściu konkurencyjnym. Drugim znaczącym czynnikiem prowokującym błędy metodologiczne jest zbyt daleko posunięta ingerencja w planowanie metodologii na poziomie *Terms of Reference*. Uniemożliwia ona projektującym badanie ewaluatorom właściwe dopasowanie metodologii do celów badania oraz cech badanej interwencji społecznej.

Nie zamierzamy minimalizować różnic między metodami ilościowymi a jakościowymi. Wręcz przeciwnie – świadomi ograniczeń i odmienności każdego z tych typów, wskazujemy na możliwości, jakie stwarza umiejętne połączenie technik badawczych, związanych z poszczególnymi podejściami. Postulowana przez nas postawa badawcza nazywana jest w skrócie **triangulacją**<sup>7</sup>. Niejednokrotnie w raportach ewaluacji *ex-post* pojęcie to zawęża się do dywersyfikacji źródeł informacji albo też, co gorsza, sprowadza się triangulację do prób wyciągania wniosków jakościowych z badań o charakterze ilościowym bądź odwrotnie – wnioskowania ilościowego na podstawie badań jakościowych.

---

<sup>7</sup> Zob. m.in. Jick T.D., „Mixing qualitative and quantitative methods: triangulation in action”, *Administrative Science Quarterly* 1979, nr 24(4), s. 602–611.

**Triangulacja** może być rozumiana jako zwielokrotnienie:

- metod i technik badawczych,
- źródeł informacji,
- typów danych,
- technik analitycznych,
- teorii wyjaśniających,
- liczby osób realizujących badania.

Opierając się na typologii Michaela Q. Pattona<sup>8</sup> można wyróżnić cztery typy triangulacji służących sprawdzeniu i weryfikacji przeprowadzanych analiz jakościowych: 1) triangulację metodologiczną (*methods triangulation*), polegającą na kontrolowaniu spójności wniosków formułowanych przy wykorzystaniu różnych metod gromadzenia danych; 2) triangulację źródeł informacji (*triangulation of sources*), polegającą na zwielokrotnieniu źródeł informacji w ramach tej samej metody ich gromadzenia; 3) triangulację analityczną (*analyst triangulation*), w której zwiększa się liczbę osób analizujących wnioski z badań; 4) triangulację teoretyczną (*theory/perspective triangulation*), polegającą na zwielokrotnieniu teorii i/lub perspektyw badawczych służących interpretacji danych.

Powszechny brak współpracy między badaczami reprezentującymi jakościowe i ilościowe paradygmaty w metodologii badań społecznych znacznie zubaża przeprowadzane analizy. Sprawa nabiera dodatkowego znaczenia w przypadku badań ewaluacyjnych, w których niejednokrotnie z zasady wyklucza się jeden bądź drugi model metodologiczny. Przykładem takiego badania może być ewaluacja *ex-post*; często bowiem z założenia wyklucza się zastosowanie w niej jakościowych metod badawczych<sup>9</sup>. Sztuczność tego podejścia ujawnia się, na przykład, przy dokonywaniu oceny trafności interwencji po realizacji danego programu, na podstawie założeń sformułowanych przed jego rozpoczęciem. Brak jakościowej analizy procesów w trakcie realizacji programu oraz bezpośredniego kontaktu z beneficjentami po jego zakończeniu powoduje, że ocena trafności sprowadza się do formalnego potwierdzenia celów projektu określonych w jego założeniach. Nie ma natomiast możliwości odniesienia efektów programu do rzeczywistych potrzeb jego beneficjentów. Można, na przykład, założyć, iż zwiększenie środków finansowych przeznaczonych na budowę autostrady spowoduje podwyższenie atrakcyjności regionu dla potencjalnych inwestorów i pozwoli zrealizować założony cel programu – rozwój społeczno-ekonomiczny regionu. Łatwo przeprowadzić ewaluację *ex-post*, wykazującą trafność, efektywność i skuteczność interwencji. Z drugiej jednak strony mało kto da się przekonać, że rozwój społeczno-ekonomiczny jakiegokolwiek regionu zależy od tego, ile

<sup>8</sup> Patton M.Q., *Qualitative Evaluation and Research Method*, second edition, Sage Publications, London 1990, s. 464–465.

<sup>9</sup> Z założenia metody te opierają się na bezpośrednim kontakcie badacza z badaną rzeczywistością; w przypadku, gdy realizacja programu została zakończona, taki kontakt jest już niemożliwy.

zbuduje się tam nowych dróg. Takie wskaźniki stosuje się mimo to powszechnie i uważa za zobiektywizowane, świadczące o rozwoju. Są one zaś zbyt ubogie. Rozwój regionu jest zjawiskiem na tyle złożonym, że sprowadzenie go do uproszczonych wskaźników spowoduje niesprawiedliwość i błędność ocen. Brak uczestnictwa badacza w realizacji programu znacząco zmniejsza jakość możliwych do sformułowania wniosków. Brak kontaktów ewaluatora z beneficjentami po zakończeniu interwencji uniemożliwia jakościową refleksję na temat konsekwencji podjętych działań, pozwalającą na pogłębioną analizę korzyści i ewentualnych strat, które przyniósł zakończony już program. Program, realizujący te same cele społeczne, ale działający w odmienny sposób, zostanie oceniony jako nietrafny, gdyż nie da się wykazać, że był realizowany za pomocą wskaźników określonych w projekcie. Dotyczy to nie tylko ewaluacji *ex-post*. Taki problem będzie występować w każdym typie ewaluacji, w którym konstrukcja narzędzia oraz wybór wskaźników i metod nie zostaną poprzedzone głębszą refleksją na temat natury analizowanych zjawisk społecznych. W ewaluacji typu *ex-post* problem ten jednak występuje najwyraźniej ze względu na ograniczanie jej do analizy danych zastanych.

Zagadnieniem, w którym niezbędne okazują się metody ilościowe jest problem reprezentatywności. Jak generalizować lub porównywać wyniki badań jakościowych, zawsze kontekstualnych, zupełnie inaczej podchodzących do tej kwestii niż czynią to badania ilościowe? Tymi i podobnymi aspektami wyboru metody badawczej zajmiemy się poniżej.

## Odległe...

Nie opisujemy szczegółowo różnic dzielących oba podejścia. Różnice te, wszak fundamentalne, zostały bowiem szeroko omówione w podręcznikach metodologii badań społecznych. Wynikają one z odmiennych podstaw filozoficznych (ontologicznych i epistemologicznych)<sup>10</sup>, których konsekwencją są różne postawy badaczy, jak też odmienności w zbieraniu danych oraz sposobach ich analizy i interpretacji<sup>11</sup>. Nawet pojęcie reprezentatywności jest zasadniczo różnie rozumiane w obu podejściach. Można zasygnalizować tu za Brymanem<sup>12</sup> podstawowe różnice między ilościowymi a jakościowymi podejściami badawczymi.

---

<sup>10</sup> Sztompka P., „Teoria socjologiczna końca XX wieku. Wstęp do wydania polskiego”, w: J. Turner, *Struktura teorii socjologicznej*, przekł. J. Szmatka, Państwowe Wydawnictwo Naukowe, Warszawa 1985, s. 9–53.

<sup>11</sup> Zob. Gubrium J.F., Holstein J.A., „Qualitative methods”, w: E.F. Borgotta, M.L. Borgotta (red.), *Encyclopedia of Sociology*, Macmillan, New York 1992; Lissowski G., „Metody ilościowe”, w: *Encyklopedia socjologii*, Oficyna Naukowa, Warszawa 2002; Hammersley M., Atkinson P., *Metody badań terenowych*, przekł. S. Dymczyk, Zysk i S-ka, Poznań 2000.

<sup>12</sup> Bryman A., *Quantity and Quality in Social Research*, Routledge, London 1988, s. 93–126.

Tabela 1. Różnice między ilościowymi a jakościowymi podejściami badawczymi

	<b>Badania ilościowe</b>	<b>Badania jakościowe</b>
Koncepcja rzeczywistości społecznej	Stacyczna, zewnętrzna wobec wszystkich aktorów: – rzeczywistość istnieje w sposób obiektywny i sposób niezależny od wszystkich uczestników życia społecznego	Zmienna, procesualna, społecznie konstruowana przez aktorów: – rzeczywistość społeczną konstruują aktorzy życia społecznego – jest ona taka, jaką widzą (tworzą) ją poszczególni aktorzy
Zakres ustaleń badawczych	Nomotetyczny: – celem prowadzonych badań jest ustalenie uniwersalnych praw rządzących życiem społecznym – na podstawie zgromadzonych danych wnioskuje się o całych populacjach – typową formą badania jest sondaż	Idiograficzny: – celem prowadzonych badań jest opis dotyczący konkretnych osób, czasu i miejsca – typową formą badania jest „studium przypadku” ( <i>case study</i> )
Związek między teorią a badaniami	Konfirmacja/weryfikacja: – teoria społeczna określa, co i w jaki sposób będzie badane – teoria → badania empiryczne	Odkrywanie: – celem badań jest opis i ewentualne tworzenie teorii, a nie jej weryfikacja – teoria może dostarczyć jedynie wstępnej orientacji w badanej problematyce – badania empiryczne → teoria
Strategia badawcza	Ustrukturowana: – jednoznacznie określona procedura badawcza, której jasność i uzasadnienie poszczególnych etapów jest warunkiem poprawności metody	Nie ustrukturowana: – brak z góry określonego planu prowadzenia badań – możliwość dokonywania zmian i reformułowania problemów badawczych pod wpływem informacji uzyskiwanych w trakcie badań
Charakter danych	„twarde”, „solidne”, „porównywalne”	„miękkie”, „głębokie”, „wyjątkowe”
Pozycja badacza w stosunku do przedmiotu badania	<i>Outsider</i> : – obserwator zdystansowany, zewnętrzny i niezależny wobec badanego przedmiotu, tak jak w naukach przyrodniczych	<i>Insider</i> : – poznanie i opisanie konkretnej rzeczywistości społecznej z perspektywy jej uczestników wymaga od badacza bliskiego i bezpośredniego kontaktu z przedmiotem badania
Związek między badaczem a badanymi	Odległy: – brak kontaktu lub jedynie sporadyczny kontakt w czasie przeprowadzania sondażu lub eksperymentu	Bliski: – w przypadku realizacji obserwacji uczestniczącej długotrwały kontakt z badanymi – w przypadku realizacji wywiadów charakter kontaktu z badanymi zależy od typu wywiadu

Źródło: Opracowanie własne.

W badaniu ilościowym zasadniczą czynnością jest pomiar, polegający na przypisaniu obserwowanym obiektom oraz relacjom między nimi modelu

liczbowego zawierającego relacje między liczbami–miarami i reprezentującego relacje empiryczne między elementami populacji. O tym, jakiego typu wnioski będzie można wywieść na podstawie uzyskanych w badaniu danych, decyduje poziom pomiaru. Stopień jednoznaczności wartości zmiennych uzyskanych w procesie pomiaru oraz stopień arbitralności jednostek tych zmiennych decyduje z kolei o tym, jakie interpretacje badacza będą uprawnione na etapie analizy. Na przykład, badając osiągnięcia szkolne uczniów, różnym poziomom wiedzy ucznia przypisuje się stopnie (w skali od 1 do 6, czy też od 2 do 5). Stopnie te mają charakter wyłącznie porządkowy. Nikogo nie trzeba przekonywać, że „odległość” między „dwóją” a „tróją” jest bez wątpienia inna niż między „czwórką” a „piątką”. Arytmetyczna różnica jest jednak identyczna, a obliczanie średniej czy też współczynnika korelacji liniowej Pearsona zakłada istnienie ustalonej jednostki pomiaru (pomiar interwałowy). Mimo że nie ma takiej jednostki w przyjętej metodzie oceniania, parametry te powszechnie oblicza się i interpretuje (!) w naszym systemie edukacji. Inaczej będzie na przykład w przypadku dochodu. Tu możemy wyznaczać i porównywać znacznie więcej parametrów statystycznych (średnią, wariancję, modele regresyjne itp.), przy czym napotykamy na ograniczenia, wynikające choćby z różnic walutowych czy też struktury porównywanych zbiorowości. Zastosowanie prostych parametrów statystycznych może stać się w takiej sytuacji źródłem poważnych błędów.

W badaniu jakościowym pomiar w zasadzie nie występuje. Celem jest tu stworzenie typologii (czasem określanej jako zmienne albo kategorie analityczne) obserwowanych obiektów i opisanie relacji między nimi. Opis ten powinien odpowiadać sposobowi porządkowania świata przez badanych. Toteż można powiedzieć, że procedura badawcza ma na celu ujawnienie i opisanie punktów widzenia ludzi w całej ich różnorodności. Przykładowo, badacz może ujawnić, iż stosowany powszechnie w badaniach podział nauczycieli ze względu na nauczane przez nich przedmioty nie odpowiada typologii stosowanej przez uczniów, dla których ważniejszym kryterium podziału jest sprawiedliwość postępowania poszczególnych nauczycieli. Stworzona na tej podstawie typologia pozwoli trafniej opisać relacje między uczniami a nauczycielami. Stosowanie narzuconych i nieadekwatnych typologii można porównać do prób opisu kolorów za pomocą kategorii „słodki”, „gorzki” itd. Brak znajomości typologii stosowanej przez aktorów życia społecznego może prowadzić do wniosków nieadekwatnych i nieużytecznych, chociaż sprawiających wrażenie poprawnych formalnie i mających charakter jakościowy.

W badaniu ilościowym proces badawczy przebiega równocześnie z rezygnacją z danych indywidualnych na rzecz coraz bardziej syntetycznych wniosków, dotyczących przedmiotu badania, a więc populacji. W wyniku pomiaru badacz określa wartości zmiennych. Na podstawie macierzy danych indywidualnych, które są przypisaniami wartości zmiennych poszczególnym



jednostkom obserwacji, badacz wyznacza rozkłady zmiennych, a następnie wyznacza miary–parametry, charakteryzujące strukturę zbiorowości ze względu na te zmienne. Traci więc informację o poszczególnych jednostkach, a nawet o częstościach reprezentowania poszczególnych cech w zbiorowości. Zyskuje jednakże możliwość wnioskowania o całej populacji.

Badania jakościowe natomiast koncentrują się na pogłębionym opisie postaw, motywów, sposobów postrzegania rzeczywistości, interpretacjach właściwych dla małych, konkretnych zbiorowości. Dąży się do pokazania specyfiki i wydobycia odmienności, charakterystycznej dla każdej grupy. Każdą z grup traktuje się jako zamknięty, niepowtarzalny „świat społeczny”. Stąd też bierze się niechęć badaczy do tworzenia generalizacji, wykraczających poza opis danej zbiorowości. W zamian za rezygnację z wnioskowania o całej populacji uzyskuje się możliwość formułowania pogłębionych interpretacji na poziomie konkretnych społeczności.

Powyższe przykłady pokazują, że wybór metody badawczej – czy to ilościowej, czy jakościowej – nie wiąże się z wyższością którejkolwiek z nich. Jest on natomiast podporządkowany oczekiwanemu zakresowi i charakterowi wniosków. W obu podejściach ujawnia się napięcie między podstawowymi celami badań społecznych: opisem, rozumieniem, wyjaśnianiem i przewidywaniem. Różnice dotyczą raczej rozłożenia akcentów niż wykluczenia któregoś z tych celów. Metody jakościowe zwracają się wyraźniej ku rozumieniu opisywanych zjawisk z perspektywy ich uczestników, ilościowe zaś zakładają konieczności rozumienia zjawisk i procesów, lecz pozwalają na ich wyjaśnianie oraz formułowanie przewidywań. W badaniach ilościowych badacz odkrywa i weryfikuje zależności ukryte „za” opiniami badanych, odwołując się do reprezentatywności przeprowadzonego pomiaru; w badaniach jakościowych natomiast zagłębia się on w opinie badanych, poszukując w nich wzorów służących pogłębionej interpretacji.

### **...a jednak bliskie**

Paradoksalnie, obie metody badawcze właśnie dzięki radykalnej odmienności założeń i celów analizy wyjątkowo dobrze się uzupełniają; można by nawet rzec, że są sobie niezbędne w całym procesie badawczym. Coraz częściej stawia się przed ewaluacją (szczególnie typu *ex-post*) cele wykraczające poza technokratyczność ocen, służących jedynie administracji publicznej. Coraz większego znaczenia nabiera społeczny aspekt ewaluacji sięgającej do takich zagadnień, jak dialog między grupami zainteresowanymi programem. Świadectwem tego może być chociażby powszechna dostępność wyników badań ewaluacyjnych dotyczących realizowanych programów społecznych. Jak pogodzić dwa – pozornie nie do pogodzenia – cele ewaluacji: „negocjowanie perspektyw, w jakich program jest postrzegany”, i wieloaspektowość

jego skutków oraz „systematyczną i obiektywną ocenę programu lub polityki”<sup>13</sup>?

Badania ewaluacyjne charakteryzuje wieloaspektowa ocena jakości i efektywności programów, projektów, a nawet polityk publicznych. Ich zalety to, z jednej strony, uwzględnianie perspektyw podstawowych grup zaangażowanych w proces będący przedmiotem badania (zarówno beneficjentów, jak i realizatorów projektu), z drugiej zaś – posługiwanie się wystandaryzowanymi wskaźnikami, konfrontowanymi z rzeczywistością. Ewaluacje, z jednej strony, wspierają tworzenie pozytywnego klimatu społecznego wokół badanego projektu, z drugiej – mają one badać efektywność ewaluowanych interwencji. Pogodzenie tych sprzecznych wymagań stawianych ewaluacji możliwe jest jedynie poprzez zastosowanie zróżnicowanych, lecz zintegrowanych metod postępowania badawczego – a więc **triangulacji**.

Istnieje kilka modeli sekwencji stosowania metod jakościowych i ilościowych w różnych momentach procesu badawczego. Przy formułowaniu pytań badawczych zarówno wyniki badań jakościowych, jak i ilościowych okazują się inspirujące. Znaczące obserwacje pewnego typu postaw mogą też stać się inspiracją do badań ilościowych nad powszechnością tych postaw w populacji. Z drugiej strony, badania wykazujące duże natężenie pewnych przekonań w populacji prowadzą nieraz do pogłębionych studiów przypadku, dotyczących kształtowania się takich przekonań w różnych środowiskach. Przygotowanie kwestionariusza z zasady poprzedza mniej lub bardziej zamaskowana forma analizy jakościowej, czy to opartej na bezpośredniej obserwacji, czy też na analizie literatury przedmiotu, danych zastanych albo też w różnym stopniu sformalizowanych wywiadów. Pytania w kwestionariuszu powinny być budowane nie tylko z myślą o poprawności merytorycznej; trzeba też, by uwzględniały kulturową adekwatność narzędzia. Dotyczy to zwłaszcza międzynarodowych badań porównawczych. Te same narzędzia/pytania mogą działać zupełnie odmiennie w różnych środowiskach. „Zrównywanie szans kobiet i mężczyzn”, „innovacyjność”, „prawica” w Polsce znaczy co innego niż w Holandii, nie mówiąc już o Hondurasie.

Za Milesem i Hubermanem<sup>14</sup> można wymienić trzy sposoby integracji metod ilościowych i jakościowych, które naszym zdaniem są przydatne w ewaluacji *ex-post* (patrz tab. 2).

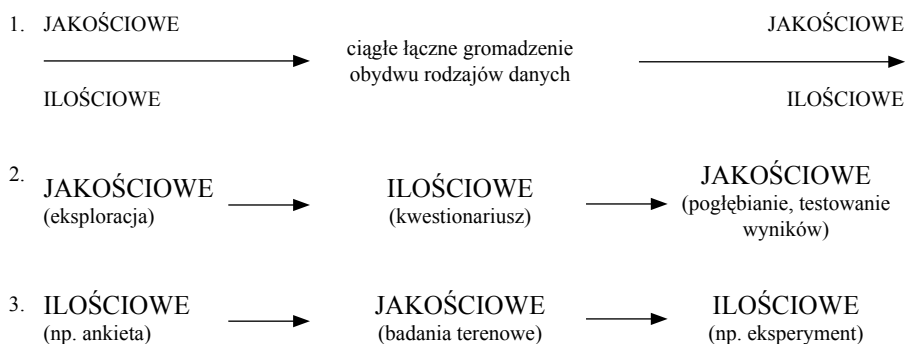
Istotą modelu pierwszego jest zachowanie odrębności i specyfiki obu podejść w momencie realizacji badań. Działania w ramach każdego z nich realizowane są równocześnie, a zarazem niezależnie od siebie – żadna z technik nie oddziałuje na drugą w trakcie przeprowadzania badania. Badacz formułuje wnioski z badań ilościowych niezależnie od wniosków z badań jakościowych. Rezultaty mogą potwierdzać się wzajemnie lub wykazywać

<sup>13</sup> Ekiert K., *Ewaluacja w administracji publicznej...*, op.cit.

<sup>14</sup> Miles M.B., Huberman A.M., *Analiza danych jakościowych*, przekł. S. Zabielski, Trans Humana, Białystok 2000.

sprzeczności. Jeśli będą ze sobą zgodne (obraz zbadanej rzeczywistości jest jednolity), to wnioski z całej ewaluacji pozwolą na bardziej jednoznaczną ocenę podjętych działań. Pokazuje to m.in. znaczenie ewaluacji *ex-post* jako „zwieńczenia” programu, którego częścią była ewaluacja. Odmienność wyników opartych na każdej z metod pozwala stwierdzić wielowymiarowość rzeczywistości, której nie byłaby w stanie uchwycić żadna z tych metod stosowana osobno. Jak pisaliśmy powyżej, oba te podejścia mają własną specyfikę i zakres wniosków, które można by w ramach danego podejścia formułować. Zatem nawet sprzeczne wnioski raczej będą uzupełniać się nawzajem niż znosić. Można tu mówić o superaddytywności metod ilościowych i jakościowych. Tak więc w żadnym wypadku podejścia te nie powinny służyć weryfikowaniu (z zwłaszcza fałsyfikowaniu) wniosków wywiedzionych na podstawie zastosowania drugiego podejścia. W związku z tym, często pojawiające się w raportach sformułowanie o weryfikowaniu wniosków z badań poprzez triangulację metod należy uznać za nadużycie metodologiczne.

Tabela 2.

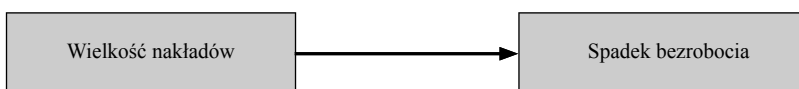


W praktyce najczęściej stosowane są modele: drugi i trzeci. Ich istotą jest to, że wnioski z jednego typu badań stanowią istotny czynnik, wykorzystywany do konstrukcji narzędzia badawczego stosowanego w ramach drugiego z podejść. Dzięki temu uzyskuje się stałe pogłębienie i/lub rozszerzenie wiedzy o badanej interwencji. Jedna metoda pozwala stawiać pytania, na które odpowiedzi poszukujemy w ramach drugiej, przy czym wyniki badań są interpretowane w ramach obu perspektyw. Znowu mamy więc do czynienia z komplementarnością omawianych podejść.

Jako ilustrację trzeciego modelu można wyobrazić sobie ewaluację programu przeciwdziałania pewnej patologii społecznej (narkomanii, przemocy wśród młodzieży, epidemii AIDS itp.). Badanie ilościowe (np. klasyczne badanie statystyczne czy analiza sieci społecznych na podstawie sondażu bądź danych GUS) określa najpowszechniejsze i najgroźniejsze przejawy danej patologii (rodzaje narkotyków, najbardziej patogenne środowiska i zacho-

wania). Kolejny etap polega na przeprowadzeniu badań jakościowych (terenowych). Ich celem jest sporządzenie szczegółowego opisu mechanizmów zachowań patologicznych oraz społecznego kontekstu ich występowania. Znaczenie takiej procedury polega ponadto na uwzględnieniu perspektywy badanych, niedostępnej ewaluatorowi w pierwszym etapie badań. Dzięki temu możliwe staje się rozszerzenie zakresu stawianych pytań badawczych, odnoszących się teraz do konkretnych środowisk, zachowań itp. Tym samym w kolejnym etapie zapewniona będzie większa trafność stawianych pytań. W modelowym przykładzie ewaluacji *ex-post* ostatni etap badań może stanowić eksperyment<sup>15</sup>. W praktyce przeprowadzenie eksperymentu jest niezwykle trudne. Badania ewaluacyjne wymagają zastosowania schematów *quasi*-eksperymentalnych. Wykorzystuje się schematy z uwzględnieniem szeregów czasowych, nierównoważnych grup kontrolnych itp.<sup>16</sup>. Poprzedzenie eksperymentu opisanymi powyżej działaniami badawczymi daje szansę na pełniejsze i bardziej wiarygodne wnioski z ewaluacji. Każde twierdzenie dotyczące efektywności programu opiera się na wnioskach o związkach przyczynowo-skutkowych. Wnioskowanie takie będzie trafniejsze, jeśli adekwatność pojęć i zjawisk, będących przedmiotem zainteresowania badacza i osób realizujących interwencję, została wcześniej zweryfikowana tak, jak to opisaliśmy powyżej.

Inny ważny obszar uzupełniania się obu podejść ilustruje poniższy przykład<sup>17</sup>. W ewaluacji hipotetycznego programu szkoleń, mających służyć zwiększeniu zatrudnienia bezrobotnych, stwierdzono związek statystyczny między wielkością środków finansowych przeznaczonych na szkolenia dla bezrobotnych a zaobserwowanym po interwencji wzrostem zatrudnienia. Związek ten wydaje się oczywisty (rys. 1).



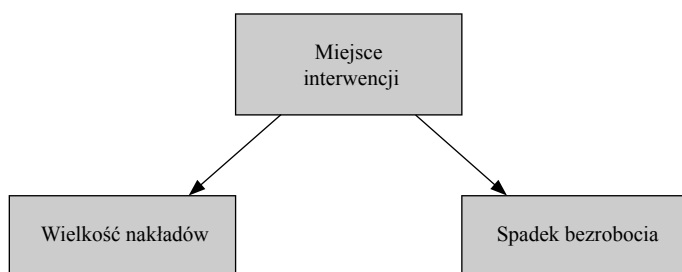
Rysunek 1.

<sup>15</sup> Jednoznaczna analiza przyczynowa może być prowadzona wyłącznie w warunkach klasycznego rozumianego eksperymentu. Większość realizowanych projektów uniemożliwia przeprowadzenie ewaluacji według klasycznego modelu badań eksperymentalnych. W rzeczywistości na ogół nie jesteśmy w stanie w losowy sposób dobrać ani grupy eksperymentalnej, ani grupy kontrolnej, ani też kontrolować w pełni oddziaływania bodźca. W powszechnie stosowanych modelach *quasi*-eksperymentalnych dowiedzenie związku przyczynowego jest również co najmniej dyskusyjne. Niezwykle często od ewaluatorów oczekuje się kategorycznej odpowiedzi na pytania, na które jednoznacznie odpowiedzieć się nie da. „Czy realizacja programu A jest przyczyną zaobserwowanego upowszechnienia cechy X w skali kraju?”. Jeżeli badacze postawieni przed takim problemem są odpowiedzialni, mogą jedynie przedstawić zestaw możliwych wyjaśnień.

<sup>16</sup> Babbie E., „Typy schematów badań ewaluacyjnych”, w: *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2003, s. 377–383.

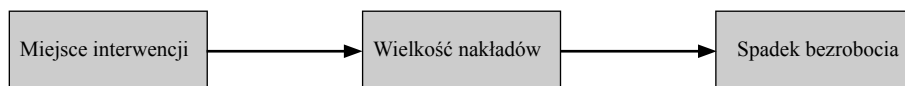
<sup>17</sup> Zob. Saris W.E., Stronkhorst H., *Causal Modelling in Nonexperimental Research*, Sociometric Research Foundation, Amsterdam 1984.

Jednakże pogłębiona analiza problemu wskazuje na inne możliwe wyjaśnienia stwierdzonej zależności. Zależność tę można bowiem przedstawić za pomocą kilku typów związków, a powyższy przykład (bezpośredni związek przyczynowy) stanowi tylko jeden z nich bądź jedną ze składowych związku. Takim wyjaśnieniem zależności jest, na przykład, zależność pozorna. Zarówno na wielkość nakładów, jak i na spadek bezrobocia wpływać mogło miejsce realizacji szkoleń. Większe nakłady przeznaczano na szkolenia w miastach. Równocześnie mieszkańcom większych miast łatwiej było znaleźć zatrudnienie. Ilustruje to poniższy rysunek (rys. 2).



Rysunek 2.

Według tego modelu, poprzedni, tak oczywisty związek przyczynowy, nie istnieje. Jest on związkiem pozornym. To jednak nie koniec możliwości interpretacji zaobserwowanej zależności. Być może rzeczywiście mamy do czynienia z wpływem, ale wpływem pośrednim. W takiej sytuacji wielkość miasta (miejsca interwencji) oddziaływałaby na ilość zainwestowanych w szkolenia pieniędzy i w ten sposób wpływałaby na redukcję bezrobocia. Tak więc pierwotną przyczyną spadku bezrobocia nie byłaby wielkość nakładów, lecz miejsce interwencji. Widać to na kolejnym schemacie (rys. 3).

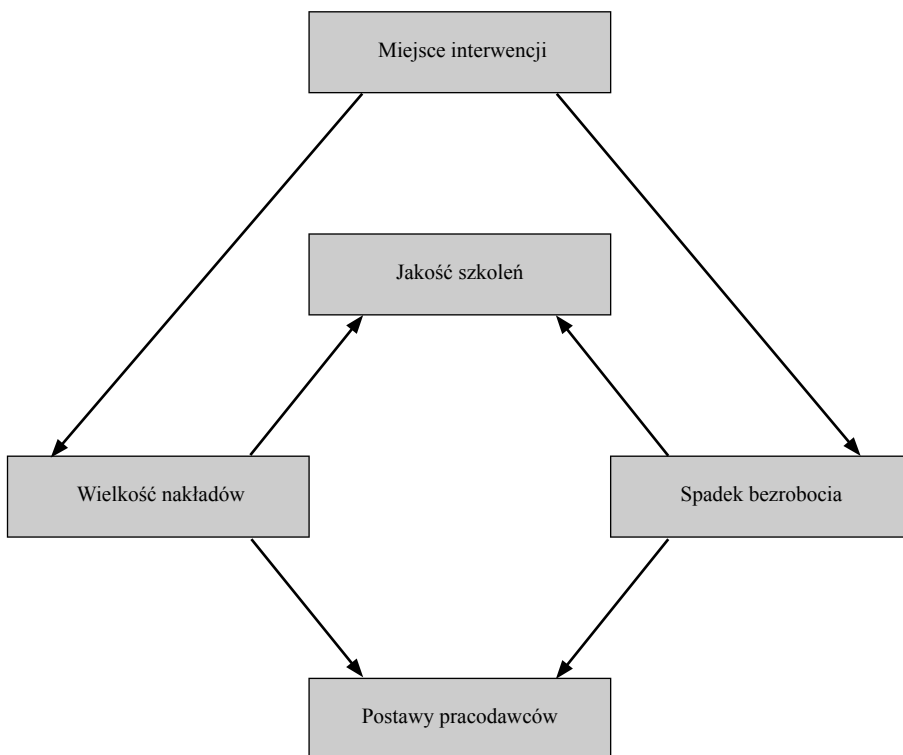


Rysunek 3.

Można by zaproponować jeszcze inne, bardziej złożone modele wyjaśnień zaobserwowanej zależności. Na przykład taki jak na poniższym rysunku (rys. 4).

Być może nakłady na program szkoleń dla bezrobotnych są tym większe, im większe jest miasto, w którym realizuje się interwencję. Z kolei wielkość nakładów wpływa na jakość organizowanych szkoleń oraz na postawy lokalnych pracodawców tworzących nowe miejsca pracy. Wzrost zatrudnienia jest

więc zdeterminowany wielkością miejscowości, jakością ośrodków szkoleniowych oraz opiniami pracodawców. Nie determinuje go zaś bezpośrednio wielkość zainwestowanych środków na szkolenia.



Rysunek 4.

Powyższy przykład przedstawia różne modele wyjaśniania zaobserwowanej zależności. Żaden z nich sam w sobie nie jest lepszy od innych. Co więcej, nie wyczerpują one wszystkich możliwości. Dobór zagadnień uwzględnianych w modelu zależy nie tylko od założeń i celów programu poddanego ewaluacji, ale, przede wszystkim, od kontekstowej analizy problematyki społecznej w badanej społeczności.

Pozostawienie ewaluatorów bez możliwości odniesienia modeli statystycznych do interpretacji jakościowej, odwołującej się do kontekstu społecznego, realizowanej polityki, uwarunkowań historycznych i kulturowych, możliwości ekonomicznych itp., powoduje, że rezultaty ich badań nie mogą wykroczyć poza poziom rozważań niezwiązanych z rzeczywistością. Są przez to nieużyteczne z punktu widzenia celów ewaluacji *ex-post*, nierozzerwalnie związanej z realizacją konkretnych programów interwencji społecznych.

Nawet określanie efektu netto – wydawałoby się, czysto ilościowy aspekt ewaluacji *ex-post* – wymaga sprzężenia obu podejść. Zarówno nakłady ponie-

sione na realizację programu (nie tylko mierzone w euro), jak i osiągnięte rezultaty zawsze tworzą zmienne społeczne, zwane cechami ukrytymi, których pomiar wymaga zestawu wskaźników niemożliwych do zidentyfikowania bez jakościowej analizy postrzeganych zjawisk. Rezultaty przeprowadzonego szkolenia wykraczają poza kwantyfikowalną wiedzę uczestników przedsięwzięcia. Kapitał społeczny powstały w trakcie szkoleń tworzy, poza nabytą wiedzą, cała sieć nieformalnych więzi oraz tzw. inteligencja emocjonalna, na którą składa się m.in. motywacja, zdolność do empatii, poczucie własnej wartości, umiejętności zespołowe itp. Przyrostowi wiedzy uczestników szkolenia towarzyszyć może jednak również drastyczny spadek motywacji. Co w takim wypadku jest efektem netto szkolenia?

Próby poznania przy użyciu technik jakościowych perspektyw podzielanych przez uczestników życia zbiorowego umożliwiają wieloaspektową interpretację wyników analiz statystycznych. Zależności uzyskiwane na podstawie analiz ilościowych mogą być lepiej rozumiane i zaprezentowane, jeśli analizy te są sprzęgnięte z refleksją uwzględniającą jakościowe ujęcie problemu. Ich wyjaśnienia to w istocie modele rzeczywistości, których nie należy traktować jako ostatecznej prawdy o rzeczywistości. Wymagają więc interpretacji na bazie przyjętej teorii społecznej. Częstym błędem jest określanie z góry znaczenia wartości mierników siły zależności. Nieraz twierdzi się, że jeśli miernik przyjmuje wartość od 0 do  $1/3$ , wówczas zależność ta jest słaba, jeśli przyjmuje wartość od  $1/3$  do  $2/3$  – średnia, jeśli zaś przyjmuje wartość od  $2/3$  do 1, to możemy powiedzieć, iż mamy do czynienia z zależnością silną. Rzecz w tym, że interpretacja niemal całkowicie zależy tutaj od charakterystyki badanej zbiorowości. Jeśli zbadamy wśród radnych i wśród zwykłych mieszkańców powiatu zależność między politycznymi identyfikacjami a działaniami politycznymi (np. głosowanie w wyborach samorządowych, referendum powiatowe itp.), to wartość 0,84 uzyskana wśród działaczy politycznych nie powinna w ogóle dziwić. Działania polityczne (na poziomie powiatu) stanowią bowiem ich zawód. Taka sama wartość miernika, uzyskana wśród zwykłych mieszkańców, oznaczałaby tak nadzwyczajną mobilizację polityczną w powiecie, jakiej praktycznie się nie spotyka. W tej grupie prawdopodobnie zależność na poziomie 0,25 można by uznać za silną. Zależy to jednak od cech danej zbiorowości. Badacz jest w stanie je ocenić, posługując się doświadczeniem, które wykracza poza warsztat statystyczny i bazuje na kontakcie z członkami grupy. Innym powszechnym nieporozumieniem pojawiającym się w praktyce ewaluacji *ex-post* jest utożsamianie wykrytych przez badacza zależności statystycznych ze związkami przyczynowymi. Ujawniona w badaniach korelacja stanowi w istocie zestawienie różnorodnych związków między badanymi cechami. Bezpośredni związek przyczynowy, często poszukiwany w badaniach ewaluacyjnych, jest zaledwie jednym z typów możliwych relacji w ramach ujawnionej zależności. Kwestię tę zilustrowaliśmy już przykładem dotyczącym uwarunkowań redukcji bezrobocia.

## Podsumowanie

Jak staraliśmy się pokazać, istnieje możliwość integracji podejścia ilościowego i jakościowego. Mimo zasygnalizowanych na wstępie różnic (albo, paradoksalnie, ze względu na nie) metody te nie są konkurencyjne, można je łączyć i z powodzeniem stosować w konkretnych ewaluacyjnych przedsięwzięciach badawczych.

Kluczową kwestią w badaniach społecznych (w szczególności zaś w badaniach ewaluacyjnych) jest świadomość, że metody zawsze są wtórne wobec stawianych problemów i pytań badawczych.

Każdy z przedstawionych powyżej punktów wyznacza pole dla realizacji triangulacji metodologicznej. Powyższe rozważania odnoszą się zaledwie do jednego z wymienionych aspektów triangulacji. W podobny sposób można by postulować różnorodność typów danych wykorzystywanych w badaniach, źródłach informacji, do których badacz powinien sięgać itp. Każdemu z tych zagadnień mógłby być poświęcony osobny artykuł.

Podobnie można by mnożyć przykłady błędów, popełnianych na skutek niezrozumienia specyfiki i możliwości poszczególnych podejść. Tutaj pragniemy jedynie jeszcze raz podkreślić wzajemną niezastępowalność metod jakościowych i ilościowych. Każda z nich ma swoje zalety, lecz również ograniczenia. Nie można, opierając się na trzydziestu zogniskowanych wywiadach grupowych (tzw. fokusach), wyciągać wniosków tak samo, jak na podstawie 400-osobowej próby losowej. Analogicznym błędem byłoby wnioskowanie jakościowe na podstawie wyników badań kwestionariuszowych z tzw. pytaniami zamkniętymi. Tego typu badania nie pozwalają na wywodzenie wniosków wykraczających poza ustaloną kafeterię. Nie pozwalają więc także na rekonstrukcję sposobu rozumienia pojęć przez badanych.

W naszych rozważaniach skupiliśmy się na opisanu zastosowań i znaczenia postulatu metodologicznego zwanego triangulacją. Staraliśmy się wykazać, iż jakość badania ewaluacyjnego i użyteczność płynących z niego wniosków zależy w dużym stopniu od umiejętnie stosowanej i właściwie rozumianej triangulacji. Wyrazem jej zrozumienia powinno być nie tylko korzystanie z metod ilościowych i jakościowych, lecz także, co ważniejsze, świadomość ich ograniczeń, odrębności i nieprzekładalności. W każdym projekcie badawczym można pogodzić obie perspektywy – ilościową i jakościową. W stosunku do siebie są one bowiem komplementarne. Z całą odpowiedzialnością możemy stwierdzić, że umiejętnie integrowane podejścia badawcze służą wzbogaceniu i podniesieniu jakości badań ewaluacyjnych, a także jakości wniosków oraz zwiększeniu efektywności ewaluowanych programów.

---



**Mikołaj Jasiński** – doktor nauk społecznych, adiunkt w Zakładzie Statystyki, Demografii i Socjologii Matematycznej IS UW, kierownik Biura ds. Jakości Kształcenia, w tym Pracowni Ewaluacji Jakości Kształcenia Uniwersytetu Warszawskiego. W ramach zadań Pracowni, współtwórca (razem z dr. Michałem Kowalskim) wdrażanego obecnie systemu ewaluacji jakości kształcenia na UW. Wykładowca na szkoleniach (MRR oraz IS UW) z dziedziny metodologii badań społecznych, w tym ewaluacji. Wykładowca w Podyplomowym Studium Ewaluacji Programów Społecznych w IS UW. Wybrane programy badawcze, w których ostatnio uczestniczył lub uczestniczył: *Badanie ewaluacyjne działań na rzecz kształcenia i szkolenia w ramach EFS* (badanie realizowane dla MRR w ramach konsorcjum Polskiego Towarzystwa Socjologicznego i Agrotec Polska, 2007–2008), *Polacy 2007 – wspólnota i tożsamość* (badanie ogólnopolskie realizowane przez Centrum Myśli Jana Pawła II, 2007–2008), *Wpływ przestrzeni ideologicznej w Sejmie RP na znaczenie ugrupowań parlamentarnych podczas głosowań* (projekt realizowany w IS UW, 2007), *Młodzież, imigranci, tolerancja* (w ramach Towarzystwa Demokratycznego „Wschód” 2007), *SHARE* (razem z WNE UW i Mannheim Research Institute for the Economies of Aging, 2006–2008), *Monitoring przestrzegania praw ucznia w szkole* (ogólnopolski program badawczy realizowany przez CODN, 2005–2006).

Autor publikacji i referatów m.in. na temat metodologii badań ilościowych, procesów decyzyjnych w zgromadzeniach politycznych, ewaluacji, zjawisk społecznych w środowisku szkolnym.

**Michał W. Kowalski** – doktor nauk społecznych, adiunkt w Instytucie Socjologii Uniwersytetu Warszawskiego (Zakład Antropologii Społecznej). Z badaniami ewaluacyjnymi związany od połowy lat dziewięćdziesiątych, członek-założyciel Polskiego Towarzystwa Ewaluacyjnego. Uczestniczył w licznych programach badawczych, prowadzi aktywną działalność dydaktyczną i szkoleniową z zakresu ewaluacji i metodologii jakościowej badań społecznych. Od 2007 roku w zespole Pracowni Ewaluacji Jakości Kształcenia na Uniwersytecie Warszawskim.

## Bibliografia

- Babbie E., „Typy schematów badań ewaluacyjnych”, w: *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2003, s. 377–383.
- Brouwer A., *Wdrażanie Europejskiego Funduszu Rozwoju Regionalnego*, Projekt Phare PL/IB/2002/OT/03C 2002.
- Bryman A., *Quantity and Quality in Social Research*, Routledge, London 1988.
- Ekiert K., *Ewaluacja w administracji publicznej, funkcje, standardy i warunki stosowania*, Departament Spraw Społecznych i Ustrojowych 2006.
- Gubrium J.F., Holstein J.A., „Qualitative methods”, w: E.F. Borgotta, M.L. Borgotta (red.), *Encyclopedia of Sociology*, Macmillan, New York 1992.
- Jick T.D., „Mixing qualitative and quantitative methods: triangulation in action”, *Administrative Science Quarterly* 1979, nr 24(4), s. 602–611.
- Hammersley M., Atkinson P., *Metody badań terenowych*, przekł. S. Dymczyk, Zysk i S-ka, Poznań 2000.
- <http://www.konferencja-ewaluacja.pl/2005/materialy.php>

- Lissowski G., „Metody ilościowe”, w: *Encyklopedia socjologii*, Oficyna Naukowa, Warszawa 2002.
- Miles M.B., Huberman A.M., *Analiza danych jakościowych*, przekł. S. Zabielski, Trans Humana, Białystok 2000.
- Patton M.Q., *Qualitative Evaluation and Research Method*, second edition, Sage Publications, London 1990.
- Saris W.E., Stronkhorst H., *Causal Modelling in Nonexperimental Research*, Sociometric Research Foundation, Amsterdam 1984.
- Silverman D., *Doing Qualitative Research. A Practical Handbook*, Sage Publications, London 2000.
- Sulek A., *Ogród metodologii socjologicznej*, Wydawnictwo Naukowe Scholar, Warszawa 2002.

Jarosław Górniak, Barbara Worek, Seweryn Krupnik

## Zastosowanie podejścia badań jakościowych w ewaluacji *ex-post*

### Wstęp

W ewaluacji *ex-post* jesteśmy zainteresowani przede wszystkim całościową oceną programu, projektu czy interwencji. Nie tylko dążymy wówczas do oceny podjętych działań, wypracowanych produktów czy bezpośrednich rezultatów tych działań, lecz także interesujemy się wpływem danej interwencji na szersze otoczenie, np. społeczno-ekonomiczne. Istotnym elementem oceny programów jest także wyjaśnienie sukcesu lub niepowodzenia, analiza samego procesu wdrażania i odkrycie możliwości jego usprawnień i innowacji.

Jednym z kluczowych zadań ewaluacji jest ustalenie związku przyczynowego między podjętą interwencją a obserwowanymi wynikami. Wymaga to zaś:

- zmierzenia efektu całkowitego interwencji jako ogólnej zmiany w zakresie zdefiniowanym przez cel (zmiany na poziomie przyjętych wskaźników),
- oddzielenia zmiany niezależnej od podjętej interwencji od zmiany, którą można przypisać wpływowi interwencji (efekt netto).

Takie rozumienie zadania zwraca uwagę raczej na ilościowe podejście do badań. Podstawą uzyskania wyników jest obiektywny, rzetelny, trafny, standaryzowany pomiar efektów programu, dokonywany na reprezentatywnej, losowej próbie osób czy innych jednostek badania, lub na podstawie specjalnie dobranej próby kontrolnej, pozwalającej oszacować efekt kontrfaktycznej sytuacji braku oddziaływania na aktualnych beneficjentów. Można więc spytać – gdzie tu jest miejsce na podejście jakościowe? Postaramy się podać odpowiedź na to pytanie.

Każda interwencja publiczna, niezależnie od swej genezy, ma za podstawę swoistą hipotezę („teorię programu”): w danych warunkach podjęcie działań wynikających z projektu interwencji doprowadzi do osiągnięcia założonych rezultatów. Polityki publiczne opierają się na takich przyczynowych konstrukcjach. Nie zawsze wyraża się je *explicite* – wręcz przeciwnie, często ukryte są pod powierzchnią formułowanych celów i przyjmowanych sposobów działania. Te ostatnie bywają zresztą ułomną operacjonalizacją istoty teorii

programu. Aby dobrze przeprowadzić ewaluację, należy właściwie uchwycić istotę analizowanej interwencji, czyli jej teorię. Teorii programu, to znaczy założeniom odnoszącym się do związków przyczynowych pomiędzy podejmowanymi działaniami a pożądanymi efektami, towarzyszy na ogół zbiór przekonań dotyczących najwłaściwszych sposobów wdrażania interwencji. Efektem takiej „teorii wdrażania” jest przyjmowanie rozwiązań w zakresie struktur wdrażania (np. centralizowane *versus* zdecentralizowane), sposobu zarządzania, organizacji monitoringu, kontroli finansowej itp. Rozpoznanie teorii wdrażania, dominującej w gronie decydentów i osób projektujących interwencję, pozwala lepiej zrozumieć politykę czy program i ich pogłębioną ewaluację<sup>1</sup>. Ewaluacja (zwłaszcza końcowa) sprawdza, czy w świetle uzyskanych efektów teorię programu można utrzymać, czy też należy ją odrzucić. Koncepcje działania, które pomyślnie przejdą ewaluację, tworzą korpus wiedzy podnoszącej skuteczność polityk publicznych. Faktycznie, zawsze mamy do czynienia ze spontanicznym procesem weryfikacji w postaci społecznej oceny skutków prowadzonych polityk. Ewaluacja jest metodologicznie kontrolowanym i usystematyzowanym sposobem weryfikacji, stosowanym przez ekspertów. Z natury rzeczy zatem przyczynia się do legitymizacji lub delegitymizacji efektywnościowego sposobu rządzenia. Bezpośrednią konsekwencją dysponowania wiedzą o warunkach dobrego działania jest możliwość formułowania rekomendacji usprawniających. Wszystkie te zadania ewaluacji można podporządkować jej trzem zasadniczym funkcjom. Są to<sup>2</sup>:

- funkcja rozliczenia i egzekwowania odpowiedzialności (*accountability*),
- funkcja poznawcza (*knowledge*),
- funkcja stymulowania usprawnień i rozwoju organizacyjnego (*development*).

Czy w realizacji tak określonych celów i funkcji ewaluacji mogą być pomocne metody jakościowe? Stanowiska badaczy są w tej kwestii podzielone: niektórzy twierdzą, że metody jakościowe sprawdzają się głównie w ewaluacji o charakterze formatywnym, zmierzającej do usprawnienia procesu wdrażania interwencji. Pozwalają one bowiem zdobyć wiedzę, dzięki której można usprawnić przebieg tego procesu, np. przez skorygowanie założonych sposobów działania, tak aby stały się bardziej adekwatne do sytuacji, w jakiej są stosowane. Inni zaś uważają, że metody jakościowe mogą być wykorzystywane w każdym typie ewaluacji, gdyż zawsze wnoszą do tego procesu określone korzyści, do których należą między innymi: możliwość ujawniania niezamierzonych skutków podejmowanych działań, spojrzenie na dokonujące się procesy oczyma ich uczestników, rozpoznanie możliwych (często sprzecznych) interesów różnych aktorów oraz sposobów postrzegania

<sup>1</sup> Por. Brickmayer J.D., Weiss C.H., „Theory-based evaluation in practice. What do we learn?”, *Evaluation Review* 2000, vol. 24, nr 4, s. 407–431.

<sup>2</sup> Por. np. Chelimsky E., „The coming transformations in evaluation”, w: E. Chelimsky, W.R. Shadish (red.), *Evaluation for the 21st Century*, Sage Publications, London 1997.

i rozumienia przez nich sytuacji, dostrzeżenie ukrytych barier i ograniczeń utrudniających realizację zamierzonych działań.

Niezależnie jednak od szerokiej możliwości wykorzystania metod jakościowych, zarówno praktyka badawcza, jak i literatura wskazują, że w ewaluacji polityk publicznych dominuje wykorzystanie podejścia „twardego”, czyli eksperymentalnego lub sondażowego, pozwalającego uzyskać „twarde”, ilościowe dane. Wbrew częstym wskazaniom na to, iż decydenci łatwo percypują wyniki badań jakościowych, zdaje się, że to wyniki badań ilościowych są łatwiej przyjmowane, uznawane za bardziej zrozumiałe i wiarygodne. Badacz jakościowy, prezentując wyniki przeprowadzonych badań, staje często przed koniecznością szczegółowego wyjaśnienia specyfiki podejścia jakościowego, wskazywania, czego od takich badań można oczekiwać, a czego nie. Jednocześnie, ponieważ wyników badań jakościowych nie można rozszerzyć na populację, decydenci czasami kwestionują ich użyteczność. Należy jednak zaznaczyć, że często sami badacze jakościowi nie są tutaj bez winy, gdyż, prezentując wyniki swoich badań, nie uwzględniają celów i perspektywy decydentów, którzy są zainteresowani uzyskaniem precyzyjnych i rzetelnych informacji, pozwalających im ocenić skutki podjętych działań, wskazać dalsze perspektywy, wyzwania i możliwe sposoby działania. Przekazywaniu tak precyzyjnych informacji nie służy narracyjne podejście do prezentowania danych, nagromadzenie cytatów z wypowiedzi respondentów i formułowanie wniosków, które nie mają dostatecznej podstawy w postaci konkretnych danych. Tymczasem poprawnie przeprowadzone badania jakościowe, z rzetelnie przeprowadzoną analizą danych, obejmującą ich redukcję, kategoryzację oraz wyprowadzanie i weryfikację wniosków, mogą być nieocenionym źródłem użytecznej wiedzy. Przydatność tego typu badań doceniono już dawno w badaniach marketingowych, gdzie ich popularność systematycznie wzrasta, a wachlarz wykorzystywanych technik staje się coraz szerszy. Niewątpliwie, taki proces zajdzie też w przypadku badań ewaluacyjnych, jednak, aby usprawnić jego przebieg, warto zwrócić uwagę na kilka zasadniczych kwestii związanych ze specyfiką metod jakościowych.

## **Metody jakościowe i ich miejsce w ewaluacji *ex-post***

Zarówno w klasycznych podręcznikach metodologicznych, jak i w publikacjach oraz opracowaniach poświęconych ewaluacji, podkreśla się podstawowy fakt prymatu problemu nad metodą badawczą<sup>3</sup>. To metody mają być dopasowane

<sup>3</sup> Por. np. Babbie E., *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2003; Sztumski J., *Wstęp do metod i technik badań społecznych*, Wydawnictwo Śląsk, Katowice 2005; Tavistock Institute, *The Evaluation of Socio-Economic Development. The Guide*, Tavistock Institute 2003; Ekiert K., *Ewaluacja w administracji publicznej*, Rządowe Centrum Studiów Strategicznych, Warszawa 2004.

wane do problemu, który rozwiązujemy, nie zaś odwrotnie. Wiedząc, przed jakimi problemami stajemy w ewaluacji *ex-post*, powinniśmy sprawdzić, czy i w jakim zakresie w ich rozwiązaniu może nam pomóc zastosowanie metod jakościowych. Spójrzmy więc najpierw, co dałoby nam zastosowanie metod jakościowych w całym procesie ewaluacji, a następnie zastanówmy się, jakie korzyści z tych metod odnieśliśmy w interesującym nas rodzaju ewaluacji.

Najogólniej rzecz ujmując, celem badań jakościowych jest uzyskanie pogłębionego zrozumienia doświadczeń ludzi, ich historii, perspektyw i planów w kontekście szerszego środowiska, w jakim funkcjonują. Charakteryzując badania jakościowe, najczęściej wskazuje się na to, że:

- pozwalają widzieć zjawiska oczyma ich uczestników,
- wykorzystuje się w nich nieustrukturyzowane lub słabo ustrukturyzowane metody, które można elastycznie dopasowywać do społecznego kontekstu badań,
- koncentrują się przede wszystkim na pytaniach: „co?”, „jak?”, „dlaczego?”,
- otrzymuje się dzięki nim wiele szczegółowych i złożonych informacji,
- proces analizy i wnioskowania ma charakter raczej indukcyjny niż dedukcyjny,
- wypracowuje się w nich wyjaśnienia, które mają zastosowanie w określonym kontekście społeczno-kulturowym, pozwalają zrozumieć procesy zachodzące w konkretnym środowisku, instytucji, społeczności; nie stanowią natomiast podstaw do tworzenia ogólnie obowiązujących praw, niezależnych od kontekstu<sup>4</sup>.

Dzięki specyficznemu podejściu i wykorzystywanym technikom, metody jakościowe pozwalają więc w sposób pogłębiony analizować działania i ich znaczenie, identyfikować różne perspektywy, ujmować dynamikę czy procesualność zjawisk, co sprawia, że mogą one być szczególnie przydatne w badaniach ewaluacyjnych. Jak wskazuje Janet Finch, to właśnie te cechy badań jakościowych czynią je atrakcyjnymi dla decydentów i sprawiają, że odgrywają one dużą rolę w ocenie polityk publicznych<sup>5</sup>. Dzieje się tak dlatego, iż metody te pozwalają lepiej opisać i zrozumieć rzeczywiste skutki podejmowanych działań, porównać założenia, które leżały u podstaw działań, z praktyką społeczną i ich wynikami oraz pomóc w ocenie dopasowania tych działań do faktycznych potrzeb<sup>6</sup>.

Zdaniem wielu autorów<sup>7</sup>, metody jakościowe są w ewaluacji szczególnie przydatne wtedy, gdy:

---

<sup>4</sup> Por. Spencer L., Ritchie J., Lewis J., Dillon L., *Quality in Qualitative Evaluation: a framework for assessing research evidence*, National Centre for Social Research 2003, s. 17.

<sup>5</sup> Finch J., *Research and Policy: the Uses of Qualitative Methods in Social and Educational Research*, Falmer Press, London 1986.

<sup>6</sup> Por. ibidem.

<sup>7</sup> Por. np. Rist R.C., „Influencing the policy process with qualitative research”, w: N.K. Denzin, Y.S. Lincoln (red.), *Handbook of Qualitative Research*, Sage Publications, Thousand Oaks 2000;

- istotne jest doprecyzowanie pytań ewaluacyjnych i celów prowadzonej ewaluacji,
- ważne jest szczegółowe zbadanie, jak faktycznie została zrealizowana dana interwencja oraz poznanie mechanizmów decydujących o przebiegu jej implementacji i czynników wpływających na ten proces,
- zależy nam na zidentyfikowaniu niezamierzonych skutków podjętych działań,
- ewaluacja ma charakter formacyjny, a uzyskane wyniki badań i same działania badawcze mogą przyczynić się do usprawnienia wdrażania programu czy przedsięwzięcia,
- osiągnięcie założonych rezultatów interwencji wymaga podjęcia złożonych działań w określonym środowisku,
- zależy nam na opracowaniu takich sposobów mierzenia rezultatów programu, aby uwzględniały one subiektywne perspektywy uczestników przedsięwzięcia,
- chcemy dokładniej wyjaśnić, jak powinniśmy zdefiniować sukces naszych działań i ich skuteczność,

potrzebujemy wsparcia informacyjnego podczas projektowania i realizacji badań ilościowych; np. gdy projektujemy standaryzowane narzędzia badawcze lub chcemy wyjaśnić zaskakujące wyniki uzyskane w badaniach ilościowych. Próbuując nieco bardziej zsyntetyzować cele, które da się zrealizować przy pomocy metod jakościowych w ewaluacji, można odwołać się do dwóch zasadniczych modeli, opracowanych w obszarze ewaluacji działań z zakresu ochrony zdrowia. Jak piszą J. Popay i G. Williams<sup>8</sup>, przywołując typologię N. Blacka, wypracowano obecnie dwa zasadnicze sposoby czy modele wykorzystywania metod jakościowych: **model wspierania** lub **uzupełniania** (*enhancement model*) oraz **model różnicy** (*difference model*). Zgodnie z pierwszym modelem, zakładającym uzupełniającą rolę metod jakościowych, badania jakościowe mają pomóc zrozumieć, dlaczego procesy przebiegają tak, a nie inaczej, dostarczyć danych pozwalających zwiększyć trafność badań ilościowych i zidentyfikować zmienne, które powinny być w tych badaniach uwzględnione, pomóc w sformułowaniu hipotez, weryfikowanych następnie w badaniach ilościowych oraz wyjaśnić niespodziewane wyniki badań ilościowych. W drugim modelu zwraca się uwagę na odmienne funk-

---

Patton M.Q., *Qualitative Research and Evaluation Methods*, Sage Publications, Thousand Oaks 2002; Davies P., „Contributions from qualitative research”, w: H.T.O. Davies, S. Nutley, P.C. Smith (red.), *What Works? Evidence-based Policy and Practice in Public Services*, Policy Press, Bristol 2000; Williams D.D., „When is naturalistic evaluation appropriate?”, w: D.D. Williams (red.), *Naturalistic Evaluation, New Directions for Program Evaluation No. 30*, Jossey-Bass, San Francisco 1986.

<sup>8</sup> Popay J., Williams G., „Qualitative research and evidence-based healthcare”, *Journal of the Royal Society of Medicine* 1998, Supplement, vol. 191, nr 35, s. 32–37. Por. także: Black N., „Why we need qualitative research?”, *Journal of Epidemiology and Community Health* 1994, nr 48, s. 425–426.

cje badań jakościowych, podkreślając, że ich celem powinno być wyjaśnianie procesów, na które często nie zwraca się uwagi, zrozumienie zwyczajnych zachowań, a także badanie tego, jak zwykli ludzie postrzegają celowość danej interwencji, jakość podejmowanych działań i ich rezultatów; istotne jest też rozumienie kultury organizacyjnej danej instytucji i wykorzystanie podejścia jakościowego do oceny złożonych interwencji<sup>9</sup>. W pierwszym modelu zakłada się więc pomocniczą rolę badań jakościowych w stosunku do badań ilościowych, w drugim zaś akcentuje się swoistą autonomię badań prowadzonych metodami jakościowymi. Należy przy tym zwrócić uwagę na niebezpieczeństwo wartościowania tych dwóch modelowych podejść i uznawania jednego z nich za lepsze, a drugiego za gorsze. Formułowanie takich ocen jest całkowicie bezzasadne, zwłaszcza w kontekście przywołanej już tutaj zasady prymatu problemu nad metodą. W zależności od problemu, który aktualnie rozwiązujemy, metody jakościowe możemy zarówno zastosować pomocniczo (będą wówczas stanowić wartościowe uzupełnienie badań ilościowych), jak i wykorzystać z powodzeniem do samodzielnego opisu i wyjaśniania zjawisk o takim zakresie i charakterze, że podejście jakościowe jest odpowiednie i wystarczające<sup>10</sup>.

Kwestią kluczową wykorzystania metod jakościowych w ewaluacji *ex-post* jest ich użyteczność w zrozumieniu mechanizmu działania programu, wyjaśnianiu sposobu, w jaki uzyskano jego efekty oraz przewidywaniu możliwości replikacji tychże efektów. W dalszej części artykułu opiszemy podejścia i narzędzia, służące realizacji następujących zadań: ewaluacji opartej na teorii programu oraz wyjaśniania i przewidywania z wykorzystaniem metod jakościowych.

## Metody jakościowe a możliwość wyjaśniania i przewidywania

Wśród omówionych wyżej zadań, przy których realizacji używa się w ewaluacji metod jakościowych, znajdują się też zadania charakterystyczne dla ewaluacji *ex-post*. Należą do nich między innymi ocena zamierzonych skutków podjętych działań, identyfikacja skutków niezamierzonych oraz opisanie i wyjaśnienie przebiegu danej interwencji i mechanizmów decydujących o jej sukcesie czy niepowodzeniu. Wykorzystanie do realizacji tych zadań metod jakościowych pozwala zdobyć pełniejszą wiedzę o przebiegu danych procesów (np. o implementacji działań ukierunkowanych na rozwijanie przedsiębiorczości wśród studentów), a także daje możliwość szerszego i pełniejsze-

<sup>9</sup> Por. Spencer L., Ritchie J., Lewis J., Dillon L., *Quality...*, op.cit., s. 36.

<sup>10</sup> Problematyka relacji pomiędzy metodami jakościowymi i ilościowymi została już omówiona w znajdującym się w tym tomie artykule Mikołaja Jasińskiego i Michała Kowalskiego. Szczególnie użyteczna jest tutaj opisywana przez Autorów koncepcja triangulacji, w tym, przede wszystkim, interesująca nas idea triangulacji metod badawczych.



go opisanie osiągniętych efektów, zarówno tych dających się łatwo zmierzyć i ocenić, jak i tych, które nie poddają się tak łatwo ocenie (np. społeczne oddziaływanie podjętych działań, wpływ danego programu nie tylko na jego bezpośrednich beneficjentów, lecz także na ich rodziny, społeczności lokalne, w których żyją i, o czym często zapominamy, na same osoby i instytucje realizujące program).

Dla zebrania materiału badawczego, koniecznego do realizacji tak określonych celów badawczych możemy wykorzystać klasyczne techniki badań jakościowych: indywidualne wywiady pogłębione, wywiady eksperckie, zogniskowane wywiady grupowe (fokusy), panelowe dyskusje eksperckie; czasem stosujemy także elementy obserwacji, pomocna może być analiza treści dokumentów programowych, projektów oraz innych materiałów związanych z realizowanym przedsięwzięciem<sup>11</sup>.

Gdy przeprowadzamy ewaluację *ex-post*, zazwyczaj część danych już zebrano; czasem konieczne jest tylko ich uzupełnienie, zwłaszcza o informacje dotyczące wpływu danego przedsięwzięcia na różne kategorie beneficjentów i innych podmiotów, mogących w jakiś sposób odczuwać jego skutki. Osoba przystępująca do analizy tych materiałów ma więc do dyspozycji wiele danych o charakterze jakościowym: nagrania wywiadów czy fokusów, transkrypcje sporządzone na podstawie tych nagrań, notatki terenowe badaczy i ich uwagi, artykuły prasowe, dokumenty urzędowe, czasem również zdjęcia czy inne obrazy, stanowiące uzupełnienie informacji tekstowych. Dane te są zazwyczaj rozproszone, a ich ilość może przytłoczyć nie tylko odbiorców wyników badań czy decydentów, lecz także samych badaczy, którym w ogromnej ilości materiałów nieraz trudno jest dostrzec rysujące się wzory i prawidłowości czy zidentyfikować związki między kluczowymi zmiennymi i wyjaśnić pojawiające się tendencje. Tymczasem takie właśnie wnioski mogą decydować o sukcesie realizowanych badań, a przede wszystkim o użyteczności ich wyników dla decydentów. Oczywiście, istotne jest też umiejętne zaplanowanie badania, zebranie danych o odpowiednio wysokiej jakości, ich rejestracja, następnie zaś wstępna redukcja (czyli np. sporządzenie transkrypcji), a potem kategoryzacja, do której możemy wykorzystać procedurę kodowania dostępną w wielu programach komputerowych do analizy danych jakościowych, m.in. w MAXqda, QDA Miner, Atlas.ti. Jednak o efekcie końcowym

<sup>11</sup> Techniki badawcze wykorzystywane w badaniach jakościowych są stosunkowo dobrze opisane w literaturze przedmiotu, dlatego też nie będziemy tutaj poświęcać im więcej uwagi. Zainteresowanych odsyłamy do podstawowych podręczników metodologii badań społecznych (np. Babbie E., *Badania społeczne...*, op.cit.; Sztumski J., *Wstęp...*, op.cit.; Konecki K., *Studia z metodologii badań jakościowych. Teoria ugruntowana*, Wydawnictwo Naukowe PWN, Warszawa 2000), a także podręczników badań marketingowych (np. Maison D., *Zogniskowane wywiady grupowe. Jakościowa metoda badań marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 2001). Dostępnych jest też wiele publikacji anglojęzycznych opisujących zastosowanie technik badań jakościowych w ewaluacji, por. np. Patton M.Q., *Qualitative Evaluation and Research Methods*, Sage Publications, Beverly Hills 1990.

całego przedsięwzięcia, a zwłaszcza o jego wartości dla decydenta, decyduje właśnie umiejętność syntetyzowania danych, formułowania wniosków i rzetelnej oceny ich jakości. Systematyczny opis tego zagadnienia zdecydowanie wykracza poza zakres tego artykułu, toteż zaprezentujemy tutaj jedynie wybrane techniki, pomocne podczas formułowania wniosków na podstawie danych jakościowych, oraz metody oceny i weryfikacji tych wniosków.

### **Techniki pomocne w procesie formułowania wniosków z badań jakościowych**

Do technik przydatnych w procesie wyciągania wniosków na podstawie danych jakościowych należy kilka opisanych przez M.B. Milesa i A.M. Hubermana sposobów reprezentacji danych: matryca wyjaśniająca skutki, matryca dynamiki przypadku oraz sieć przyczynowa<sup>12</sup>. Techniki te nie tylko służą opisowi danych, lecz także, zgodnie z założeniami autorów, mają ułatwiać jakościową analizę przyczynową, czyli analizę pozwalającą rozpoznawać mechanizmy wpływające na taki, a nie inny przebieg zjawisk. Dzięki niej można wyjaśnić przebieg złożonych procesów; jednocześnie jest ona wrażliwa na kontekst lokalny, daje szansę na uchwycenie procesualności zjawisk, pozwala bowiem uwzględnić czasowy wymiar przebiegu zjawisk, pokazując, że jedno wydarzenie lub stan poprzedza inne. Podejście zaproponowane przez Milesa i Hubermana nie jest oczywiście jedynym możliwym sposobem ułatwienia systematycznej analizy danych jakościowych; podejść takich znamy znacznie więcej. Wymieniłem tutaj chociażby podejście zakładające konstruowanie map pojęciowych, które pozwala odtwarzać procesy kognitywne badanych i wyjaśniać podejmowane przez nich decyzje<sup>13</sup>. W niniejszym tekście decydujemy się jednak na bliższe omówienie kilku metod przedstawionych przez Milesa i Hubermana, gdyż wydają się nam one najbardziej adekwatne do zadań, jakie stoją przed ewaluacją *ex-post*, czyli do opisania i wyjaśnienia przebiegu interwencji, jej rezultatów i wpływu.

Matryca wyjaśniająca skutki jest sposobem takiego uporządkowania danych, aby ich układ pozwalał nam uzyskać odpowiedź na pytanie, jak osiągnięto określone rezultaty i z czego one wynikają. Możemy na przykład analizować krótko- i długoterminowe skutki działań podejmowanych przez różne grupy osób zaangażowanych w dane przedsięwzięcie, np. osób zarządzających projektem, wykonawców tego projektu, pracowników administracji,

<sup>12</sup> Miles M.B., Huberman A.M., *Analiza danych jakościowych*, przekł. S. Zabielski, Trans Humana, Białystok 2000.

<sup>13</sup> Informacje na temat tworzenia map koncepcyjnych można znaleźć m.in. w artykule: Worek B., Perek-Białas J., „Tworzenie map koncepcyjnych. Jakościowa technika rekonstrukcji procesów decyzyjnych”, w: J. Garczarczyk (red.), *Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność*, Wydawnictwo Akademii Ekonomicznej, Poznań 2006.

Tabela 1. Matryca wyjaśniająca skutki: bieżąca pomoc

Lokalizacja	Ocena	Typ pomocy	Skutki użytkowników [stany użytkowników]	Skutki długoterminowe [zdolny/niezdolny do]	Wyjaśnienia badacza
Nadzorca	0	Żaden	„Przychodzili tylko dla pieniędzy”	Tworzenie/poszerzanie programu na własną rękę	Nadzorca rządził pieniędzmi nie związanymi z projektem 1.
Administrator budynku Rok 1–2	+	Kontakty z szerszą społecznością. Popiera, zachęca. „Zderzak” w kontaktach.	„Posiadamy tu pewną pomoc”. „Nie jesteśmy sami”. „Walczymy o swoje sami”.	Zostaje użytkownikiem więcej czasu dla programu, problemów uczniów, miejsca, transportu. Musieli walczyć o swoje sami, nauczyli się jak walczyć.	Administrator wspierał program innowacji, dostarczając konieczną pomoc nauczycielom.
Rok 3–4	0	Żaden	„Straciliśmy przyjaciela”.		Względna obojętność administratora prawdopodobnie umocniła nauczycieli.
Kierownicy programów	++	Podawali pomysły, pomagali. Dawali szkielec, projekt. Planowanie działań. Pozytywny pogład. Zasoby.	„Znają się na tym”. „To wszystko doręczne”. „Jak to może działać?”. „Ha, ja to mogę zrobić”. „Możemy dostać sprzęt”.	Ekspansja programu do zewnętrznych miejsc poza kampusem. Szerze wiązanie społeczności lokalnej.	Wydział żądał więcej pieniędzy, aby prowadzić program: jeden użytkownik wymagał pokierowania i pomocy, otrzymał jedno i drugie, i poszło.
Inni użytkownicy	+	Pomagali w planowaniu. Podawali pomysły. Dawali zachęte.	Informacje o działaniu programu. Wypełnianie luk. „Nie jestem sam: jest pomoc”.	Zdecydowani użytkownicy poza kampusem wiedzą, jak to zrobić.	Jeden użytkownik był bardzo doświadczony w tym podejściu i spowodował przyspieszenie.
Materiały (przewodniki)	+	Podsuwanie szczegółowych działań, nadzór, odpowiednie podejście.	To jest rzeczywiście dobry materiał, dobrze opracowany, wypróbowany, użyteczny.	Dobra postawa do opracowania programu własnego.	Materiały spełniły swoją rolę – pobudzenie i pomoc w rozwoju lokalnych programów.

Źródło: Miles M.B., Huberman A.M., *Analiza...*, op.cit., s. 153.

zewnętrznych konsultantów itp. Taka matryca pozwala zrozumieć czasowy przebieg zdarzeń, ale daje też możliwość wstępnego wglądu w zależności rysujące się pomiędzy rodzajami podejmowanych działań a ich efektami (przykładową matrycę prezentujemy w tab. 1). Jednocześnie, matryca nie pozwala jeszcze objąć całości zachodzących procesów, bowiem trudno tutaj dostrzec związki pomiędzy poszczególnymi typami działań podejmowanych przez różnych aktorów a ich skutkami. Dlatego M.B. Miles i A.M. Huberman zalecają, aby taką matrycę traktować jako wstęp do dalszych analiz, w których można wykorzystać matrycę dynamiki przypadku oraz sieć przyczynową.

Matryca dynamiki przypadku jest kolejnym sposobem uporządkowania i zaprezentowania danych. Pozwala pokazać czynniki generujące zmiany oraz dające się zaobserwować skutki tych zmian. Konstrukcja matrycy dynamiki przypadku polega na przedstawieniu, z jednej strony, „żądań”, „napięć” czy „wymagań”, charakteryzujących daną interwencję czy innowację, z drugiej zaś – na wskazaniu możliwych uwarunkowań tych napięć czy wyzwań, informacji o sposobach organizacji radzenia sobie z nimi oraz o przyjętych rozwiązaniach organizacyjnych. Przykładową matrycę dynamiki przypadku prezentuje tab. 2.

Jeśli dane zostały wcześniej odpowiednio zakodowane, a przyjęty klucz kategoryzacyjny tak skonstruowany, aby możliwa była identyfikacja elementów koniecznych do sporządzenia takiej matrycy, sama jej konstrukcja nie powinna być zbyt czasochłonna i skomplikowana. Ponieważ matryca ma wspomóc proces interpretacji danych i ułatwić formułowanie wniosków, należy ją uzupełnić opisem, w którym będzie zawarte omówienie i wyjaśnienie dających się zaobserwować związków.

**Sieć przyczynowa** to reprezentacja najważniejszych niezależnych i zależnych zmiennych badania terenowego (zawartych w ramkach) i relacji między nimi (pokazanych przy użyciu strzałek)<sup>14</sup>. Układ tych relacji ma pokazywać kierunek oddziaływania, a nie tylko współwystępowanie zmiennych. Mapa powinna dać możliwość zaobserwowania, że pewne czynniki wywierają wpływ na inne: „X przywołuje do zaistnienia Y albo sprawia, że Y staje się większe lub mniejsze”<sup>15</sup>. Przykład sieci prezentuje rys. 1. Aby taka sieć była użyteczna, trzeba do niej dołączyć opis relacji pomiędzy zaprezentowanymi czynnikami.

Sieci przyczynowe można tworzyć na dwa sposoby: wykorzystując podejście indukcyjne albo dedukcyjne. W podejściu indukcyjnym odkrywamy powtarzające się zjawiska w trakcie kontaktu z badanymi i obserwacji zachodzących procesów. Na bieżąco formułujemy hipotezy robocze, które są weryfikowane podczas dalszych prac badawczych. Mapa przyczyn opisująca i wyjaśniająca funkcjonowanie określonego przedsięwzięcia powstaje

<sup>14</sup> Miles M.B., Huberman A.M., *Analiza...*, op.cit., s. 156.

<sup>15</sup> Ibidem.

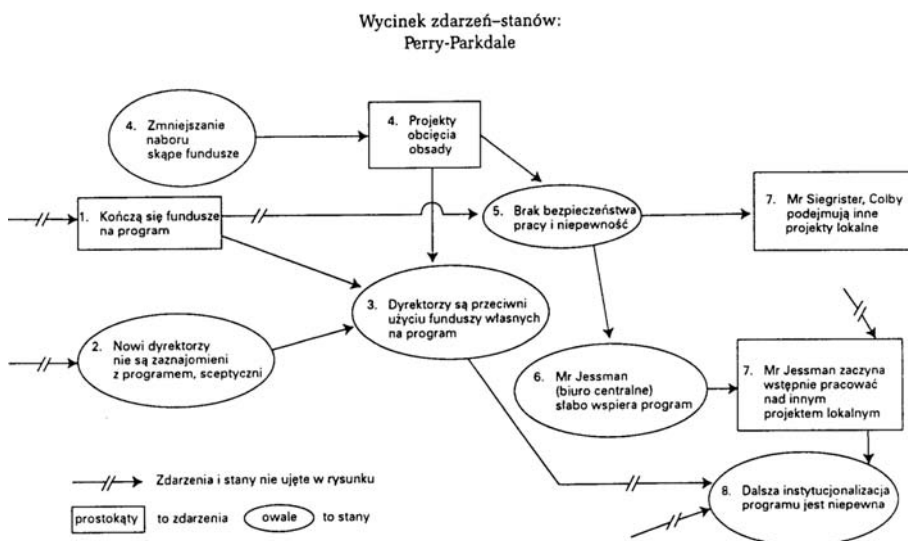
Tabela 2. Matryca dynamiki przypadku. Wdrażanie projektu partnerskiego w ramach inicjatywy EQUAL

Napięcia, trudności	Powody (w opinii badacza)	Jak sobie z tym radzono?	Jak rozwiązano problem? (typ zmian organizacyjnych)
Problemy z przyjmowaniem odpowiedzialności za realizację celów, jakie postawiono przed partnerstwem.	<ul style="list-style-type: none"> <li>- brak znajomości celów projektu i logiki przedsięwzięcia</li> <li>- niejasne zdefiniowanie celu projektu</li> </ul>	<ul style="list-style-type: none"> <li>- organizacja wspólnych spotkań</li> <li>- identyfikacja celów</li> <li>- dyskusja nad celami</li> </ul>	<ul style="list-style-type: none"> <li>- wprowadzono procedurę regularnych spotkań i warsztatów dla partnerów projektu (P) (K)</li> </ul>
Przekonanie, że obowiązujące w partnerstwie procedury monitoringu są niepotrzebnym obciążeniem.	<ul style="list-style-type: none"> <li>- obciążenie pracą</li> <li>- silne przywiązanie do autonomii instytucji</li> <li>- niechęć do kontroli zewnętrznej</li> </ul>	<ul style="list-style-type: none"> <li>- wspomaganie partnerów</li> <li>- usprawnianie procedur</li> </ul>	<ul style="list-style-type: none"> <li>- wielokrotne korekty i upraszczanie procedur</li> <li>- opracowanie zasad sprawozdawczości (P)</li> </ul>
Niesprawna komunikacja, częste opóźnienia w przekazywaniu i odbieraniu informacji.	<ul style="list-style-type: none"> <li>- niejasne zasady komunikacji</li> <li>- różne systemy obowiązujące w różnych instytucjach</li> </ul>	<ul style="list-style-type: none"> <li>- identyfikacja istniejących barier</li> <li>- wspomaganie partnerów przez lidera projektu</li> </ul>	<ul style="list-style-type: none"> <li>- opracowanie i upowszechnienie procedur komunikacji (P)</li> <li>- wiodąca rola lidera w systemie przekazu informacji i monitorowaniu tego systemu (S)</li> </ul>
Wątpliwości co do słuszności kierunków działań zapisanych w projekcie.	<ul style="list-style-type: none"> <li>- autonomia</li> <li>- dążenie do realizacji własnych celów</li> <li>- trudności we wdrożeniu celów</li> </ul>	<ul style="list-style-type: none"> <li>- rozmowy, spotkania</li> <li>- dyskusja o celach</li> <li>- wsparcie</li> </ul>	<ul style="list-style-type: none"> <li>- analiza kontekstu lokalnego i identyfikacja koniecznych modyfikacji działań (P)</li> </ul>
Problemy z rozliczaniem „złe działających” partnerów.	<ul style="list-style-type: none"> <li>- władza</li> <li>- koordynacja</li> <li>- odpowiedzialność</li> </ul>	<ul style="list-style-type: none"> <li>- jasne zdefiniowanie roli lidera i jego kompetencji</li> <li>- określenie sankcji związanych z niewywiązywaniem się z zadań</li> </ul>	<ul style="list-style-type: none"> <li>- instytucjonalizacja działań lidera (S)</li> <li>- opracowanie zasad regulujących współpracę (P)</li> <li>- wzmocnienie poczucia identyfikacji z partnerstwem (K)</li> </ul>
„Zbyt aktywni” partnerzy nie są akceptowani przez pozostałych członków partnerstwa.	<ul style="list-style-type: none"> <li>- obciążenie pracą</li> <li>- problemy związane z zarządzaniem partnerstwem</li> </ul>	<ul style="list-style-type: none"> <li>- łagodzenie konfliktów poprzez nieformalne dyskusje</li> </ul>	<ul style="list-style-type: none"> <li>- częstsze interakcje i wymiana informacji między partnerami (K)</li> <li>- wsparcie słabszych, mających trudności partnerów przez lidera projektu (S)</li> </ul>

(S) Zmiany strukturalne, (P) Zmiany proceduralne, (K) Zmiana klimatu

Źródło: Opracowanie własne.

więc krok po kroku podczas zbierania danych i gromadzenia doświadczeń. Najpierw mamy surowe dane i obserwacje, wskazujące na współwystępowanie pewnych zjawisk i na ich powiązanie z innymi, następnie dane te są kategoryzowane, a badacz przedstawia relacje między nimi w postaci mapy. W podejściu dedukcyjnym natomiast przystępujemy do badań, mając opracowaną siatkę pojęciową oraz określone w ramach hipotez relacje pomiędzy pojęciami. Te hipotezy poddaje się potem weryfikacji w terenie. W praktyce badawczej, a zwłaszcza w badaniach ewaluacyjnych, które mają zazwyczaj dość jasno zarysowane cele, zaleca się stosowanie podejścia dedukcyjnego, pozwala ono bowiem znacznie zaoszczędzić czas i zwiększyć efektywność poszukiwań badawczych. Brak wstępnej konceptualizacji i hipotez roboczych może spowodować, że będziemy tracić czas i pieniądze na zbieranie niepotrzebnych danych.



Rysunek 1. Sieć przyczynowa: relacje zdarzenia–stany

Źródło: Miles M.B., Huberman A.M., *Analiza...*, op.cit., s. 163.

## Ocena trafności sformułowanych wniosków

Niezależnie od tego, jakimi posługujemy się metodami badawczymi, słusznie oczekujemy, że w wyniku przeprowadzonego procesu badawczego i analitycznego sformułujemy trafne wnioski, które – najprościej rzecz ujmując – pozwolą nam uzyskać prawdziwe czy zobiektywizowane odpowiedzi na zadane pytania. Trafność jest istotna w każdym typie badań, zarówno w badaniach naukowych, jak i stosowanych: czy to marketingowych, czy ewaluacyjnych. W tych ostatnich nabiera jednak szczególnego znaczenia, gdyż od

jakości sformułowanych wniosków zależy jakość podejmowanych decyzji, a decyzje te niejednokrotnie dotyczą wielomilionowych inwestycji o charakterze komercyjnym lub publicznym. Kwestia trafności w takiej sytuacji przestaje być tylko problemem teoretycznym czy metodologicznym, stając się wielkim wyzwaniem praktycznym; chodzi bowiem o zagwarantowanie, iż wnioski i zalecenia, które sformułujemy, będą oparte na pełnej i rzetelnej ocenie dostępnych informacji.

Tradycyjnie uważa się, że badania jakościowe cechują się dużą trafnością, ponieważ dane, które zbieramy, są słabiej zdeterminowane charakterem zastosowanego narzędzia, rodzajem zadanych pytań, w większym stopniu natomiast umożliwiają badanym wypowiedzenie się i przedstawienie ich sposobu rozumienia i postrzegania rzeczywistości. Niesłusznym byłoby jednak stwierdzenie, że sam charakter zastosowanych metod gwarantuje trafność sformułowanych wniosków. Problem ten doskonale rozumieją badacze jakościowi, którzy przeprowadzali badania w jakiejś społeczności czy instytucji. Doświadczyli oni bowiem tego, że zanim udało im się dotrzeć do faktycznego stanu rzeczy, musieli zapoznać się z wieloma punktami widzenia, poznawali wiele różniących się od siebie znacząco interpretacji, z których każda była przedstawiana jako „prawdziwa” czy „uzasadniona”. Kluczowym więc pozostaje pytanie, jak zagwarantować obiektywność wniosków formułowanych w oparciu o bardzo subiektywne dane.

W badaniach jakościowych wypracowano liczne procedury pozwalające rozwiązać ten problem, a w każdym razie znacząco przybliżające nas do jego rozwiązania. Wiele z nich opisują (w cytowanej już tutaj wielokrotnie książce) M.B. Miles i A.M. Huberman, jak też inni badacze jakościowi, na przykład D. Silverman czy J. Kirk i M. Miller<sup>16</sup>. Istotne znaczenie mają także – opisane w innym rozdziale niniejszej publikacji – procedury triangulacji; i to zarówno triangulacji metod, jak i triangulacji danych czy informatorów. Nie sposób opisać tutaj wszystkich takich działań (czy raczej procedur), warto jednak wspomnieć o tych, które wydają się najciekawsze z punktu widzenia badań ewaluacyjnych. Należy do nich weryfikacja wniosków, które sformułował badacz, przez samych badanych lub ich odpowiednio dobranych przedstawicieli i poproszenie ich o skorygowanie tych wniosków czy rekomendacji. M.B. Miles i A.M. Huberman podejście to nazywają „**uzyskiwaniem informacji zwrotnych od informatorów**”<sup>17</sup>. Takie podejście nie tylko pozwala na ocenę sformułowanych wniosków, lecz także umożliwia badanym stanie się współautorami wniosków, włączając ich w proces uzgadniania dalszych

<sup>16</sup> Miles M.B., Huberman A.M., *Analiza...*, op.cit.; Silverman D., *Interpretacja danych jakościowych*, przekł. M. Głowacka-Grajper, J. Ostrowska, Wydawnictwo Naukowe PWN, Warszawa 2007; Kirk J., Miller M.L., *Reliability and Validity in Qualitative Research*, Sage Publications, Beverly Hills 1986.

<sup>17</sup> Miles M.B., Huberman A.M., *Analiza...*, op.cit., s.166. Podobne podejście zalecają także np. L. Spencer, J. Ritchie, J. Lewis i L. Dillon w publikacji *Quality...*, op.cit.

działań czy podejmowania decyzji. Osoby oceniające wnioski powinny być odpowiednio dobrane, aby skład grupy oceniającej odpowiadał zróżnicowaniu podmiotów zaangażowanych w projekt czy inicjatywę.

Pozytywna weryfikacja wniosków przez badanych nie gwarantuje jednak ich trafności; może się bowiem okazać, że zarówno badacz, jak i badani oceniający wnioski są w swoich ocenach tendencyjni albo nie mają wystarczającej wiedzy, pozwalającej im właściwie ocenić daną sytuację. Ponadto oceniający (badani, badacze i odbiorcy wniosków, czyli decydenci) nie zawsze zgadzają się ze sobą w ocenie wniosków. Spośród wielu możliwych przyczyn wymieńmy tu, wspomniane już wcześniej, bazowanie na niekompletnych informacjach, ale też czynniki odnoszące się do specyficznej sytuacji oceniających: np. urzędnicy protestują przeciwko krytycznej ocenie instytucji, w której są zatrudnieni, gdyż taka ocena odnosiłaby się także pośrednio do ich pracy, a mieszkańcy danej miejscowości nie zgadzają się ze stwierdzeniem, że społeczność tę cechuje bierność i roszczeniowość.

Jak piszą M.B. Miles i A.M. Huberman, najlepszym sposobem na sprawdzenie trafności sformułowanych wniosków jest wykorzystanie tych wniosków do sporządzenia predykcji, następnie poddanie tej predykcji pod ocenę dobrze znających problem ekspertów, ostatecznie zaś porównanie tej oceny z faktycznym kierunkiem rozwoju wydarzeń. I tak na przykład, jeśli poprawnie sporządziliśmy analizę jakościową danego przedsięwzięcia w trakcie jego trwania, powinniśmy móc przewidzieć, w jakim kierunku to przedsięwzięcie będzie się rozwijało. Jeżeli należycie zanalizowaliśmy rezultaty programu realizowanego w danej społeczności, powinniśmy być w stanie określić, czy istnieje szansa na kontynuację działań z tego zakresu. Jednakże, jak piszą autorzy, takich predykcji dokonuje się rzadko, wymagają one bowiem bardzo szczegółowej analizy, precyzyjnego określenia zmiennych, powiązań między nimi oraz zdecydowanej i rzetelnej oceny badacza. Tymczasem, badacze jakościowi (i nie tylko oni) zamiast identyfikować związki przyczynowo-skutkowe między zmiennymi i objaśniać, że przy podejmowaniu określonych działań sytuacja w przyszłości będzie wyglądała tak, a nie inaczej, wolą używać bezpiecznej formuły, iż dane „wskazują to i to” czy „sugerują to i tamto”, albo, jak dosadnie piszą M.B. Miles i A.M. Huberman, uciekają się do „tchórzliwego stylu”, pisząc, iż „nie jest rzeczą nierozsądną, aby przypuścić, że...”<sup>18</sup>. Oczywiście takie wnioski nie zadowolają odbiorców badań ewaluacyjnych.

Przedstawione tutaj sposoby oceny trafności wniosków formułowanych na podstawie danych jakościowych są zaledwie wybranymi przykładami podejścia do tego złożonego i często wśród badaczy dyskutowanego zagadnienia. Literatura metodologiczna zawiera wiele innych, równie cennych propozycji i wskazówek, których nie sposób tutaj należycie omówić<sup>19</sup>. Najważniejsze

<sup>18</sup> Miles M.B., Huberman A.M., *Analiza...*, op.cit., s. 170.

<sup>19</sup> Por. np. Silverman D., *Interpretacja...*, op.cit., s. 252–269; Spencer L., Ritchie J., Lewis J., Dillon L., *Quality...*, op.cit., s. 70–82.



wszakże pozostaje to, że wnioski, jakie formułujemy na podstawie danych jakościowych, powinny opierać się na danych poprawnie zbieranych, a następnie poddanych rygorystycznej ocenie i interpretacji. Na podstawie tak opracowanych wniosków można formułować jasne rekomendacje.

## Ewaluacja oparta na teorii programu

Jak wskazywaliśmy na początku, każdy ewaluator, podejmując się oceny czy to małego programu, czy dużej interwencji publicznej, musi zawsze wyjść od indywidualnej hipotezy zakładającej, że w danych warunkach podjęcie działań wynikających z projektu interwencji doprowadzi do osiągnięcia założonych skutków. Interesującym przykładem takiej perspektywy badawczej, opartej głównie na wykorzystaniu metod jakościowych w ocenie programów, jest **ewaluacja oparta na teorii programu**. Podkreśla ona, że zanim przystąpimy do zadawania pytań ewaluacyjnych, powinniśmy znać teorię programu, czyli wyobrażenie na temat tego, jak program może (bądź powinien) działać, lub – jak wskazuje H.T. Chen – „dookreślenie, co musi być zrobione, aby osiągnąć pożądane efekty, jakich innych efektów programu można oczekiwać i w jaki sposób te cele i efekty zostaną wygenerowane”<sup>20</sup>.

Jak twierdzi F.L. Leeuw<sup>21</sup>, użyciu tej perspektywy sprzyjają trzy czynniki: duża ilość zasobów zaangażowanych w program, istotność i siła wpływu programu oraz niepewność co do wpływu programu i jego efektów ubocznych. Zastosowanie omawianego podejścia jest tym bardziej wskazane, im bardziej interesuje nas, oprócz odpowiedzi na pytanie: „czy program działa?”, odtworzenie kontekstu jego działania zarówno w otoczeniu wewnętrznym, jak i zewnętrznym, oraz dowiedzenie się, w jaki sposób program prowadzi do zamierzonych i niezamierzonych efektów.

Fundamentalny model ewaluacji oparty na teorii programu składa się z trzech podstawowych etapów: 1) wygenerowanie teorii programu, 2) sformułowanie i uhierarchizowanie pytań badania ewaluacyjnego, 3) odpowiedź na pytania ewaluacyjne<sup>22</sup>. Jak widać, różni się on od standardowej procedury

---

<sup>20</sup> Chen H.T., *Theory-Driven Evaluations*, Sage Publications, Newbury Park 1990, s. 43. Warto tu zwrócić uwagę na kilka rozróżnień dotyczących teorii programu. Po pierwsze, należy odróżnić teorię programu od teorii wdrażania samego programu. W sytuacji, gdy program nie przyniósł pożądanych skutków, kluczowa jest odpowiedź na pytanie, gdzie leżał błąd. W samej teorii programu czy też w teorii jego wdrażania? Badacze zwracają uwagę na rozróżnienie oficjalnej i faktycznej teorii programu. Oficjalna, zapisana w dokumentach programowych teoria jest dostosowana do zewnętrznych wymogów programu, np. może być tworzona z myślą o potencjalnych kontrolach. Natomiast faktyczna teoria programu jest tym konstruktem myślowym, który w rzeczywistości kieruje działaniami osób odpowiedzialnych za realizację programu. Dużym wyzwaniem dla ewaluatorów może być więc dotarcie do faktycznej teorii programu.

<sup>21</sup> Leeuw F.L., „Reconstructing program theories: methods available and problems to be solved”, *American Journal of Evaluation* 2003, vol. 24, nr 1, s. 5–20.

<sup>22</sup> Donaldson S.I., *Program Theory-Driven Evaluation Science. Strategies and Applications*, Lawrence Erlbaum Associates, New York 2007.

ewaluacji *ex-post* tym, że generuje się teorię programu (a raczej odkrywa się ją czy wypracowuje) jeszcze zanim zostaną sformułowane pytania badawcze.

J. Davidson<sup>23</sup> rozwija ten model proponując, by najpierw ocenić potrzeby informacyjne odbiorców ewaluacji, następnie, poprzez analizę danych zastanych i wywiady z interesariuszami, zidentyfikować wszelkie zmiany, które może wywołać program, i aby poszukać wszystkich możliwych przyczyn tych zmian. Do tak zbudowanego wstępnego modelu działania programu należy, zdaniem autorki, dodać wszystkie omówione w literaturze przedmiotu i pojawiające się w programach zależności, a następnie przeprowadzić test tak zbudowanej teorii<sup>24</sup>. S. Sidani<sup>25</sup> idzie jeszcze dalej, rozbudowując model do całego zestawu kluczowych pytań (zaprezentowanych na poniższym schemacie), na które należy odpowiedzieć, by wygenerować kompletną teorię programu<sup>26</sup>.

W każdym z przytoczonych powyżej wariantów generowania teorii programu jako podstawowe techniki badawcze wymienia się analizę oficjalnych dokumentów programowych oraz wywiady (indywidualne lub grupowe) z interesariuszami i ekspertami. Bazowanie na paradygmacie idiograficznym w ewaluacji opartej na teorii programu wiąże się niewątpliwie z paralelizmem cech charakteryzujących obydwa podejścia, np.<sup>27</sup>:

- konstruktywizm – punktem wyjścia w badaniu jest sposób postrzegania samego programu przez najważniejszych interesariuszy,
- podmiotowość beneficjentów – pod uwagę bierze się proces decyzyjny beneficjentów,
- kontekstowość – program nie może być rozpatrywany w oderwaniu od swojego kontekstu,
- przyczynowość – jednym z celów ewaluacji jest zrozumienie, w jaki sposób podjęte działania doprowadziły do obserwowanych efektów,
- intersubiektywizm – w ramach badania wykracza się poza oficjalne teorie i cele programu poprzez czerpanie informacji na temat programu z różnych źródeł.

---

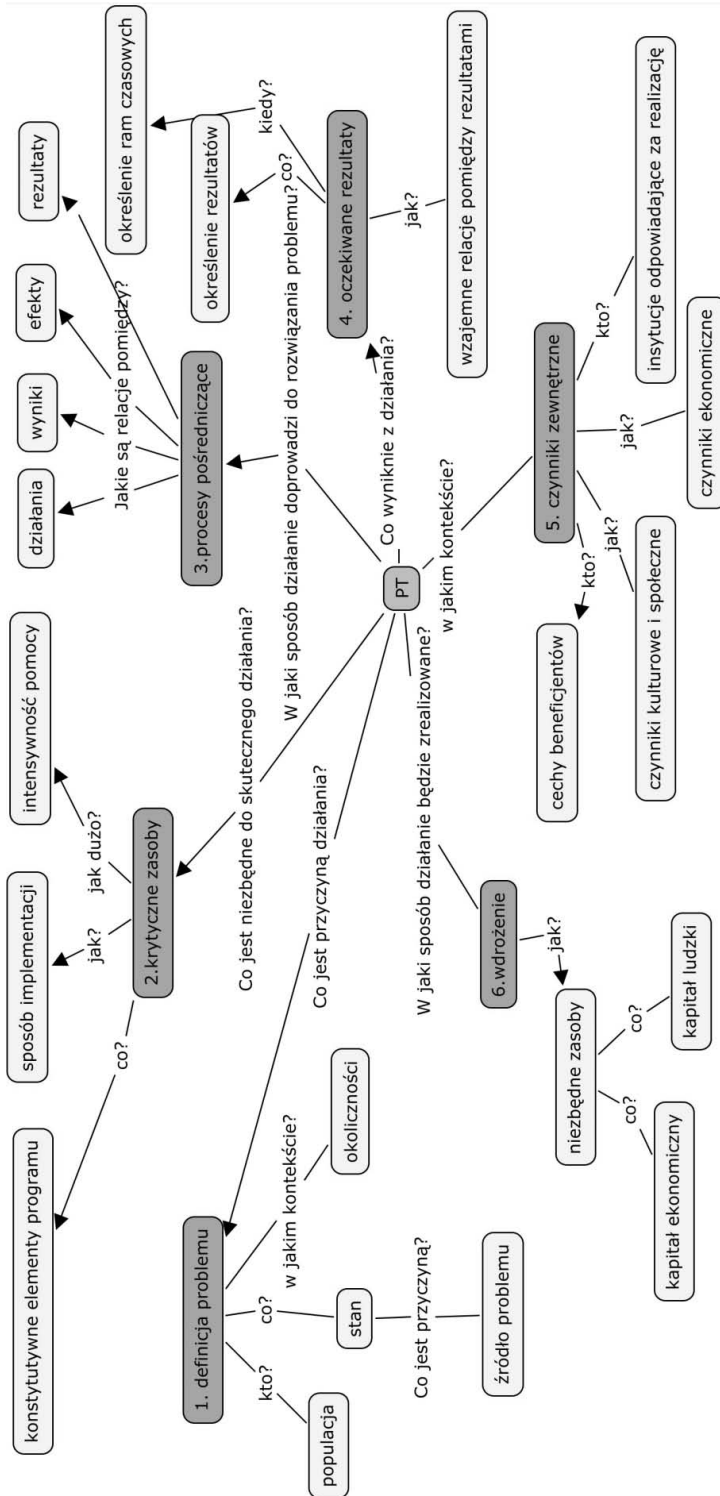
<sup>23</sup> Davidson J., „Ascertaining causality in theory-based evaluation”, *New Directions for Evaluation* 2000, nr 87, s. 17–26.

<sup>24</sup> Najlepiej byłoby, gdyby test został przeprowadzony na podstawie innych danych niż te, na których budowano model, np. danych ilościowych (zastanych bądź uzyskanych w trakcie badania terenowego), dokumentów sprawozdawczych, wcześniejszych wyników z badań ewaluacyjnych w obszarze danej interwencji bądź interwencji o podobnym charakterze itp.

<sup>25</sup> Sidani S., Sechrest L., „Putting program theory into operation”, w: *American Journal of Evaluation* 1999, vol. 20, nr 2, s. 227–238.

<sup>26</sup> Płodnym poznawczo uzupełnieniem teorii programu jest zestawienie go z „antyteorią”, tzn. opiniami krytyków danego programu; por. Carvalho S., White H., „Theory-based evaluation: the case of social funds”, *American Journal of Evaluation* 2004, vol. 25, nr 2, s. 141–160.

<sup>27</sup> Por. Stame N., „Theory-based evaluation and types of complexity”, *Evaluation* 2004, nr 10(1), s. 58–76; a także przedstawione we wcześniejszej części artykułu warunki użyteczności badań jakościowych w ewaluacji.



Rysunek 2. Elementy teorii programu

Źródło: Opracowanie własne na podstawie: Sidani S., Sechrest L., „Putting program theory...”, op.cit.

Przede wszystkim jednak umożliwia realizację kluczowych celów ewaluacji, jakimi są odtworzenie zachodzących zjawisk oraz wyjaśnienie genetyczne i przyczynowe ich pojawienia się i funkcjonowania. Jak wskazują R. Pawson i N. Tilley<sup>28</sup>, badania ewaluacyjne powinny odnajdywać i weryfikować przypisane do programów modele składające się ze szczegółowo opisanych podmiotów działania, mechanizmów zmiany, kontekstu w jakim działanie ma miejsce i efektów działania. Zdaniem autorów, programów nie można dzielić na skuteczne i nieskuteczne. Większość programów jest bowiem skuteczna dla pewnej grupy podmiotów i w określonym kontekście.

Ewaluacja oparta na teorii programu jest zapewne narzędziem, które warto rozważyć w trakcie przeprowadzania ewaluacji *ex-post*, jednak na koniec tego skrótowego opisu podejścia warto zestawić ze sobą koszty i korzyści jej zastosowania. Obok niewątpliwych zalet (lepsze zrozumienie mechanizmu działania programu, kumulacja wiedzy o podobnych programach<sup>29</sup>, a wreszcie dokładniejsza ocena ewaluowanego programu) badacz musi uwzględnić konieczność wykorzystania w większej mierze dwóch najważniejszych zasobów badania ewaluacyjnego: czasu (badaczy i interesariuszy) oraz pieniędzy.

## Zakończenie

Polityki publiczne powstają w wyniku splotu inicjatywy i twórczych idei polityków ze zracjonalizowanym, opartym o wiedzę ekspercką działaniem administracji i ograniczeniami narzucanymi przez warunki działania. Cele polityk publicznych nie wynikają automatycznie z diagnoz prowadzonych przez surowych, obiektywnych analityków. Są one wynikiem wizji politycznych, ideologii rozwojowych, gry interesów i konkurencji politycznej. Ekspercka diagnoza społeczna i analiza polityk publicznych powinna sprawdzać prawidłowość diagnoz i definicji sytuacji przyjmowanych przez polityków, ustalać zależności przyczynowe, bilansować zasoby niezbędne do działania, tworzyć scenariusze wynikające z proponowanych działań. Stanowi ona jednak zawsze tylko wsparcie dla inicjatywy polityków. Opieranie się na zweryfikowanych pozytywnie teoriach programów podnosi prawdopodobieństwo sukcesu, choć ze względu na zmienność warunków działania nie gwarantuje idealnego powtórzenia efektów. W zakresie „odkrywania” teorii programu metody jakościowe wydają się być narzędziem co najmniej przydatnym, a nawet pierwszorzędym.

Ewaluacja ma dużo wyższy potencjał falsyfikacyjny (odrzućcia interwencji nieskutecznych, a często wręcz przeciwnie skutecznych) niż konfirmacyjny (potwierdzania, że osiągnięte efekty są wynikiem interwencji, a nie czynników od niej niezależnych). Wywołuje to pewną nieufność decydentów wobec

<sup>28</sup> Por. Pawson R., Tilley N., *Realistic Evaluation*, Sage Publications, London 1997.

<sup>29</sup> Por. ibidem.

ewaluacji; wolą oni analizę polityk publicznych, która jest prospektywna (podczas gdy ewaluacja w ścisłym tego słowa znaczeniu jest retrospektywna), oparta na koncepcjach teoretycznych i wiedzy eksperckiej oraz nie tak bardzo związana procedurami metodologicznymi. Także ewaluatorzy nie chcą pozostawać w roli krytycznych falsyfikatorów i rozwijają koncepcje ewaluacji bliższej uczestnikom procesu wdrażania rozwiązań, która staje się faktycznie moderowaną przez ewaluatorów autorefleksją uczestników interwencji i jej beneficjentów nad ich rolą w realizowanym programie i prowadzi do ciągłego ulepszania prowadzonych działań. Takie koncepcje pojawiają się w ramach tzw. nurtu konstruktywistycznego czy też ewaluacji partycypacyjnej i zwane są też ewaluacją czwartej generacji<sup>30</sup>.

Nie negując samej możliwości uwzględnienia partycypacji wszystkich zainteresowanych stron w procesie ewaluacji i pragmatycznej wartości takiego podejścia, trzeba stwierdzić, że ewaluacja nie może odchodzić od swej istoty, czyli oceny efektów (potencjalnych i ostatecznych) interwencji. W tym artykule staraliśmy się pokazać, jakie warunki powinny spełniać metody jakościowe, by sprostać wymaganiom wynikającym ze wszystkich funkcji ewaluacji. Dlatego nacisk został położony na dwie kwestie: po pierwsze, wiarygodność i obiektywizm badań jakościowych wyrażający się w dążeniu do ich specyficznej rzetelności i trafności (w tym zwłaszcza prognostycznej); po drugie, znaczenie badań jakościowych w rekonstrukcji teorii programu. Kwestie te są bardzo ważne w realizacji funkcji pomiaru efektów programu (rozliczeniowej) i funkcji poznawczej (gromadzenia wiedzy o mechanizmach programów i ich implementacji oraz wartości). Wykorzystanie badań jakościowych w realizacji funkcji rozwojowej (formatywnej) jest dość oczywiste i eksponowane, niekiedy nawet jako pierwszorzędne dla ewaluacji.

Badania jakościowe zbyt często traktuje się jako ekspresję anarchizmu metodologicznego. Mamy nadzieję, że ten artykuł dowodzi tezy przeciwnej: jeśli badania jakościowe mają stanowić narzędzie podnoszenia skuteczności i jakości programów, jeśli mają uzyskać akceptację ze strony decydentów, muszą być oparte o metodologię, która nie budzi wątpliwości co do obiektywizmu i trafności wyników. Badania jakościowe stanowią oczywiste wsparcie w procesie budowania koncepcji ewaluacji, definiowania wskaźników, narzędzi badawczych i interpretacji wyników. Tutaj staraliśmy się podkreślić jednak ich samodzielną rolę, dopełniającą ustalanie wartości programu w procesie triangulacji metod i danych.

---

<sup>30</sup> Szerszy krytyczny komentarz na temat tego podejścia zawarty jest w artykule: Górniak J., „Ewaluacja w cyklu polityk publicznych”, w: *Ewaluacja funduszy strukturalnych. Perspektywa regionalna*, MSAP, Uniwersytet Ekonomiczny w Krakowie, Kraków 2007.

**Jarosław Górniak** – doktor habilitowany, profesor Uniwersytetu Jagiellońskiego, kierownik Zakładu Socjologii Gospodarki i Edukacji Instytutu Socjologii UJ, prodziekan Wydziału Filozoficznego UJ, profesor w Wyższej Szkole Europejskiej im. Ks. J. Tischnera w Krakowie. Socjolog i ekonomista, specjalista w zakresie socjologii gospodarki i organizacji, metod badań i analizy danych, metodologii ewaluacji i analizy polityk publicznych. Ekspert Zespołu koordynacyjnego ds. ewaluacji *ex-ante* dokumentów unijnych na lata 2007–2013 w Ministerstwie Rozwoju Regionalnego. Konsultant w zakresie programowania i ewaluacji, a także w zakresie systemów analizy danych w zarządzaniu i marketingu.

**Barbara Worek** – doktor nauk społecznych, adiunkt w Zakładzie Socjologii Gospodarki i Edukacji Instytutu Socjologii UJ i w Wyższej Szkole Europejskiej im. Ks. J. Tischnera w Krakowie. Studiowała socjologię na Uniwersytecie Jagiellońskim i teologię na Papieskiej Akademii Teologicznej w Krakowie. Przez wiele lat pracowała jako specjalista ds. badań jakościowych w Instytucie Badań Rynku i Opinii Publicznej CEM, zajmując się moderowaniem grup fokusowych, analizą danych i sporządzaniem raportów z badań. Doświadczenia badawcze zdobywała uczestnicząc w polskich i międzynarodowych projektach badawczych dotyczących partycypacji obywatelskiej, polityki publicznej, zagadnień wykluczenia społecznego, rynku pracy i kształcenia ustawicznego. Autorka i współautorka licznych raportów badawczych, w tym raportu poświęconego ewaluacji wdrażania Działań 1.2 i 1.3 SPO RZL. Prowadzi zajęcia z metodologii badań oraz proseminaria i seminaRIA magisterskie poświęcone zastosowaniu metod jakościowych w diagnozie społecznej i ewaluacji.

**Seweryn Krupnik** – absolwent socjologii Uniwersytetu Jagiellońskiego, doktorant w Zakładzie Socjologii Gospodarki i Edukacji UJ, wykładowca w Wyższej Szkole Europejskiej im. Ks. J. Tischnera, stypendysta Antioch College i The Ryoichi Sasakawa Young Leaders Fellowship Fund, w latach 2004–2005 studiował na Uniwersytecie Johanna Wolfganga Goethego we Frankfurcie. Prowadzi zajęcia z obszaru badań społecznych. Brał udział w trzech edycjach międzynarodowego badania studentów ESN Survey jako konsultant metodologiczny. Obecnie koordynuje badania przedsiębiorców w ramach projektu diagnozującego stan kształcenia ustawicznego w Małopolsce.

## Bibliografia

- Babbie E., *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2003.
- Black N., „Why we need qualitative research?”, *Journal of Epidemiology and Community Health* 1994, nr 48, s. 425–426.
- Brickmayer J.D., Weiss C.H., „Theory-based evaluation in practice. What do we learn?”, *Evaluation Review* 2000, vol. 24, nr 4, s. 407–431.
- Carvalho S., White H., „Theory-based evaluation: the case of social funds”, *American Journal of Evaluation* 2004, vol. 25, nr 2, s.141–160.
- Chelimsky E., „The coming transformations in evaluation”, w: E. Chelimsky, W.R. Shadish (red.), *Evaluation for the 21st Century*, Sage Publications, London 1997.

- Chen H.T., *Theory-Driven Evaluations*, Sage Publications, Newbury Park 1990.
- Davidson J., „Ascertaining causality in theory-based evaluation”, *New Directions for Evaluation* 2000, nr 87, s. 17–26.
- Davies P., „Contributions from qualitative research”, w: H.T.O. Davies, S. Nutley, P.C. Smith (red.), *What Works? Evidence-based Policy and Practice in Public Services*, Policy Press, Bristol 2000.
- Donaldson S.I., *Program Theory-Driven Evaluation Science. Strategies and Applications*, Lawrence Erlbaum Associates, New York 2007.
- Ekiert K., *Ewaluacja w administracji publicznej*, Rządowe Centrum Studiów Strategicznych, Warszawa 2004.
- Finch J., *Research and Policy: the Uses of Qualitative Methods in Social and Educational Research*, Falmer Press, London 1986.
- Górniak J., „Ewaluacja w cyklu polityk publicznych”, w: *Ewaluacja funduszy strukturalnych. Perspektywa regionalna*, MSAP, Uniwersytet Ekonomiczny w Krakowie, Kraków 2007.
- Kirk J., Miller M.L., *Reliability and Validity in Qualitative Research*, Sage Publications, Beverly Hills 1986.
- Konecki K., *Studia z metodologii badań jakościowych. Teoria ugruntowana*, Wydawnictwo Naukowe PWN, Warszawa 2000.
- Leeuw F.L., „Reconstructing program theories: methods available and problems to be solved”, *American Journal of Evaluation* 2003, vol. 24, nr 1, s. 5–20.
- Maison D., *Zogniskowane wywiady grupowe. Jakościowa metoda badań marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 2001.
- Miles M.B., Huberman A.M., *Analiza danych jakościowych*, przekł. S. Zabielski, Trans Humana, Białystok 2000.
- Patton M.Q., *Qualitative Research and Evaluation Methods*, Sage Publications, Thousand Oaks 2002.
- Patton M.Q., *Qualitative Evaluation and Research Methods*, Sage Publications, Beverly Hills 1990.
- Pawson R., Tilley N., *Realistic Evaluation*, Sage Publications, London 1997.
- Popay J., Williams G., „Qualitative research and evidence-based healthcare”, *Journal of the Royal Society of Medicine* 1998, Supplement, vol. 191, nr 35, s. 32–37.
- Rist R.C., „Influencing the policy process with qualitative research”, w: N.K. Denzin, Y.S. Lincoln (red.), *Handbook of Qualitative Research*, Sage Publications, Thousand Oaks 2000.
- Sidani S., Sechrest L., „Putting program theory into operation”, w: *American Journal of Evaluation* 1999, vol. 20, nr 2, s. 227–238.
- Silverman D., *Interpretacja danych jakościowych*, przekł. M. Głowacka-Grajper, J. Ostrowska, Wydawnictwo Naukowe PWN, Warszawa 2007.
- Spencer L., Ritchie J., Lewis J., Dillon L., *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*, National Centre for Social Research 2003, s. 70–82.
- Stame N., „Theory-based evaluation and types of complexity”, *Evaluation* 2004, nr 10(1), s. 58–76.
- Sztumski J., *Wstęp do metod i technik badań społecznych*, Wydawnictwo Śląsk, Katowice 2005.
- Tavistock Institute, *The Evaluation of Socio-Economic Development. The Guide*, Tavistock Institute 2003.
- Williams D.D., „When is naturalistic evaluation appropriate?”, w: D.D. Williams (red.), *Naturalistic Evaluation, New Directions for Program Evaluation No. 30*, Jossey-Bass, San Francisco 1986.
- Worek B., Perek-Białas J., „Tworzenie map koncepcyjnych. Jakościowa technika rekonstrukcji procesów decyzyjnych”, w: J. Garczarczyk (red.), *Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność*, Wydawnictwo Akademii Ekonomicznej, Poznań 2006.





*Agnieszka Haber, Maciej D. Kryszczuk*

## **Wybrane problemy doboru technik badawczych w ewaluacji *ex-post***

Zagadnienie doboru technik badawczych, ich sprawnej integracji oraz organizacji jest – po ustaleniu celów i kryteriów – niewrażliwym punktem każdego, rzetelnie projektowanego procesu ewaluacyjnego. Pomimo zbieżności metodologicznej, w wielu istotnych obszarach schematy procesów badawczych, które znajdują zastosowanie w praktyce ewaluacyjnej, odbiegają od modeli stosowanych w badaniach rynkowych oraz akademickich. Poza różnicami fundamentalnymi, związanymi z innymi funkcjami, na sposoby realizacji badania mają wpływ takie czynniki, jak chociażby: cel, przedmiot, kategoria respondenta, czas i tempo realizacji, aktorzy zaangażowani w proces badawczy, środki finansowe, zakres i złożoność analiz, etyka postępowania i wreszcie użyteczność i odpowiedzialność publiczno-prawna. Odmienne kontekst instytucjonalny badań ewaluacyjnych tworzy ramy, w których konstytuują się rozmaite praktyki i wzory postępowania, specyficzne dla różnych środowisk zajmujących się realizacją badań empirycznych. Przy doborze właściwych technik, kluczowe jest więc określenie metodologii ewaluacji oraz poznanie unikalnych i charakterystycznych cech kontekstu instytucjonalnego, w którym dokonuje się oceny interwencji publicznej<sup>1</sup>.

### **Specyfika badań ewaluacyjnych**

Podobnie jak akademickie i rynkowe ośrodki realizacji badań, tak instytucje prowadzące studia ewaluacyjne mają swoją historię i specyfikę, która wytycza ramy tej dziedziny aktywności społecznej. Podstawową funkcją ewaluacji *ex-post* jest całościowa bądź częściowa ocena realizacji, zakresu oddziaływania oraz trwałości danej interwencji publicznej. Wielość funkcji ewaluacji powoduje, iż – w odróżnieniu od badacza komercyjnego czy akademika – ewaluator przyjmuje bardzo wiele ról. Ewaluator jako konsultant

<sup>1</sup> Model procesu badawczego to powtarzalny i dynamiczny wzór postępowania, w którym wyróżnić można różne etapy, tj. sposoby finansowania, planowania, strukturalizacji, realizacji, interpretacji czy publikacji wyników. Dobór właściwych technik, ich ewentualna triangulacja, komplementarność oraz kontekst instytucjonalny określają charakterystyczne właściwości poszczególnych modeli stosowanych w badaniach rynkowych, ewaluacyjnych i akademickich.

ogranicza się do realizacji wytycznych kontraktowych, natomiast w roli animatora–doradcy prowadzi głównie badania interaktywne, na bieżąco wspierając użytkowników w ich technicznych działaniach i organizując pole dialogu. Ewaluator w roli edukatora posiada również duży autorytet naukowy i wiedzę szerszą od operatorów danej interwencji. Wreszcie ewaluator–naukowiec unika interakcji, a zlecających i innych interesariuszy traktuje wyłącznie jako przedmiot badania. Taki ewaluator bada i ocenia program *ex cathedra* – z perspektywy naukowej, bazując wyłącznie na „obiektywnych” danych. W tej roli ewaluator ma tendencje do odnoszenia interwencji – głównie jej efektów – do szerszych teorii i procesów<sup>2</sup>.

Pierwsze pokolenie ewaluatorów było mocno osadzone w tradycji zdominowanej przez dążenie do naśladowania nauk przyrodniczych, co oznaczało ukierunkowanie studiów ewaluacyjnych na analizy statystyczne, testowanie hipotez czy konstruowanie teorii. Postępujący od końca lat sześćdziesiątych coraz wyraźniejszy rozdział badań ewaluacyjnych oraz akademickich, nastawionych z definicji na wiedzę formalną (tzw. badania podstawowe), spowodował zwrot ewaluacji w kierunku logiki odkrycia, wolnej od sztywnych reguł właściwych logice prezentacji. Jak pisze MacDonald, „[...] Teoria ewaluacji w ciągu ostatniego ćwierćwiecza kładła coraz większy nacisk na logikę odkrycia, można twierdzić, że dokonywało się to kosztem respektu dla logiki prezentacji. W obszarze studiów oceniających rozróżnienie proponowane przez Nisbeta znalazło wyraz w podziale na informacje umożliwiające rozumienie (*vindicative*) i dane o charakterze wskaźnikowym (*indicative*), przy czym te pierwsze dotyczą zmian już zrealizowanych, drugie zaś mają służyć jako materiał wspierający podejmowaniu decyzji”<sup>3</sup>.

Zmianom w rozumieniu funkcji studiów ewaluacyjnych towarzyszył rozwój nowych koncepcji badawczych – swoisty metodologiczny zwrot, który można by określić mianem nowego ujęcia relacji pomiędzy paradygmatem ilościowym i jakościowym. Stake, postulując nastawienie na konkretne projekty w praktyce ewaluacyjnej, zauważył, iż podejście jakościowe może równie dobrze służyć uzyskaniu cennych generalizacji: „[...] Pełne studium oceniające przynosi wynik w postaci «opowieści», czasem wspartej danymi o charakterze statystycznym. Opowiada ono o tym, co się wydarzyło. Ujawnia punkty widzenia i przekonania rozmaitych grup i osób w sposób, mam nadzieję, obiektywny. Opowiada o zaletach i brakach. Dodatkową korzyścią, jaką przynosi pełne studium ewaluacyjne, jest to, że jego wyniki poddają się uogólnieniu (Z opowieści wynika...) mogą służyć usprawnieniu następnych programów [...]”<sup>4</sup>.

<sup>2</sup> Por. artykuł K. Olejniczaka w niniejszej publikacji „Teoretyczne podstawy ewaluacji *ex-post*” (role ewaluatorów).

<sup>3</sup> MacDonald B., „Mandarins and lemons – the executive investment in program evaluation”, *Phi Delta Kappa CEDR Quarterly* 1982, nr 2, s. 14.

<sup>4</sup> Stake R.E., „Toward a technology for the educational evaluation”, *Teacher College Record* 1967, t. 68, s. 527.

Ukierunkowanie studiów ewaluacyjnych na praktykę, bez ambicji wnożenia wkładu w ogólne teorie zmian społecznych, należy powiązać ze zwróceniem większej uwagi na określone miejsce i czas interwencji, co skutkuje zdobywaniem tzw. „wiedzy osobistej” (*tacit knowledge*), która pozwala instytucjom prowadzącym badania ewaluacyjne na szybszą naukę oraz skuteczniejsze reagowanie na zmiany określonego otoczenia. Jeszcze dobitniej o zwrocie metodologicznym w ewaluacji pisze Cronbach: „Powinniśmy zastanowić się nad zmianą naszych priorytetów badawczych, które dotychczas polegały na poszukiwaniu uogólnień. Sugerowałbym odwrócenie naszych praktyk. Obserwator zbierający dane w konkretnej sytuacji, dostrzegając określone reakcje w specyficznym kontekście, jest w stanie sformułować ocenę lub hipotezę dotyczącą tejże sytuacji. [...] Wraz z przechodzeniem od jednej sytuacji do kolejnej, jego podstawowym celem staje się opisanie reakcji w zmieniających się okolicznościach – niewykluczone, że dostrzeże on przy tym to, co stanowi o unikalności kontekstów. [...] W miarę kumulowania się danych tego rodzaju, możliwe stanie się zrozumienie tego, jak niekontrolowane okoliczności mogą powodować brak oczekiwanych rezultatów. Oznacza to m.in., że uogólnienia stanowią końcowy efekt procesu badawczego, a wyjątki należy traktować równie poważnie jak zawarte w tych uogólnieniach prawidłowości”<sup>5</sup>.

Oprócz ściśle uwarunkowanych i kontrolowanych eksperymentów, każde badanie ewaluacyjne można uznać za swoiste i niepowtarzalne. Swoistość ta wynika m.in. z celu i zakresu, rodzaju zastosowanych wskaźników i narzędzi, typu prawidłowości do ustalenia, cech respondenta, organizacji i przebiegu badania, zespołu badawczego, środków przeznaczonych na dany projekt oraz instytucji realizującej badanie. To właśnie instytucje mają skłonność do instytucjonalizacji pewnych praktyk badawczych, które są z biegiem czasu doskonalone i nabierają charakteru odrębnych kanonów metodologicznych. W nauce mówimy przeważnie o szkołach czy paradygmatach – w badaniach opinii bądź też w ewaluacji zwykło się mówić o modelach czy schematach badawczych. Robson wyszczególnia jedenaście takich modeli ewaluacji, które w praktyce bardzo często zachodzą na siebie – przez co należy je traktować jako komplementarne w procesie ewaluacji<sup>6</sup>:

- 1) Analiza systemowa polega na pomiarze ilościowym danych na wejściu i wyjściu oraz na badaniu efektywności i wydajności;
- 2) Zoperacjonalizowanie zadania, czyli ustalenie, w jakim stopniu osiągnane są mierzalne, z góry określone cele;
- 3) Ewaluacja przyjmująca za podstawę potrzeby polega na badaniu, na ile zaspokojone zostały potrzeby klienta;

<sup>5</sup> Cronbach L.J., „Beyond the two disciplines of scientific psychology”, *American Psychologist* 1975, nr 30, s. 124–125.

<sup>6</sup> Robson C., „Projektowanie ewaluacji”, w: L. Korporowicz (red.), *Ewaluacja w edukacji*, Oficyna Naukowa, Warszawa 1997, s. 153.

4) Ekspertyza – jej celem jest stwierdzenie, na ile program odpowiada standardom, które ewaluator–ekspert uznaje za doskonałe;

5) Inspekcja kwalifikacji opiera się na tym, że eksperci z zewnątrz oceniają, w jakim stopniu program spełnia obowiązujące standardy profesjonalne;

6) Model konfrontacyjny, w którym dwa zespoły ewaluatorów prowadzą spór na temat zalet i wad programu oraz zastanawiają się, czy program powinien być kontynuowany;

7) Model transakcyjny koncentruje się na przebiegu danego programu;

8) Podejmowanie decyzji – ewaluacja jest zdeterminowana rodzajem decyzji, które należy podjąć;

9) Analiza rozbieżności porównuje zakładane rezultaty programu z jego efektami faktycznymi;

10) Ewaluacja wyjaśniająca koncentruje się na metodach jakościowych, analizie indukcyjnej oraz badaniu naturalistycznym;

11) Ewaluacja dialogiczna kładzie nacisk na uwzględnienie wszystkich stron zaangażowanych w proces i zainteresowanych wynikami ewaluacji.

Skoro podstawowy cel ewaluacji stanowi ocena procesu realizacji interwencji publicznej, rolą badań ewaluacyjnych jest ustalenie rzeczywistego wyniku interwencji, opisanie czynników wpływających na zakładany skutek oraz grup beneficjentów, a na koniec wartościująca ocena wyników danej interwencji. Aby zrealizować powyższe cele, trzeba określić kryteria i ramy metodologiczne, w których ewaluator dokonuje wyboru właściwych i adekwatnych do danej interwencji technik badawczych. Opracowanie metodologii, czyli schematu badawczego, wymaga z kolei znajomości poszczególnych technik badawczych, ich wad i zalet, ograniczeń i możliwości.

## Kryteria doboru technik badawczych

Historia badań społecznych obfituje w punkty zwrotne oraz rewolucje paradygmatyczne, określające stan współczesny tej dziedziny ludzkiej aktywności. Począwszy od pionierskiej działalności badawczej Georga Gallupa, założyciela i dyrektora American Institute of Public Opinion (1935), a skończywszy na rewolucji informacyjnej, która zmieniła wiele tradycyjnych technik badawczych – wiedza i doświadczenie praktyczne zdobyte w tej dziedzinie określiło kanon badań, wykorzystywany przez ewaluatorów na całym świecie. Należy pamiętać, iż badania ewaluacyjne są jednym ze sposobów zastosowania technik badawczych, a nie odrębną metodą<sup>7</sup>. Do oceny efektów danej interwencji publicznej można zastosować szeroką paletę sprawdzonych i skutecznych metod zdobywania informacji oraz analizy danych. Przywołując znane

<sup>7</sup> Babbie E., *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2006.

powiedzenie Paula Feyerabenda: *everything goes* – wszystko, co się sprawdza (w sensie skuteczności oceny i działania), warte jest aplikacji w badaniach ewaluacyjnych. To, jakie techniki warto i należy stosować w konkretnej sytuacji, zależy od wielu kryteriów: począwszy od rodzaju interwencji, która podlega ocenie, celu badania ewaluacyjnego, a skończywszy na szeregu kryteriów kontekstualnych, takich jak np. specyfika beneficjenta.

Pierwszym kryterium doboru metodologii badań ewaluacyjnych jest (1) **uzyskanie rzetelnych i trafnych informacji na temat rodzaju interwencji publicznej**, która ma być oceniana. Duże zróżnicowanie tematów projektów, jak również skala interwencji wymuszają na ewaluatorach z jednej strony elastyczność metodologiczną, z drugiej zaś – otwartość na wiedzę specjalistyczną. Ewaluatorzy nie zawsze mają możliwość poznania specyfiki danej interwencji publicznej na podstawie *desk research* (analizy dokumentacji programowej, danych statystycznych, raportów monitoringowych oraz raportów z wcześniejszych ewaluacji). Taka sytuacja wymaga sięgnięcia w procesie ewaluacji po wiedzę ekspercką, wykraczającą poza zakres samego programu (ekspertyzy, analizy, konsultacje, wywiady eksperckie, panele ekspertów itp.).

Gdy przedmiot ewaluacji (w tym m.in. rodzaj interwencji publicznej, jej specyfika, cele interwencji, poniesione nakłady, podjęte działania oraz osiągnięte produkty, kontekst społeczny i gospodarczy) został już rozpoznany, kolejnym kryterium, jakie należy wziąć pod uwagę przy doborze technik badawczych, jest (2) **wyznaczenie celu badania ewaluacyjnego**. Określenie, czy podejmowana ewaluacja ma na celu stymulowanie usprawnień, rozwój organizacyjny, udoskonalenie kolejnych edycji interwencji (**badanie diagnostyczne**) czy też rozliczenie interwencji i w konsekwencji oszacowanie rezultatów i oddziaływania oraz sprawdzenie zależności pomiędzy nakładami i efektami (**badanie weryfikacyjno-sprawdzające**) warunkuje (3) **zakres badania i charakter poszukiwanych informacji**. I tak, badania skoncentrowane głównie na funkcji formatywnej, nastawione na zrozumienie procesu działania interwencji, rozumienie faktów, eksplorację ukrytego znaczenia badanych zjawisk zachodzących w obszarze programu, innymi słowy mające, przede wszystkim, udzielić odpowiedzi na pytania typu „dlaczego?”, „jak?”, „w jaki sposób?”, będą skupione wokół metod jakościowych<sup>8</sup>. Badania nastawione zaś na **ewaluację konkluzywną**, kontrolną, pomiar i identyfikację faktów oraz odpowiedzi na pytania typu „co?”, „ile?”, „jak często?”, będą zorganizowane wokół metod ilościowych<sup>9</sup>. Praktyka badawcza pokazuje jednak,

<sup>8</sup> Na przykład, badanie służące ocenie jakości systemu zarządzania programem czy ocenie poziomu odpowiedzialności za realizację danej interwencji powinno opierać się na technikach, które pozwolą na określenie czynników wpływających na pracę osób odpowiedzialnych za realizację. Można zatem zastosować m.in. wywiady pogłębione z pracownikami, zarządem, beneficjentami itd., a w przypadku badań realizowanych w trakcie wdrażania programu – również obserwacje uczestniczące, badania typu Mystery Shopping – tajemniczy klient itp.

<sup>9</sup> Na przykład, jeśli celem ewaluacji jest określenie związków przyczynowo-skutkowych między podejmowanymi działaniami a efektami interwencji, konieczne jest zastosowanie technik eks-

że obok celu ewaluacji, formalnych wytycznych, potrzeb informacyjnych zlecających badanie oraz innych odbiorców raportu, **kluczowym czynnikiem dla zakresu badania** jest przede wszystkim **inwencja samych ewaluatorów**. To badacz musi ostatecznie doprecyzować: czy ocena dotyczyć ma całości programu, czy tylko wybranego aspektu lub problemu związanego z interwencją, czy badana będzie synergia z innymi programami, jakie analizy będą konieczne, jak będzie skonstruowany schemat oceny itp.<sup>10</sup>. Oczywiście, im większy zakres ewaluacji, tym więcej informacji trzeba zebrać i przetworzyć, tym bardziej zostanie rozbudowany schemat badawczy, przez co będzie wymagał sięgnięcia po wiele różnych technik badawczych, które muszą być ze sobą komplementarne.

Kolejnym kryterium wpływającym na dobór technik badawczych jest (4) **moment dokonywanej oceny**. Na każdym etapie cyklu procesu ewaluacyjnego (ewaluacje *ex-ante*, *mid-term*, *on-going*, *ex-post*) stosuje się różne metody i techniki badawcze<sup>11</sup>. W przypadku ewaluacji *ex-post* program jest zakończony, mamy dostęp do całej dokumentacji projektów, dużą ilość informacji zgromadzonych w ramach systemu kontroli i monitoringu oraz wcześniejszych badań ewaluacyjnych, pełną populację beneficjentów, kompletne bazy danych, możemy badać wszystkich odbiorców programu z uwzględnieniem efektów i oddziaływań oddalonych w czasie. Jednocześnie jednak na tym etapie (czyli od roku do pięciu lat po zakończeniu interwencji) często mamy utrudniony dostęp do kluczowych aktorów, odpowiedzialnych np. za konstrukcję, zarządzanie, wdrażanie programu, a tym samym ograniczone możliwości realizacji np. pogłębionych wywiadów, organizacji grup foko-

---

perymentalnych lub *quasi*-eksperymentalnych, które pozwolą na oszacowanie efektu netto, poprzez kontrolowany dobór próby kontrolnej, np. za pomocą techniki PSM (*Propensity Score Matching*); na ten temat zob. m.in.: Rubin D.B., Thomas N., „Matching using estimated propensity scores: relating theory to practice”, *Biometrics* 1996, nr 52, s. 249 – 264; Dehejia R.H., Wahba S., „Propensity score-matching methods for nonexperimental causal studies”, *The Review of Economics and Statistics* 2002, nr 84(1), s. 151–161.

<sup>10</sup> Przykładowo, chcąc ocenić efekt wsparcia konkurencyjności małych i średnich przedsiębiorstw, można wyjść poza oficjalne dane (obligatoryjne dla beneficjentów) i opinie przedstawicieli instytucji wdrażających oraz sprawdzić, jak realizację projektu oceniają sami zainteresowani oraz ich bezpośrednie otoczenie, np. inni przedsiębiorcy – konkurencja, społeczności lokalne, przedsiębiorcy nie objęci wsparciem itp. W tym celu trzeba byłoby zaprosić celowo dobraną próbę beneficjentów do uczestnictwa w zogniskowanych wywiadach grupowych bądź też zaprojektować tzw. spotkanie konfrontacyjne, w którym biorą udział także reprezentanci instytucji odpowiedzialnej za realizację danego programu lub też grupy spełniające warunki grupy kontrolnej ze schematu klasycznego eksperymentu. Albo też, jeśli beneficjentem programu są osoby bezrobotne z małej wsi, które uzyskały pomoc i skutecznie z niej skorzystały, ocena danego działania wspierającego powinna obejmować wpływ pomocy nie tylko na samego beneficjanta, jego rodzinę, lecz także na jego bezpośrednie otoczenie społeczne (np. takie elementy jak efekty propagandowe i edukacyjne, dotyczące aktywności tychże osób w ich otoczeniu społecznym). Dodatkowe skutki pozytywne bądź negatywne muszą stanowić równie istotny obszar zainteresowania ewaluatora.

<sup>11</sup> Więcej zob. Olejniczak K., „Ewaluacja jako narzędzie zarządzania w sektorze publicznym”, w: G. Gorzelak, A. Tucholska (red.), *Rozwój, region, przestrzeń*, Ministerstwo Rozwoju Regionalnego, Warszawa 2007, s. 283.

wych itd. Z perspektywy *ex-post* również stosunkowo trudno jest odtwarzać szereg zagadnień kontekstualnych (np. ocenę jakości kilkudniowego szkolenia w perspektywie pół rocznej) czy też związków przyczynowych.

Obok kryteriów wymienionych powyżej, przy planowaniu metodologii badań ewaluacyjnych, należy uwzględnić również szereg **kryteriów kontekstualnych**:

(5) **podmiot zlecający** – kto zleca ewaluację i za nią odpowiada (np. Instytucja Zarządzająca Programem, Instytucja Wdrażająca, Komisja Europejska itp.). Poza elementarnymi kwestiami, takimi jak omówione wcześniej poziom badania, zainteresowania informacyjne oraz cel badania<sup>12</sup>, z punktu widzenia doboru technik istotne jest, z jednej strony, doświadczenie badawcze instytucji zlecającej, z drugiej zaś – tryb samego zamówienia (zgodnie z Prawem Zamówień Publicznych obowiązującym instytucje publiczne), wpływający na rodzaj kontraktu, jego zakres, a tym samym elastyczność umowy zawartej z wykonawcą (czyli możliwość dokonywania ewentualnych zmian w celu dostosowania badania terenowego do realiów już w trakcie jego realizacji<sup>13</sup>);

(6) **podmiot realizujący** – kto realizuje badania. Poszczególne instytuty badawcze różnią się od siebie doświadczeniem w realizacji poszczególnych technik, zapleczem technicznym i osobowym (np. liczbą stanowisk CATI – Computer Assisted Telephone Interviewing, komputerów do badania CAPI – Computer Assisted Personal Interview, ankietatorów itp.), a tym samym szybkością realizacji poszczególnych technik; ponadto – cenami (różnice cen poszczególnych technik badawczych pomiędzy poszczególnymi firmami mogą nawet sięgać 40%), doświadczeniem w zakresie danego obszaru tematycznego ewaluacji (np. badań społecznych i gospodarczych, badań sektorowych, np. w zakresie ochrony środowiska, infrastruktury, sektora małych i średnich przedsiębiorstw, rozwoju zasobów ludzkich, badań organizacji itd.) oraz ekspertami, z którymi współpracują;

(7) **perspektywę podmiotowo-celową** – do kogo dana interwencja jest adresowana, kto jest jej **beneficjentem**, jakie są indywidualne cechy/profil beneficjentów. Należy uwzględnić, z jednej strony, **adekwatność metodologiczną** poszczególnych technik badawczych, czyli przydatność ze względu na problematykę, przedmiot badania, jej informatywność i wiarygodność,

<sup>12</sup> Zob. *Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, <http://www.funduszezstrukturalne.gov.pl/NR/rdonlyres/4E41CBAE-F849-48BE-BAC9-0614753930AA/38773/PlanewaluacjiNSROwrzesien2007180907.pdf>, *System ewaluacji Narodowego Planu Rozwoju na lata 2004–2006 i Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, <http://www.funduszezstrukturalne.gov.pl/NR/rdonlyres/4E41CBAE-F849-48BE-BAC9-0614753930AA/38771/SystemewaluacjiWrzesie180907.pdf>.

<sup>13</sup> W przypadku dużych zamówień publicznych elastyczność umowy z wykonawcą badania ewaluacyjnego jest dość ograniczona. Zob. Ustawa z dnia 29 stycznia 2004 r. Prawo zamówień publicznych (Dz. U. z 2006 r. Nr 164, poz. 1163, Nr 170, poz. 1217 i Nr 227, poz. 1658 oraz z 2007 r. Nr 64, poz. 427 i Nr 82, poz. 560).

**z drugiej – ich adekwatność psychologiczną**, czyli przydatność techniki ze względu na badaną zbiorowość, jej cechy emocjonalne, intelektualne, a także szeroko rozumiane możliwości komunikacyjne respondentów itp. (specyficzne grupy respondentów wymagają zastosowania niekonwencjonalnego aparatu badawczego lub istotnej modyfikacji tradycyjnych technik badawczych<sup>14</sup>);

(8) **źródła, jakość i dostępność danych** gromadzonych przez instytucje zarządzające interwencją oraz Główny Urząd Statystyczny. Badacz musi ustalić, „co wiemy kiedy przystępujemy do ewaluacji?”, oraz, w konsekwencji, „jakie dane musimy pozyskać w wyniku badań terenowych?”, „od kogo?” i „jakie techniki będą najbardziej adekwatne do zebrania tych informacji?”;

(9) **jakość, dostępność, zakres bazy danych o beneficjentach interwencji**. Od dostępności danych o beneficjentach zależą **możliwe kanały komunikacji** z respondentami, a tym samym możliwość **wyboru najbardziej adekwatnej techniki**<sup>15</sup>. Z drugiej strony, dostęp do każdej dodatkowej zmiennej, dotyczącej np. rodzaju, wielkości wsparcia, indywidualnych cech beneficjenta (czyli przykładowo wielkości przedsiębiorstwa, wieku, wykształcenia itp.) **umożliwia m.in. zawężenie i profilowanie badania, skrócenie narzędzia oraz precyzyjniejszy/celowy dobór respondentów do badania** itd. Ponadto należy wskazać, iż:

1) możliwość przetwarzania danych i użycia ich w technikach wykorzystujących oprogramowanie komputerowe zależy od tego, czy dane te występują w wersji papierowej, czy elektronicznej,

2) czas potrzebny na procedury uzyskiwania i obróbki danych zależy od tego, czy były one zbierane przez jedną instytucję, np. ministerstwo, czy przez sieć rozproszonych instytucji, np. Powiatowe Urzędy Pracy,

3) jakość i ewentualne obciążenie błędami, a w konsekwencji użyteczność danych zależy od tego, czy były one na bieżąco wprowadzane bezpośrednio przez beneficjentów (np. generator wniosków), czy wklepywane *ex-post* ze źródeł papierowych, oraz czy aktualizowano je w trakcie realizacji programu.

Na koniec warto również odnotować, że wybór poszczególnych technik badawczych wiąże się jeszcze z dwoma kryteriami kontekstualnymi, które,

<sup>14</sup> Przykładowo, osób niepełnosprawnych ruchowo nie powinno się zapraszać na badanie fokurowe, które odbywa się w lokalizacji uciążliwej ze względu na transport, a osób głuchoniemych nie można badać techniką indywidualnego wywiadu telefonicznego (CATI). Te oczywiste i zdroworozsądkowe sugestie stanowią niejednokrotnie elementarny problem przy doborze właściwej techniki oraz jej poprawnego zastosowania.

<sup>15</sup> Na przykład, przy założeniu kompletności danych dostęp do adresów e-mailowych beneficjentów umożliwia realizację badania jedną z najtańszych technik – CAWI (Computer Assisted Web Interview); dostęp do telefonów umożliwia kontakt z respondentem za pomocą techniki – CATI, dysponowanie zaś wyłącznie adresem pocztowym niejako skazuje badacza na kosztowne techniki PAPI (Paper And Pen Interviewing), CAPI, IDI (In-Depth Interview) bądź ankietę pocztową.



pomimo iż są wtórne wobec powyższych, to często bardzo silnie warunkują ostateczne decyzje metodologiczne. Są to:

(10) **zasoby czasowe** przewidziane na projektowanie, realizację, analizę danych, ocenę i raportowanie wyników. Wybór poszczególnej techniki wiąże się z koniecznością zarezerwowania określonej ilości czasu na jej realizację oraz – w sytuacji użycia technik takich jak PAPI, ankieta pocztowa, inne ankiety papierowe – na wprowadzenie danych do baz umożliwiających analizy statystyczne; w przypadku wszystkich wywiadów indywidualnych i grupowych trzeba poświęcić określoną ilość czasu na transkrypcję;

(11) **zasoby finansowe**, jakimi dysponujemy do realizacji ewaluacji. Kwestia kalkulacji kosztów zastosowania poszczególnych technik badawczych jest wyjątkowo ważna. Z jednej strony, zasoby finansowe, jakimi dysponujemy, determinują możliwość zastosowania metod wzajemnie się uzupełniających (jakościowych i ilościowych), z drugiej zaś – różnice cen pomiędzy poszczególnymi technikami, np. pomiędzy CATI a PAPI albo CATI a CAWI, mogą sięgać nawet 50%. Włączenie do modelu dodatkowych technik istotnie podnosi koszty ewaluacji. Niejednokrotnie argumenty przemawiające za wykorzystaniem konkretnej techniki zderzają się z realiami finansowymi, które ograniczają autorów badania i powodują, iż pierwotnie zakładana argumentacja merytoryczna skazana jest na kompromis z argumentacją ekonomiczną. Nic w tym nowego i dziwnego, jeśli zachowane są reguły poprawności metodologicznej. Pamiętajmy, iż na „oszczędnym” i zarazem właściwym doborze próby, wygrał prekursor badań sondażowych George Gallup – jego oszacowanie wyników wyborczych okazało się prawdziwe (i tańsze od szacunków konkurencji). Projektowanie budżetu badań, w tym znajomość cen i warunków realizacji poszczególnych technik, należy do kanonu wiedzy, którą – poza umiejętnościami technicznymi i warsztatowymi – powinien posiadać każdy badacz.

### **Możliwości i ograniczenia poszczególnych technik badań w ewaluacji *ex-post***

Decyzje o wyborze poszczególnych technik badawczych wiążą się nie tylko z operacjonalizacją problematyki badawczej czy wymienionymi powyżej kontekstualnymi czynnikami projektu badawczego, lecz także z zaletami i ograniczeniami wynikającymi z samych technik. Poza czysto metodologicznymi korzyściami, znajomość cech poszczególnych technik badawczych sprzyja optymalizacji kosztów, nakładu czasu oraz pracy, które muszą zostać poniesione przez ewaluatorów oraz instytucję realizującą badanie.

Podjmując ostateczne decyzje dotyczące wyboru określonej techniki, musimy uwzględnić: **(1)** czasochłonność i koszty techniki badania; **(2)** rodzaj poszukiwanych informacji, np. ze względu na możliwość porównania wyni-

ków i ich generalizacji; **(3)** możliwości zastosowania określonego narzędzia ze względu na poszukiwane informacje i standaryzację narzędzia<sup>16</sup> oraz **(4)** czynniki istotne z uwagi na komunikację z respondentem. Poniższa tabela pokazuje schematyczny podział technik badawczych<sup>17</sup>, które najczęściej wykorzystuje się w procesie ewaluacji *ex-post* oraz najważniejsze cechy poszczególnych technik, jakie powinny być brane pod uwagę przy projektowaniu ewaluacji.

Warto w tym miejscu zaznaczyć, iż, aby uznać trafność narzędzi pomiarowych, należy skupić się nie tylko na dowodzie naukowym (formalnym) ich poprawności, lecz także na sposobie wykorzystania wyników badań. Z jednej strony, celem badań może być **wyjaśnienie**, z drugiej zaś – zastosowanie wniosków dla **celów praktycznych**. Trzeba też zwrócić uwagę na konieczność analizy konsekwencji społecznych (użyteczności) wykorzystania danego narzędzia pomiarowego. Określając trafność stwierdzeń opartych na wynikach, jakie otrzymujemy dzięki narzędziom pomiarowym, bazować należy zarówno na 1) dowodach (*evidence*) poprawności narzędzi (czy to w celu wyjaśnienia, czy zastosowania wyników w praktyce), jak i na 2) konsekwencjach (*consequences*) ich zastosowania (czy to w celu wyjaśnienia, czy zastosowań praktycznych). W tym sensie to nie narzędzie charakteryzuje się trafnością lub jej brakiem, lecz wnioski, które wynikają z użycia narzędzia pomiarowego.

---

<sup>16</sup> Na przykład, oprogramowanie komputerowe (do CATI, CAPI, CAWI) zapewnia utrzymanie prawidłowego przebiegu wywiadu, weryfikuje poprawność wprowadzanych danych, wariantuje logiczny układ kwestionariusza zgodnie ze specyfiką kolejnych odpowiedzi respondenta (program z regułami przejścia), a tym samym większy stopień standaryzacji zachowań ankietera (w przypadku technik realizowanych przy pomocy ankietera) oraz większy stopień kontroli pracy ankietera – czas wywiadu, kontroli pracy z komputerem (liczba cofnięć, przejść, skasowań, pomyłki wciśnięcia klawiszy). Ponadto, techniki wykorzystujące komputery eliminują tradycyjne kwestionariusze papierowe, oszczędzają czas i koszty druku, dystrybucji i przechowywania kwestionariuszy, skracają proces przenoszenia danych na nośniki elektroniczne, wstępną selekcję i weryfikację zgromadzonego materiału badawczego na ekranie, a automatyczne kodowanie uzyskiwanych informacji w czasie wywiadu, daje możliwość szybkiej realizacji badań powiązanych, np. przy realizacji badania z zastosowaniem PSM (Propensity Score Matching), doboru respondentów, tzw. „bliźniaków”.

<sup>17</sup> Obszerniejsze omówienie poszczególnych technik można znaleźć w każdym podręczniku do metodologii badań społecznych, np. Babbie E., *Badania społeczne...*, op.cit.; Nowak S. (red.), *Metody badań socjologicznych*, Państwowe Wydawnictwo Naukowe, Warszawa 1965.

Tabela 1. Czynniki warunkujące wybór technik badawczych

Wybrane techniki badawcze	Standardyzowane					Niestandardyzowane	
	PAPI	CATI	CAWI	CAPI	Ankieta pocztowa	IDI	FGI
<b>Czynniki wpływające na realizację, koszty badania oraz poszukiwane informacje</b>							
Szeroki zakres zastosowań	X	X		X	X	X	
Względnie wysoki koszt jednostkowy wywiadu	X			X		X	X
Względnie niski koszt jednostkowy wywiadu		X	X		X		
Krótki czas realizacji badania		X	X			X	X
Możliwość generalizacji wyników z próby	X*		X*	X*	X*	X	X
<b>Czynniki istotne ze względu na charakter poszukiwanych informacji oraz możliwości techniczne narzędzia</b>							
Technika wykorzystująca ankietę papierową	X				X		
Ustna komunikacja z respondentem	X	X		X		X	X
Pisemna komunikacja z respondentem			X		X		
Wykorzystanie oprogramowania komputerowego umożliwiającego: 1) umieszczenie w ankiecie skomplikowanych filtrów i rotacji, 2) szybką obróbkę zgromadzonych danych		X	X	X			
Długi wywiad (możliwość zadania wielu pytań)	X			X	X	X	
Możliwość zadania refleksyjnych pytań (otwartych, pogłębionych)	X**		X**	X**	X**	X	X
Brak możliwości zadania refleksyjnych pytań (pytania skrótkowe, schematyczne, nieopłębione, z zamkniętymi kaferiami odpowiedzi)		X					
Możliwość prezentacji respondentowi bodźców multimedialnych			X	X		X	X
Kontakt z respondentem za pośrednictwem ankietera/moderatora	X	X		X		X	X
Bepośrednia kontrola pracy ankietera/moderatora		X					X
Wysoki stopień standaryzacji pracy ankietera		X		X			
Pracochłonność obróbki zgromadzonych danych	X				X	X	X

Tabela 1 – cd.

Wybrane techniki badawcze	Standaryzowane				Niestandaryzowane		
	PAPI	CATI	CAWI	CAPI	Ankieta pocztowa	IDI	FGI
<b>Czynniki istotne ze względu na komunikację z respondentem</b>							
Kontrola kontaktu z indywidualnym respondentem	X	X		X		X	X
Możliwość badania słabiej ugruntowanych postaw lub tematów bardziej złożonych						X	X
Większa spontaniczność odpowiedzi	X	X		X		X	X
Wyższe poczucie anonimowości badanych		X	X		X		
Możliwość podejmowania tematów drażliwych (kontrowersyjnych, osobistych i intymnych)							X
Sklonność badanych do wyboru skrajnych punktów skali		X		X			
Szybkie, mało przemyślane odpowiedzi, przerywanie odczytywania kafeletri		X		X			

\* pod warunkiem odpowiedniego stopienia zwrotów; zakłada się, że mamy zmienne determinujące odpowiedzi i jesteśmy w stanie je kontrolować

\*\* zastosowanie ograniczone; można go użyć przy założeniu, że badacz ograniczy się do kilku pytań otwartych

Źródło: Opracowanie własne.

## **Konteksty instytucjonalne: ewaluacja a inne typy badań społecznych**

Na koniec rozważań na temat zagadnień związanych z określaniem metodologii badań ewaluacyjnych warto wspomnieć o kontekście instytucjonalnym, w którym przychodzi pracować ewaluatorom. Do jakiego stopnia jest to kontekst specyficzny, warunkujący parametry badania, spróbujemy ustalić, porównując wybrane aspekty praktyki w różnych środowiskach badawczych.

Ze względu na cele, funkcje, przedmiot, wszechstronność i znaczenie społeczne, badania ewaluacyjne można umieścić pomiędzy badaniami akademickimi a marketingowymi. W klasycznym ujęciu badania ewaluacyjne służą ocenie jakości i wartości konkretnych programów społeczno-gospodarczych, zarówno prywatnych, jak i publicznych. W proceduralnym wymiarze procesu ewaluacji zazwyczaj postuluje się stosowanie podejścia naukowego popartego danymi oraz przestrzeganie rygoru badawczego wynikającego z kanonu badań nauk społecznych. Pomijając akademicki spór o to, czym dla kogo jest ów kanon oraz co i kiedy powoduje kanoniczne zmiany, z różnic w celach oraz kontekstach instytucjonalnych, w których realizuje się dany model procesu badawczego, wynikają oraz instytucjonalizują się różne role i postawy społeczne: akademika, komercyjnego badacza czy ewaluatora.

Specyfika badań ewaluacyjnych polega przede wszystkim na ich publiczno-interwencyjnym charakterze. W odróżnieniu od wizerunku publicznych badań naukowych projekty ewaluacyjne będą częściej postrzegane – skądinąd słusznie – jako instrument bieżącej polityki społeczno-gospodarczej. Abstrahując od potencjalnych skutków politycznej instrumentalizacji programów publicznych<sup>18</sup>, ewaluator stoi przed trudnym zadaniem obiektywizacji swojej roli w procesie wdrażania danego programu. Jeśli badanie ewaluacyjne realizuje instytucja odpowiadająca za zarządzanie programem, powstaje dwuznaczna sytuacja, w której kluczową rolę odgrywa wiarygodność samooceny badacza. Tak więc można postawić hipotezę, iż – z etycznego punktu widzenia – specyfika badań ewaluacyjnych jest czynnikiem wzmacniającym samokontrolę metodologiczną ewaluatorów.

Samokontrolę metodologiczną można rozumieć na dwa sposoby. Po pierwsze, chodzi o stworzenie podczas badania takiej atmosfery, w której znikną potencjalne wątpliwości co do faktycznie pozytywnej, gdyż ukierunkowanej na poprawę danej interwencji publicznej, intencji ewaluatora. Niezmiernie istotne w tej sytuacji jest takie opracowanie narzędzi badawczych, żeby wyraźnie oddzielić informacyjny sens ewaluacji od funkcji audytu czy kontroli. Wypełniając ankietę bądź też biorąc udział w grupie

<sup>18</sup> Zbyt widoczne zaangażowanie polityków w promowanie czy popularyzację danego działania powoduje, iż będzie ono utożsamiane z daną partią czy konkretnym politykiem. W takiej sytuacji oceny danego programu mogą zostać zniekształcone oceną partii czy polityka, który staje się jego medialną ikoną.

fokusowej, respondent (beneficjent) powinien mieć poczucie uczestnictwa w procesie udoskonalania i poprawy efektywności instrumentów bieżącej polityki społecznej. Specyfiką i zarazem atutem badań ewaluacyjnych jest możliwość wywierania presji na uczestnictwo beneficjentów w procesie badawczym. Sposobami, które wyraźnie poprawiają realizowalność badania, są zwykle zapowiednie, monity (pisemne przypomnienie o udziale w badaniu) oraz rozmowy telefoniczne z beneficjentami. Umiejętną perswazję stosuje się z powodzeniem także w badaniach naukowych i komercyjnych, choć jej możliwości są w badaniach ewaluacyjnych ograniczone. Drugim elementem wzmacniającym samokontrolę badacza–ewaluatora jest dywersyfikacja źródeł informacji i technik badawczych. Wnioski z ewaluacji powinny uwzględniać różne punkty widzenia, czemu sprzyja poszerzenie zakresu rzeczywistości społecznej, której podlega badanie. Triangulacja metod i źródeł informacji jest wskazana dla podwyższenia miarodajności oceny danej interwencji publicznej.

Jednym z istotnych czynników dla realizacji badania ewaluacyjnego jest kategoria beneficjentów, stanowiących specyficzną grupę respondentów. Grupy beneficjentów (w roli respondentów) odznaczają się przeważnie cechami wyjątkowymi lub rzadkimi z punktu widzenia ogólnokrajowych badań społecznych. Są to przeważnie osoby objęte programami, których celowość opiera się na pewnej wizji polityki społecznej i planach rozwoju gospodarczego: bezrobotni, bezdomni, niepełnosprawni, właściciele firm *ergo* osoby przedsiębiorcze. Instytucjonalne zobligowanie do udziału w badaniu ewaluacyjnym nie może być wzmacniane ani zakłócanie przez element kontroli i przymusu, potencjalnie wynikający z bycia beneficjentem. Zadaniem ewaluatora jest podtrzymywać prawdziwe wrażenie i poczucie dobrowolności udziału w badaniu oraz szczególnie podkreślać podmiotowość respondenta.

Inny aspekt związany z kategorią respondenta to jego rola społeczna. Dla ewaluatora respondentem jest zwykle beneficjent danego programu (podmiot instytucjonalny lub osoba fizyczna), dla firmy badawczej zaś będzie nim konsument produktu, a dla akademika – z reguły członek grupy, reprezentujący szerszą kategorię w palecie standaryzowanych i uniwersalnych ról społecznych. Z tym aspektem wiążą się inne czynniki kontekstualne, które wpływają na charakter pracy i społeczne postrzeganie osób realizujących badania ewaluacyjne (np. prestiż instytutu badawczego czy sytuacja polityczna w kraju).

Chociaż jest to sytuacja patologiczna z metodologicznego punktu widzenia, warto zaznaczyć, że określone instytucje badawcze – zwłaszcza instytucje państwowe oraz ośrodki badania opinii publicznej – podlegają mniej lub bardziej nasilonym ocenom politycznym, które mogą przekładać się na ich społeczne postrzeganie (wizerunek instytucji badającej). Burzliwe wydarzenia, kampanie wyborcze, afery czy przełomy polityczne uczulają społeczeństwo na pewne kwestie, wywołując dodatkowo „efekt sponsora”; ma on jednak

– jak wynika z analizy wyników badań z roku 1992 – niewielkie znaczenie dla deklaracji wzięcia udziału w danym badaniu<sup>19</sup>.

Od dawna metodolodzy zwracali także uwagę na regularnie pojawiające się różnice w wynikach uzyskiwanych przez różne instytucje prowadzące badania, a w szczególności przez ośrodki badawcze. Piotr Klarkowski po wyeliminowaniu potencjalnych źródeł zniekształceń wyników pomiaru (takich jak odmienność technik badawczych, różnice między próbami czy dystans czasowy między badaniami) zaobserwował występowanie istotnych i systematycznych różnic pomiędzy ośrodkami badawczymi w Polsce. Choć właściwych przyczyn tzw. „efektu ośrodka” możemy się jedynie domyślać (rzetelność pracy ankieterów, doświadczenie i styl pracy badacza w danym ośrodku...), z analiz Klarkowskiego płyną następujące wnioski: „1) wyniki badań pochodzące z jednego pomiaru i wykonane przez jedną instytucję badawczą są obciążone odchyleniem od wyniku rzeczywistego o nieznaną wielkość; 2) replikacja badań przez różne firmy badawcze na różnych próbach jest najbezpieczniejszą strategią poszukiwania wyniku zbliżonego do wyniku prawdziwego; 3) w przypadku badania realizowanego równolegle należy wprowadzić do modelu wyjaśniającego różnice między ośrodkami jako zmienną niezależną; 4) badania ewaluacyjne, polegające na testowaniu hipotez o istotności zmian w następstwie oddziaływań, powinny być replikowane przez różne firmy badawcze, gdyż różnice między pomiarami tej samej firmy na różnych próbach reprezentacyjnych są niestabilne, co prowokuje do pochopnego przyjmowania wniosku o zmianach”<sup>20</sup>. Ewaluator powinien pamiętać, że zarówno w przypadku wizerunku danego ośrodka badawczego, jak i w kwestii informacji o sponsorze badania, nie wiadomo dokładnie, jak dalece oba czynniki kontekstualne wpływają na treść wypowiedzi respondentów, a tym samym, jaki jest stopień zniekształcenia wyników pomiaru, w tym także badania ewaluacyjnego.

Dla realizacji ewaluacji istotny jest także charakter współpracy stron projektujących, realizujących i zlecających dany projekt. Naukowe instytucje badawcze mają przeważnie, w aspekcie profesjonalnym, bliższy kontakt z autorami badania, którzy również są akademikami obeznanymi z metodolo-

<sup>19</sup> Dla przykładu, Gostkowski i Przybyłowska ustalili, że przeszło 46% respondentów chciałoby być poinformowanymi o tym, kto jest finansowym sponsorem sondażu, najczęściej uzasadniając taką opinię dezaprobatą wobec anonimowości sponsora. Ci, którzy nie byli zainteresowani taką informacją, uzasadniali to stałością i niezależnością opinii, ogólnym brakiem zainteresowania „polityką”, generalną dezaprobatą wobec sondaży lub akceptacją pluralizmu. Najczęściej informację o sponsorze chcieli uzyskać respondenci z wyższym poziomem wykształcenia, mieszkańcy miast, zamożniejsi oraz ci, którzy już wcześniej zetknęli się osobiście z badaniami sondażowymi. Znikoma część respondentów oświadczyła, że dowiedziawszy się, iż sponsorem jest nieakceptowana przez nich partia – odmówiłaby wywiadu. Zob. Gostkowski Z., Przybyłowska I., „Problem sponsora sondażu w oczach respondentów CBOS w roku 1992”, *ASK. Społeczeństwo. Badania. Metody* 1996, nr 2(4).

<sup>20</sup> Klarkowski A., „Szacowanie i kontrolowanie wpływu ośrodka badawczego w analizach dynamiki nastrojów społecznych”, *ASK. Społeczeństwo. Badania. Metody* 1996, nr 1(4), s. 70.

gią i najczęściej sami opracowują narzędzia, niż firmy badawcze z komercyjnymi klientami, co prawda dobrze orientującymi się w treści badania, natomiast mającymi zazwyczaj niewielką wiedzę metodologiczną. W praktyce, znajomość warsztatu u zlecającego badanie jest ważnym czynnikiem wpływającym na jakość oraz sprawność realizacji danego projektu badawczego. W ewaluacji badanie zamawia przeważnie administracja publiczna, a jej przedstawiciele rzadko kiedy mają wiedzę i kwalifikacje metodologiczne. Nakłada to na ewaluatora obowiązek dodatkowych konsultacji, których efektem musi być pogodzenie możliwości technicznych i finansowych z oczekiwaniami instytucji zlecającej. Wspólnemu ustaleniu celów badania oraz wskaźników empirycznych powinna towarzyszyć kwestia doboru właściwych technik, tak aby spełniały one warunki poprawności metodologicznej, przy jednoczesnym szacowaniu kosztów i czasu realizacji badania.

Nierzadko oczekiwania strony zlecającej badanie – czy to będzie administracja publiczna, czy klient komercyjny – mijają się z technicznymi możliwościami i warunkami, w których ma być realizowana „koncepcja badawcza”. Dotyczy to nie tylko czasu i kosztów, lecz także problematyki samych narzędzi badawczych, nieprzystających do realiów danej grupy beneficjentów czy społeczeństwa.

Wydaje się, że świadomość różnic w praktykach badawczych jest powszechna zarówno wśród przedstawicieli środowisk ewaluacyjnych, jak i reprezentantów komercyjnych instytutów badania opinii oraz środowiska akademickiego. Nie zmienia to jednak faktu, iż obowiązującym standardem normatywnym jest deklarowanie przez środowiska nie-akademickie postulatów przestrzegania rygoru badawczego, wynikającego z kanonu badań nauk społecznych.

## Podsumowanie

Ewaluacja to systematyczne badanie wartości programu, które ma za zadanie w ściśle określonym zakresie dostarczyć informacji przede wszystkim na temat jego skuteczności i efektywności.

Aby odpowiedzieć na postawione pytania dotyczące rzeczywistości programu, konieczne jest pozyskanie określonego typu danych, spełniających szereg warunków gwarantujących ich należyłą jakość oraz poprawność. Innymi słowy, dane trzeba pozyskiwać: 1) z zachowaniem określonych reguł, 2) z uwzględnieniem warunków takich jak specyfika badań ewaluacyjnych, specyfika sektorowa i regionalna, kontekst instytucjonalny, oraz 3) z uwzględnieniem właściwych kryteriów doboru technik badawczych (ich możliwości, ograniczeń) i zasad ich wykorzystania w trakcie procesu badawczego.

Jak wskazywaliśmy, zagadnienie doboru technik badawczych, ich sprawnej integracji oraz organizacji jest – po ustaleniu celów i kryteriów – kluczo-



wym punktem każdego procesu ewaluacyjnego. Jednak wielu zamawiającym i wykonawcom nadal zbyt często wydaje się, że to problem wtórny z punktu widzenia przygotowania projektu ewaluacyjnego oraz niedostrzegalny z perspektywy produktu końcowego – raportu. A przecież od tej decyzji zależy w dużej mierze trafność i użyteczność oceny końcowej, która świadczy o jej faktycznej jakości.

Mamy nadzieję, że w tym krótkim artykule udało nam się przybliżyć istotę i znaczenie zagadnienia doboru technik badawczych w procesie ewaluacji oraz zainicjować lub rozwinąć Państwa przemyślenia w zakresie konstrukcji własnych projektów ewaluacyjnych.

---

**Maciej D. Kryszczuk** – doktor nauk humanistycznych w zakresie socjologii, adiunkt w Wyższej Szkole Przedsiębiorczości i Zarządzania im. L. Koźmińskiego, przewodniczący Sekcji Socjologii Pracy PTS, członek zespołu badawczego ZPANS przy IFiS PAN, niezależny ekspert ds. badań, absolwent Szkoły Nauk Społecznych przy IFiS PAN, autor szeregu publikacji z zakresu socjologii struktury społecznej, społeczeństwa informacyjnego oraz komputeryzacji pracy, wykładowca m.in. Collegium Civitas, WSPiZ im. Koźmińskiego oraz WSHE w Łodzi.

**Agnieszka Haber** – absolwentka Wydziału 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 Zasobów Ludzkich, Rozwoju sektora MSP oraz projektów infrastrukturalnych wdrażanych przez Agencję. Od 2005 roku organizator międzynarodowej cyklicznej Konferencji Ewaluacyjnej, odbywającej się pod auspicjami MRR oraz PARP. Od 2005 roku uczestnik grup ds. ewaluacji kolejnych programów: EFS, PO IG, PO KL, PO PRPW. Autorka prowadzi wykłady i warsztaty z zakresu ewaluacji interwencji publicznych w Wyższej Szkole Przedsiębiorczości i Zarządzania im. L. Koźmińskiego, Krajowej Szkole Administracji Publicznej oraz Uniwersytecie Warszawskim.

## Bibliografia

- Babbie E., *Badania społeczne w praktyce*, przekł. W. Betkiewicz i in., Wydawnictwo Naukowe PWN, Warszawa 2006.
- Cronbach L.J., „Beyond the two disciplines of scientific psychology”, *American Psychologist* 1975, nr 30, s. 116–127.
- Dehejia R.H., Wahba S., „Propensity score-matching methods for nonexperimental causal studies”, *The Review of Economics and Statistics* 2002, nr 84(1), s. 151–161.
- Domański H., Sawiński Z., Sztabiński F., Sztabiński P., „Badania społeczne w latach dziewięćdziesiątych. Kontynuacje i zmiany”, *ASK. Społeczeństwo. Badania. Metody* 1995, nr 1(4).

- Gostkowski Z., Przybyłowska I., „Problem sponsora sondażu w oczach respondentów CBOS w roku 1992”, *ASK. Społeczeństwo. Badania. Metody* 1996, nr 2(4).
- Klarkowski A., „Szacowanie i kontrolowanie wpływu ośrodka badawczego w analizach dynamiki nastrojów społecznych”, *ASK. Społeczeństwo. Badania. Metody* 1996, nr 1(4).
- Korporowicz L. (red.), *Ewaluacja w edukacji*, Oficyna Naukowa, Warszawa 1997.
- Kwiatkowski P., „Socjologia a badania opinii publicznej”, *ASK. Społeczeństwo. Badania. Metody* 1996, nr 2(4).
- MacDonald B., „Mandarins and lemons – the executive investment in program evaluation”, *Phi Delta Kappa CEDR Quarterly* 1982, nr 2.
- Nowak S. (red.), *Metody badań socjologicznych*, Państwowe Wydawnictwo Naukowe, Warszawa 1965.
- Olejniczak K., „Ewaluacja jako narzędzie zarządzania w sektorze publicznym”, w: G. Gorzelak, A. Tucholska (red.), *Rozwój, region, przestrzeń*, Ministerstwo Rozwoju Regionalnego, Warszawa 2007.
- Patton M.Q., *Qualitative Evaluation and Research Method. Second Edition*, Sage Publications, London 1990.
- Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, <http://www.fundusze-strukturalne.gov.pl/NR/rdonlyres/4E41CBAE-F849-48BE-BAC9-0614753930AA/38773/PlanewaluacjiNSROwrzesien2007180907.pdf>.
- Robson C., „Projektowanie ewaluacji”, w: L. Korporowicz (red.), *Ewaluacja w edukacji*, Oficyna Naukowa, Warszawa 1997.
- Rubin D.B., Thomas N., „Matching using estimated propensity scores: relating theory to practice”, *Biometrics* 1996, nr 52, s. 249–264.
- Sawiński Z., „Sondaże telefoniczne”, *ASK. Społeczeństwo. Badania. Metody* 1996, nr 1(4).
- Stake R.E., „Toward a technology for the educational evaluation”, *Teacher College Record* 1967, t. 68, s. 523–540.
- Sztabiński P., Sztabiński F., Sawiński Z., *Podręcznik ankietera*, SMG/KRC, IFiS PAN, Warszawa 2000.
- Ustawa Prawo zamówień publicznych, z dnia 29 stycznia 2004 r., Dz. U. z 2006 r. Nr 164, poz. 1163, Nr 170, poz. 1217 i Nr 227, poz. 1658 oraz Dz. U. z 2007 r. Nr 64, poz. 427 i Nr 82, poz. 560.

Jarosław Górniak, Karolina Keler

## Wskaźniki w ewaluacji *ex-post* programów publicznych

### Wprowadzenie

Obecność niniejszego artykułu w tomie poświęconym ewaluacji *ex-post* zwołania nas z obowiązku omawiania pojęcia i funkcji ewaluacji oraz różnic w podejściach do sposobu jej prowadzenia. Niezależnie od tego, jakie stanowisko zajmiemy w toczącej się dyskusji, jedno pozostaje niewątpliwe: w toku ewaluacji trzeba ustalić, czy (i w jakim stopniu) zostały zrealizowane cele interwencji. Bez tego nie będą miały dobrej podstawy ani ocena skuteczności projektu czy programu, ani identyfikacja związków przyczynowych pomiędzy instrumentami a efektami, ani działania usprawniające. Błędem jest mniemanie, że dokonując ewaluacji można sobie pozwolić na zaniechanie ustalenia efektów podjętych działań: zaprzecza to w ogóle istocie ewaluacji. Wprowadzanie w życie projektów czy programów ma na celu uzyskanie zaplanowanej zmiany (w szczególnych przypadkach może chodzić też o powstrzymanie zmiany). Kluczowym zagadnieniem ewaluacji *ex-post* jest pomiar wielkości zmiany wywołanej podjętą interwencją i w konsekwencji stwierdzenie, czy przyjęta „teoria programu” okazała się być adekwatna<sup>1</sup>.

Ewaluacji nie da się oddzielić od procesu programowania działań. To, czy działanie zostało dobrze zaprogramowane, można sprawdzić m.in. przez określenie, czy program jest w ogóle podatny na ewaluację, czyli czy możliwe jest ustalenie stopnia osiągnięcia przez niego stawianych celów. Cele, stany rzeczy, ku którym zmierza program czy projekt ujmowane są w postaci pojęć. Jeśli chcemy posługiwać się tymi pojęciami w sposób niepowodujący nieporozumień, powinny być one precyzyjnie zdefiniowane. Cele często mają charakter złożony: do ich określenia używa się wielu pojęć, a każde z nich ujmuje określony wymiar przyszłej sytuacji stanowiącej cel działania. Żeby określić, czy cel został osiągnięty lub w jakim stopniu zbliżyliśmy się do jego osiągnięcia (gdy dystans do celu jest co najmniej stopniowalny) musimy mieć możliwość wyznaczenia istotnych wymiarów celu w kategoriach „zmien-

---

<sup>1</sup> Na temat ewaluacji opartej na teorii por. Birckmayer J.D., Weiss C.H., „Theory-based evaluation in practice. What do we learn?”, *Evaluation Review* 2000, vol. 24, nr 4, s. 407–431. Kwestię tę poruszamy też w części poświęconej celom interwencji publicznej.

nych” poddających się pomiarowi, czyli ująć go w postaci „wskaźników” (o czym piszemy szczegółowo dalej). Nie jest zatem niczym zaskakującym, że Komisja Europejska wymaga od jednostek projektujących, by ustanowiły wskaźniki dla celów zakładanych w programach czy projektach. Wskaźniki te mają służyć monitoringowi (w przypadku, gdy istnieje możliwość bieżącego śledzenia wartości wskaźnika) oraz okresowym i końcowym ewaluacjom zaplanowanych przedsięwzięć. Z perspektywy metodyki zarządzania projektami naturalne jest oczekiwanie, że efekty wszelkich działań angażujących zasoby organizacyjne powinny być śledzone przy pomocy wskaźników osiągnięć (tzw. *performance indicators*), wśród których zwykle wyróżnia się te o kluczowym znaczeniu (*key performance indicators*).

Wskaźniki sprawiają poważne kłopoty jednostkom programującym. Wynika to zwykle z dużej złożoności celów i małej precyzji przy ich formułowaniu lub braku jasnych definicji pojęć użytych w ich opisie. Niekiedy wskaźniki nie są ustanawiane w oparciu o precyzyjną eksplikację celów i po namyśle nad najlepszym sposobem ich pomiaru, lecz wybiera się je z różnego rodzaju katalogów czy indykatywnych przewodników metodologicznych. Powoduje to odebranie ich od istoty tego, co powinny mierzyć. Wreszcie, przyczyną problemu projektodawców z wskaźnikami bywa niezrozumienie ich istoty i funkcji.

Wskaźniki (a także ich brak) sprawiają kłopoty także ewaluatorom, gdy cele i wskaźniki programu nie są jasno zdefiniowane i nie spełniają wymogów poprawnej ewaluacji. Ewaluatorzy stają wówczas przed pytaniem: czy precyzować cele i ustalać wskaźniki na własną rękę, by móc przeprowadzić oceny skuteczności i efektywności programu? Poza tym, ewaluator przeprowadzający ewaluację *ex-post* staje wobec konieczności zmierzenia się z ustaleniem użyteczności programu: na ile program rozwiązał istotne problemy społeczne czy zaspokoił potrzeby beneficjentów (niezależnie od tego, czy ujęto je w celach programu czy projektu). Także ten aspekt ewaluacji wymaga operowania wskaźnikami, tyle że będą to raczej wskaźniki opisujące kontekst społeczny programu, a nie bezpośrednio jego cele.

Zasadniczym zadaniem niniejszego tekstu jest określenie funkcji wskaźników w ewaluacji *ex-post* polityk publicznych oraz założeń i zasad dotyczących ich stosowania dla tego celu. Jednakże, jak wynika z zasygnalizowanych wyżej problemów, omówienie roli wskaźników w ewaluacji *ex-post* wymaga odwołania się do całościowej logiki programowania polityk publicznych, w szczególności zaś do procesu definiowania celów interwencji, a także do zadań ewaluacji *ex-ante* i śródkresowej (*mid-term*). Dopiero w tak szerokim kontekście można przedstawić zagadnienie wskaźnika jako narzędzia w ramach ewaluacji *ex-post*. Ustalenie wskaźników rozpoczyna się bowiem już w fazie formułowania celów i planów działania, następnie podczas ewaluacji *ex-ante* dokonuje się ich weryfikacji. W przypadku programów o dłuższym horyzoncie czasowym taka weryfikacja powinna też zostać przeprowadzona w trakcie pierwszej ewaluacji śródkresowej. Ustalenie w wyniku tego procesu adekwatnego

zestawu wskaźników jest kluczowe dla możliwości realizacji pełnej i poprawnej ewaluacji *ex-post*, a w konsekwencji dla ustalenia efektów interwencji.

Problematyka wskaźników jest obecna w ogromnej liczbie publikacji. Publikacje te mają często charakter „praktyczny”: dostarczają zasad doboru wskaźników lub ich konstruowania, podają przykładowe lub zalecane listy wskaźników, klasyfikują je i wiążą z poszczególnymi elementami logiki programu. W efekcie uzyskujemy obraz bogaty, ale nieuporządkowany, w którym podstawowe, uniwersalne kwestie zostają zatarte. Aby dobrze budować czy dobierać wskaźniki, warto zadać sobie pytanie o ich istotę i ogólne właściwości. To pozwoli ujrzeć w innym świetle znane „typologie” czy charakterystyki wskaźników i formułowane w publikacjach zasady stosowania.

### **Wskaźnik jako narzędzie pomiaru efektu interwencji**

Aby uchwycić zmianę wywołaną interwencją i, w konsekwencji, ocenić interwencję, musimy być w stanie zmianę tę zarejestrować. Niekiedy można to zrobić bezpośrednio, innym razem trzeba wnioskować na podstawie obserwacji różnych zjawisk powiązanych z tym, które nas właściwie interesuje. Wskaźniki to zmienne poddające się pomiarowi i pozostające w związku z interesującymi nas cechami analizowanych obiektów czy stanów rzeczy, przez co wykorzystuje się je do stwierdzenia, czy dany stan rzeczy zachodzi lub, gdy mamy do czynienia z charakterystykami dającymi się ująć ilościowo, jaki jest jego rozmiar. To ważne, by uświadomić sobie, że słowem „wskaźnik” określamy zmienną, czyli pewną cechę badanych obiektów czy stanów, która może przyjmować różne wartości (przynajmniej dwie, np. istnienie lub nieistnienie danej właściwości), a nie obiekt lub wielkość liczbowa. Liczbę traktujemy jako wartość przypisaną wskaźnikowi w konkretnej sytuacji (np. przed interwencją lub po niej), ale nie jest ona rozumiana w tym ujęciu jako sam „wskaźnik”. Często kwestie te są mylone, podobnie jak zapomina się, że w roli wskaźników nie mogą występować charakterystyki teoretyczne, abstrakty określane też mianem konstruktów, niepoddające się bezpośredniemu pomiarowi. Do takich konstruktów należą m.in. często występujące w rozmaitych programach pojęcia, jak np.: konkurencyjność, spójność społeczna, ubóstwo, innowacyjność itp.

Wskaźnik może być związany z pojęciem charakterystyki, której pomiarowi służy – tzw. *indicatum* – na mocy definicji wiążącej ogólniejsze pojęcie (np. ubóstwo) ze zmiennymi poddającymi się pomiarowi. Przykładowo bycie ubogim może być definiowane jako bycie członkiem gospodarstwa domowego o dochodach na osobę (ekwiwalentnego dorosłego) niższych niż 60% mediany dochodów ekwiwalentnych w analizowanej populacji. Pragniemy jednak podkreślić, że nie jest to jedyne dopuszczalne powiązanie<sup>2</sup>. Wskaźnikiem może być zmienna, która pozostaje w zależności statystycznej

(korelacyjnej) ze zmienną wskazywaną (*indicatum*). Sama zmienna wskazywana potencjalnie bywa empirycznie dostępna, ale z powodów prawnych lub finansowych kłopotliwa w pomiarze.

Oprócz wyżej wspomnianych związków, w literaturze metodologicznej opisuje się też tzw. wskaźniki inferencyjne<sup>3</sup>. Wskaźniki takie traktowane są jako zmienne o wartościach będących skutkiem oddziaływania zmiennych ukrytych (latentnych, teoretycznych), których nie można poddać bezpośrednio pomiarowi. Przykładami stosowania wskaźników inferencyjnych mogą być zmienne ujmowane przez tzw. skale, jak np. skala poczucia bezpieczeństwa. Nieco odmienny status mają tzw. indeksy złożone, np. indeks rozwoju społecznego UNDP (*Human Development Index*). Złożone statystyczne aspekty tak skonstruowanego pomiaru i szczegółowa dyskusja różnic pomiędzy skalami a indeksami wykraczają poza ramy tego artykułu. W tym miejscu chodzi nam bowiem jedynie o wskazanie bogactwa relacji, które mogą występować pomiędzy zmiennymi będącymi przedmiotem zainteresowania a wskaźnikami służącymi do empirycznego określenia ich poziomu.

W kontekście przyjętego tu rozumienia wskaźnika spójrzmy na jego definicję zawartą w jednym z dokumentów Komisji Europejskiej (KE): „An indicator can be defined as the measurement of an objective to be met, a resource mobilised, an effect obtained, a gauge of quality or a context variable” („Wskaźnik może być zdefiniowany jako miara celu, który ma zostać osiągnięty, zaangażowanych zasobów, uzyskanego efektu, miernik jakości lub zmienna kontekstowa”)<sup>4</sup>. Słowo „miara”, oddające angielski termin „measurement”, oznacza tu skalę, przy pomocy której ustalamy wartość, jaką przyjmuje badany obiekt. Różne obiekty mogą przyjmować różne wartości, mamy tu więc do czynienia ze zmienną, a nie konkretną wartością. Taka interpretacja pozwala traktować tę definicję jako zbliżoną do definicji podanej przez nas. Widać jednak, że definicja wskaźnika przyjęta przez KE nie jest dobra, gdyż większą wagę przywiązuje do tego, jakie ważne dla programu elementy można poddawać pomiarowi przy pomocy wskaźników, niż do tego, jak rozumieć sam termin „wskaźnik”.

Wspomnieliśmy już o tym, że wskazuje się na różne rodzaje powiązań między wskaźnikiem a *indicatum* (definicyjne, empiryczne i inferencyjne)<sup>5</sup>.

<sup>2</sup> Opisane powiązanie może wręcz rodzić rozmaite komplikacje, które są omawiane w literaturze poświęconej metodologii badań i filozofii nauki po hasłem „operacjonalizmu”; dyskusję tę tutaj pomijamy.

<sup>3</sup> Por. np. Nowak S., *Metodologia badań społecznych*, Państwowe Wydawnictwo Naukowe, Warszawa 1985.

<sup>4</sup> European Commission, *Indicators For Monitoring And Evaluation: A Practical Guide*, The New Programming Period, 2007-2013: Methodological Working Papers, Draft Working Paper [X], DG Regional Policy 2006, [http://ec.europa.eu/regional\\_policy/sources/docoffic/working/doc/indic\\_sf2007-2013\\_en.pdf](http://ec.europa.eu/regional_policy/sources/docoffic/working/doc/indic_sf2007-2013_en.pdf).

<sup>5</sup> Definicyjne – na mocy definicji wiążącej wskaźnik z *indicatum*. Empiryczne (rzeczowe, korelacyjne) – wymagające uzasadnienia empirycznego powiązania dwóch zmiennych. Inferencyjne

Warto podkreślić, że różne sytuacje badawcze wymagają korzystania z różnych rodzajów powiązań i wyboru określonego typu wskaźnika. Na bazie tego, co zostało wyżej powiedziane, ustalanie wskaźników dla celów oznacza wybór mierzalnych zmiennych reprezentujących istotne wymiary celu. Jak należy to rozumieć? W szczególnych przypadkach, w sytuacji związku definicyjnego może być tak, że cel daje się ująć bezpośrednio przy pomocy mierzalnej zmiennej wyczerpującej definicję cechy, ze względu na którą ten cel jest ustanawiany. Dzieje się tak rzadko, a właściwie tylko wtedy, gdy mamy do czynienia z celem jednowymiarowym, spotykanym raczej przy definiowaniu celów najniższego szczebla (najbardziej szczegółowych). W przypadku celów wielowymiarowych (złożonych) lepiej użyć większej liczby wskaźników. Może się zdarzyć, że spotkamy się z celami sformułowanymi w kategoriach cech niepoddających się bezpośrednio pomiarowi; przy określaniu celów interwencji publicznych nierzadko bowiem mamy do czynienia ze zmiennymi teoretycznymi, określanymi mianem konstruktów lub, według innej konwencji, zmiennych ukrytych, nie podlegających bezpośredniemu pomiarowi, np.: konkurencyjność, postęp techniczny, zatrudnialność i wiele innych. W takich sytuacjach, gdy ograniczymy się do wskazania jednego wskaźnika definicyjnego, zawężymy cel do jednego wymiaru problemu. Ze względu na fakt, że cele zwykle nie dają się prosto przełożyć na pojedyncze miary, użycie tylko jednego wskaźnika musi być świadomą decyzją, wynikającą np. ze wskazania najistotniejszego, ważnego z punktu widzenia przyjętych wartości, wymiaru problemu. Wynika z tego rekomendacja, by przy określaniu wskaźników do złożonych celów korzystać także z relacji inferencyjnych czy stosować zestawy wskaźników.

Warto zwrócić uwagę na to, że wskaźniki nie stanowią doskonałego odzwierciedlenia interesujących nas właściwości. Wskaźniki obarczone są błędem pomiarowym wynikającym z ograniczonej rzetelności naszych narzędzi; metodologia zna jednak sposoby zwiększania rzetelności pomiaru. Wszystkich wskaźników, z wyjątkiem pojedynczo użytych wskaźników definicyjnych, dotyczy dodatkowy błąd wynikający z tego, że wskaźnik nie jest tożsamy ze zmienną wskazywaną, a ujmuje ją tylko częściowo (dążymy zaś do tego, by ujmował ją jak najlepiej). Błędne też może okazać się założenie, iż oddaje on trafnie istotę mierzonego celu. Dlatego dużej wagi nabiera zagadnienie oceny i podnoszenia jakości wskaźników. Ponadto, powiązanie wskaźnika ze zmienną wskazywaną wymaga podstaw teoretycznych (zależność definicyjna lub inferencyjna) lub dowodów empirycznych (zależność empiryczna). Nauki społeczne są skarbnicą wiedzy na temat takich zależności.

Konkludując, autorzy programów czy projektów w ramach polityk publicznych odpowiadają za sformułowanie (definiowanie) celów, umożli-

---

– wymagające większej ilości wskaźników i algorytmów statystycznych potrzebnych do określenia zmiennej ukrytej.

wiające ustalenie sposobu pomiaru efektów interwencji. W praktyce to ustalenie wskaźników dla charakterystyk, które chcemy zmienić poprzez interwencję, oraz wyznaczenie ich planowanych wartości określa ostatecznie, co ma być faktycznie osiągnięte. Rozważmy tę kwestię na przykładzie. Załóżmy, że celem naszej interwencji jest obniżenie poziomu ubóstwa w określonej społeczności. Aby sprawdzić, czy w efekcie realizacji programu ubóstwo spadnie, musimy je zmierzyć, czyli posłużyć się wskaźnikiem (lub wskaźnikami) trafnie dobranym do postawionego celu. Zanim to jednak nastąpi, musimy najpierw, formułując cele, dokładnie sprecyzować, co mamy na myśli, gdy mówimy o ubóstwie, którego rozmiar zamierzamy zmniejszyć. (1) Należy doprecyzować, o jaki rodzaj ubóstwa nam chodzi: subiektywne czy obiektywne, absolutne czy relatywne, i w jakiej sytuacji będziemy klasyfikować osobę jako ubogą. (2) Należy określić, jak trzeba mierzyć ubóstwo zgodnie z przyjętą przez nas definicją i wybrać wskaźnik lub wskaźniki, które trafnie oddadzą jego istotę. (3) Można ustalić wskaźniki pomocnicze, uzupełniające obraz uzyskiwany przy pomocy wskaźników głównych. Ostatnia uwaga dotyczy zwłaszcza celów złożonych (wielowymiarowych), w ramach których określenie wskaźników kluczowych nadaje szczególną wagę poszczególnym wymiarom celu. Gdy zadbamy o precyzję celów, będziemy mieć pewność, że przypadkowo ustalone wskaźniki i oparta o nie ewaluacja nie będą niezgodne z naszymi zamierzeniami.

## Cele interwencji publicznych

„develop indicators within the discussion on the action in general, not to «add» indicators when the discussion has been finished”  
(„określ wskaźniki w toku dyskusji o całości działania, nie «dodawaj» wskaźników, gdy dyskusja została zakończona”)<sup>6</sup>

## Pojęcie celu – hierarchia celów

W programowaniu polityk i projektów publicznych szczególnie ważne jest właściwe sformułowanie celów działania. Stanowi to jednocześnie jedną z największych trudności procesu programowania polityk publicznych. Cele często są mylone ze środkami działania i samymi działaniami. Często także formułuje się je w sposób niejasny i nieprecyzyjny. Cele powinny być określone w kategoriach stanu rzeczy, który chce się osiągnąć. W metodyce programowania, gdy mowa jest o celach, język angielski rozróżnia pojęcia *goal* i *objective*. W języku polskim nie mamy dobrych odpowiedników tych pojęć. Oba tłumaczymy jako cel, przez co w dokumentach programowych autorzy nierzadko ograniczają się do podawania tylko **goals, czyli celów sformuło-**

<sup>6</sup> European Commission, *Indicators For Monitoring And Evaluation...*, op.cit., s. 11.



**wanych w kategoriach ogólnej, jakościowej wizji.** Wizja ta jest, oczywiście, koniecznym i ważnym punktem wyjścia do działania, stanowiącym podstawę formułowania programów i mobilizującym do jakościowej refleksji nad ich przebiegiem i wynikami. Jednak istotą planowania interwencji i bazą dla ewaluacji powinny być raczej **cele konkretne, wymierne, racjonalnie sformułowane i osiągalne w danych okolicznościach oraz posiadające wyraźną perspektywę czasową – takie cele zwykle określa się mianem *objectives*.**

Dobrze jest ujmować je w odniesieniu do wartości tych samych zmiennych w punkcie wyjścia, czyli w kategoriach zmiany (od stanu wyjściowego do momentu zakończenia działań). Prawdłowo ujęte cele interwencji powinny dawać się łatwo wyrazić w kategoriach wskaźników, czyli zmiennych, których wartości (a dokładniej: zmiana wartości) pozwolą nam określić, czy cel został osiągnięty lub w jakim stopniu się do niego zbliżyliśmy. Wartości docelowe wskaźników (po angielsku *targets*, co również, niestety, tłumaczy się jako „cele”) nie są tożsame z celami (w znaczeniu *objective*), ale tak je konkretyzują, że podczas planowania od razu określamy, jaka wartość docelowa wskaźnika będzie dla nas satysfakcjonująca. W planowaniu interwencji warto nie tylko podjąć trud sformułowania wizji tego, co się chce zrealizować, lecz także zarysować konkretny, wymierny obraz zmiany, którą zamierza się osiągnąć, wyrażony w kategoriach wartości wskaźników docelowych (*targets*)<sup>7</sup>. Tak sformułowane cele podlegają kontroli wykonania w ewaluacji *ex-post*.

Celem jest – podkreślmy – pewien oczekiwany stan rzeczy, do którego powinny doprowadzić podejmowane działania, a nie samo dążenie, zamiar czy usiłowanie. Oczywiście, projektodawcy będą proponować pewne działania, jednakże wszystkie osoby zaangażowane w program powinny mieć na uwadze cele tych działań i to te cele należy uważać za naczelną punkt odniesienia, którym jest pożądaną przyszły stan rzeczywistości. Rzeczywistość określamy przy pomocy jej wymiarów/charakterystyk. Cele można opisać ze względu na jeden wymiar lub wiele wymiarów: mogą być one proste lub złożone. Cel złożony daje się faktycznie ująć jako kombinacja celów ustanowionych dla poszczególnych jego wymiarów. Ponadto, osiągnięcie pewnych celów może wymagać zrealizowania innych działań, zorientowanych na właściwe sobie cele instrumentalne wobec celu nadrzędnego, co prowadzi do tzw. drzewa celów.

W przypadku programów i projektów, finansowanych w oparciu o fundusze strukturalne, nieraz zauważamy problemy z właściwym ustaleniem wymaganych przez metodykę Komisji Europejskiej wskaźników: produktu, rezultatu i wpływu. Trudności wiążą się tu nie tylko z zapewnieniem odpowiednich, z metodologicznego punktu widzenia, właściwości wskaźników, lecz także z rozróżnieniem produktów, rezultatów i wpływu na poziomie definicji celów programu czy projektu. Błędy i wątpliwości w ustalaniu wskaź-

<sup>7</sup> Atkinson A.B., Marlier E., Nolan B., „Indicators and targets for social inclusion in the UE”, *Journal of Common Market Studies* 2004, vol. 42, nr 1, s. 47–75.

ników, wyrażonych w kategoriach tych trzech rodzajów efektów, obnażają zwykle chaos w procesie planowania i brak jasności co do wyników, jakie ma przynieść planowana interwencja publiczna. Podstawą do ustalenia dobrych wskaźników jest przede wszystkim jasne określenie:

- co ma być dostarczone przez interwencję, na co zostaną zużyte środki w toku zaplanowanych działań – czyli zdefiniowanie produktów,
- jakie zmiany w sytuacji lub cechach beneficjentów interwencji mają zostać osiągnięte bezpośrednio w wyniku podjętych działań – czyli zdefiniowanie rezultatów,
- jak taka zmiana sytuacji lub cech beneficjentów bezpośrednich wpłynie na ich kondycję w dłuższym okresie czasu – czyli ustalenie, jaki jest planowany wpływ lokalny w terminologii przyjętej przez Komisję Europejską, który należałoby raczej określić mianem trwałego rezultatu, w odróżnieniu od rezultatu bezpośredniego;
- jak dana interwencja wpłynie na rozwiązanie określonego problemu społecznego, do którego jest skierowana – czyli definicja wpływu globalnego w zakresie takich kwestii, jak np. rozwiązanie problemu bezrobocia, ograniczenie przestrzennej mobilności pracowników, brak inwestycji w produkcję innowacyjną, niewielka liczba patentowanych wynalazków, zbyt długi okres oczekiwania na wyrok sądowy czy jego egzekucję, wykluczenie społeczne określonych grup itp.

Interwencje publiczne podejmuje się po to, by w sposób możliwie trwałe rozwiązywać problemy społeczne, dlatego kluczowe jest ustalenie ich celów w kategoriach oczekiwanego wpływu globalnego. Programy publiczne mają służyć celom ogólnym, ujmowanym w postaci oczekiwanych zmian w obszarze zidentyfikowanym jako problem społeczny, czyli w kategoriach wpływu. Programy operacyjne to na ogół konglomeraty tematycznie zbliżonych programów zwanych priorytetami, z których każdy skierowany jest na określony problem lub problemy społeczne. Przy tak strategicznym charakterze priorytetów także i one powinny mieć ustanowione cele w kategoriach wpływu globalnego. Planowane działania, będące środkami do osiągnięcia celów, powinny przede wszystkim zostać ukierunkowane na dostarczenie określonych produktów i wywołanie bezpośrednich zmian po stronie beneficjentów, a zatem cele trzeba tu ujmować właśnie w tych kategoriach.

Różne działania podejmowane w obszarze priorytetu prowadzą nieraz do zbliżonych rezultatów, które można, w związku z tym, w sposób zagregowany ująć jako szczegółowe cele bezpośrednie. Wartość docelową takich celów szczegółowych wyrażałoby się jako poziom rezultatów zsumowanych na poziomie priorytetów. Posłużenie się wskaźnikami rezultatu ułatwia monitoring wdrożenia całego priorytetu, zwłaszcza wtedy, gdy oczekujemy stopniowego pojawiania się rezultatów w toku realizacji.

W przypadku niektórych programów także produkty poszczególnych działań (lub ich suma na poziomie programu czy priorytetu) mogą mieć szczególne

znaczenie dla osiągnięcia celów programu (np. kilometry wybudowanych dróg różnych kategorii) – wtedy cele warto przedstawić także w kategoriach wskaźników produktu. Ważne jest przy tym, by nie utracić z pola widzenia hierarchii celów: kluczowy jest wpływ interwencji, do którego powinna prowadzić realizacja pośrednich celów w postaci rezultatów, a instrumentem uzyskania tych rezultatów ma być dostarczenie określonej wielkości produktów podejmowanych działań. W monitoringu i bieżącym zarządzaniu programami kierujemy się przede wszystkim wskaźnikami produktu i rezultatu (śledząc także zaangażowanie zasobów), natomiast w ewaluacji *ex-post* zawsze powinno się podejmować próbę zbadania lokalnego wpływu programu, czyli długotrwałych rezultatów interwencji, a w miarę możliwości także wpływu globalnego.

Ta logika obowiązuje także w przypadku projektów, które są prostsze od programów, w ramach których są planowane. Projekty powinny ostatecznie służyć rozwiązaniu problemów społecznych, ale trzeba, by czuwały nad tym jednostki kierujące wdrożeniem programu, tak selekcjonując projekty, żeby ich produkty i bezpośrednie rezultaty dawały szansę na jak największy stopień osiągnięcia planowanych wpływów programu. Projekty jako operacyjne jednostki interwencji publicznej muszą wpisywać się w określone w programie cele nadrzędne ujęte w kategoriach wpływu (co powinny zapewnić specyfikacje kryteriów dotyczących wyboru projektów), lecz w zakresie swoich celów koncentrują się na dostarczeniu produktów i bezpośrednich rezultatów na poziomie beneficjentów. Na ogół zatem nie ma potrzeby, by w dokumentacji projektów określano wskaźniki wpływu projektu, natomiast powinny zostać określone wskaźniki rezultatu i produktu.

### **Zadania ewaluacji *ex-post* a formułowanie celów**

W ewaluacji nie tylko mierzymy stopień osiągnięcia wartości docelowych, lecz także powinniśmy, w miarę możliwości, ustalić stopień, w jakim osiągnięcie to jest rzeczywiście wynikiem oddziaływania interwencji, a nie efektem procesów spontanicznych (nie kontrolowanych przez decydenta). Takie zadanie często bardzo trudno wykonać – jeżeli jest to w ogóle możliwe. Dlatego można sformułować minimalne zadanie ewaluacji jako badanie, czy zmiana zachodzi w pożądanym kierunku. Ewaluacja ma też inne cele, w tym dokonanie oceny procesu wdrażania, aby usprawnić działania przyszłe; jednak omówienie wszystkich zadań ewaluacji *ex-post* wykracza poza ramy tego artykułu. Z dotychczasowych rozważań wynika, że dla ewaluacji *ex-post* ustanowienie właściwego zestawu wskaźników operacjonalizujących cel interwencji i umożliwiających pomiar uzyskanej zmiany jest kwestią kluczową.

W przypadku ewaluacji *ex-post* ważne jest sprawdzenie efektów netto programu czy projektu, czyli tego, jaką przyniósł on wartość dodaną, która by się nie pojawiła, gdyby nie było interwencji. Na przykład, czy przeprowadzone

projekty szkoleniowe przyniosły w efekcie taką trwałą zmianę sytuacji na rynku pracy beneficjentów, która nie pojawiłaby się, gdyby nie podjęto działania (albo nadwyżkę ponad to, co stanowiłoby poprawę wynikającą z samodzielnego rozwoju sytuacji rynkowej)? Czy udostępnienie dotacji określone przedsiębiorstwu, działającemu na lokalnym rynku, nie spowodowało wyparcia przez nie konkurencji i w rezultacie spadku, a nie wzrostu zatrudnienia czy poziomu aktywności gospodarczej na tym rynku? W ewaluacji *ex-post* stawia się pytanie: na ile słuszne jest założenie o tym, że dane działania poprzez swoje produkty i rezultaty pozwalają rozwiązać określone problemy społeczne? W sytuacji, gdy odpowiedź na takie pytania jest bardzo trudna lub niemożliwa, minimalnym wymogiem stawianym ewaluacji *ex-post* powinno być ustalenie trwałego rezultatu netto (zmiany po stronie beneficjentów/układów docelowych pomniejszonej o oszacowaną wartość zmiany, która miałyby miejsce w przypadku braku interwencji), określanego także mianem „wpływu lokalnego”. Odpowiedź na takie pytania pozwala zweryfikować przyjętą jawnie bądź *implicite* teorię programu – zbiór przekonań o zależnościach przyczynowych pomiędzy podejmowanymi interwencjami a celami, którym służą. W ewaluacji musi zostać zatem postawione zagadnienie celów programu, ujętych w kategoriach wpływu. Oczywiście, ewaluator pyta także o mechanizmy wdrożenia, a więc będzie się interesować również tym, jakie mechanizmy utrudniały uzyskanie produktów i rezultatów interwencji, a jakie temu uzyskaniu sprzyjały. Wszystko po to, aby w przyszłości skorygować teorię wdrażania programu, czyli zbiór założeń dotyczących tego, w jaki sposób najskuteczniej i najefektywniej można osiągnąć zamierzone efekty.

Zapewnienie jasności w zakresie celów programów i projektów jest punktem wyjścia do określenia sposobu ich pomiaru, czyli ustalenia wskaźników. Tymczasem w dokumentach programowych nierzadko miewamy do czynienia z problemem właściwego ujęcia celów w kategoriach wpływu, w odróżnieniu od rezultatu czy produktu. Ponadto, cele bywają mylone z aktywnością, dążeniami czy usiłowaniami, które mają do nich prowadzić (w określeniach celów trafiają się takie sformułowania, jak: „tworzenie warunków”, „podejmowanie wysiłków”, „podejmowanie działań” itp.). Wszystko to bardzo utrudnia prowadzenie ewaluacji *ex-post*, a w sytuacji braku adekwatnych wskaźników – ich trafny dobór dla potrzeb ewaluacji (taki dobór wskaźników po zakończeniu programu jest sam w sobie tylko protezą w stosunku do prawidłowego ustanawiania wskaźników na etapie projektowania).

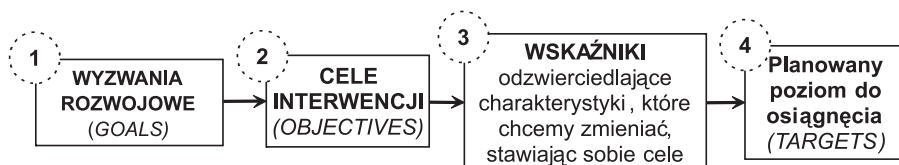
### Cele–wyzwania i cele–zadania

W programach cele powinny być formułowane na różnym poziomie ogólności:

- jako cele-wyzwania: na ogół pojedyncze lub nieliczne i często wieloaspektowo ujęte jakościowe wizje, cele ogólne (*goals*),

- jako cele-zadania: ujęte w sposób konkretny, mierzalny i określony w czasie uszczegółowienie celów interwencji (*objectives*).

Konkretyzację celów i ich mierzalność zapewnia określenie ich wskaźników, czyli poddających się pomiarowi zmiennych reprezentujących kluczowe wymiary celu, i przyjęcie planowanych wartości tych wskaźników, czyli wartości docelowych (*targets*), w określonym czasie.



Rysunek 1. Wyzwania, cele interwencji i planowany poziom realizacji celów a wskaźniki

Źródło: Opracowanie własne.

W ramach tej logiki istnieje niebezpieczeństwo utożsamienia konieczności określenia wskaźników do celów interwencji (wyrażonych poprzez określone zmienne) z podaniem planowanego poziomu osiągnięcia celu–*targets* (wartość na określonej zmiennej). Zdefiniowanie wskaźnika (zmiennej) jest warunkiem dla możliwości podania jego wartości docelowej; podanie tej wartości nie jest jednak konieczne dla poprawności formułowania siatki celów, a raczej pożądane z punktu widzenia możliwości „rozliczenia” realizacji programu czy konkretnego projektu.

Wymaga się, by cele-zadania, identyfikowane angielskim terminem *objective*, spełniały kryteria określane jako *SMART*, a więc były:

- *specific* – konkretne (co chcemy osiągnąć),
- *measurable* – mierzalne (trzeba umieć zmierzyć to, co się chce osiągnąć),
- *achievable* – osiągalne,
- *realistic* – realistyczne przy istniejących zasobach
- *time-bounded* – określone w czasie

Już na pierwszy rzut oka widać, że aby cel (*objective*) mógł być *SMART*, powinien zostać opisany przy pomocy mierzalnych zmiennych pozwalających ustalić wartość określonego stanu rzeczy, czyli wskaźników. Z drugiej strony, nie ma mowy o dobrych wskaźnikach, jeśli cele nie są ustalone w sposób nie tylko merytorycznie sensowny, lecz także formalnie poprawny.

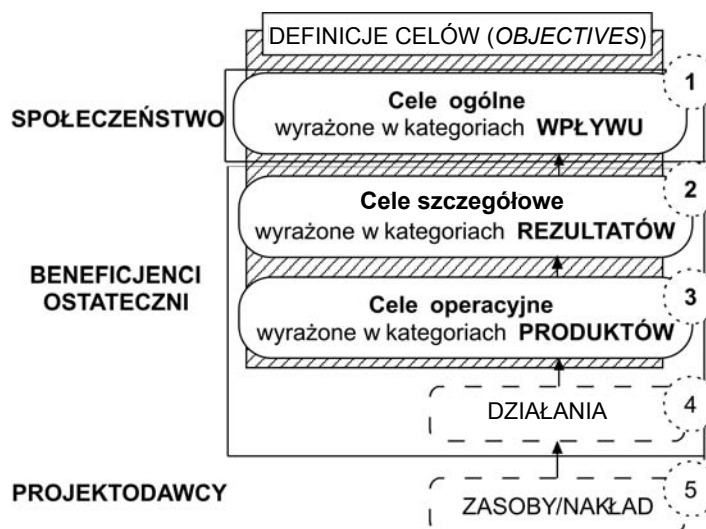
### **Wpływ społeczny a „wykonanie planu”: podział celów interwencji (*objectives*)**

Drugą zasadniczą logiką (nazwijmy ją – „pionową”) przyjmowaną przy okazji formułowania programów interwencji publicznych jest powiązanie

ogólnych celów interwencji na poziomie ogólnospołecznym z celami poszczególnych projektów na najniższym poziomie dotyczącym beneficjentów. W ramach logiki „pionowej” (1) cele ogólne formułowane są w kategoriach wpływu i dotyczą poziomu ogólnospołecznego. To są właściwe punkty zainteresowania polityki społecznej. W praktyce, programy wdrażane w ramach polityki społecznej obejmują tylko część społeczeństwa, grupy tzw. beneficjentów. (2) Cele szczegółowe formułowane w kategoriach rezultatów oraz (3) cele operacyjne wyrażane w kategoriach produktów dotyczą już właśnie owych beneficjentów programu.

Podążając ścieżką logiki wdrożenia: realizatorzy interwencji podejmują pewne (4) działania (które (5) wiążą się z kosztami). Wykonanie tych działań ma docelowo mieć charakter (3) produktu. Produktem może być np. przeprowadzenie określonego szkolenia, które kończy się dla beneficjentów uzyskaniem danego zaświadczenia lub certyfikatu. Produkt sam w sobie jest ważnym elementem projektowania działań, ale cele operacyjne (produkty) są tylko narzędziem służącym do osiągnięcia celów szczegółowych (rezultatów) polegających na wywołaniu pożądanej zmiany po stronie beneficjenta. (2) Rezultat definiuje się jako zmianę dotyczącą beneficjentów, wywołaną przez interwencję. To, co stanowi tę zmianę, czyli rezultat, powinno wynikać z odpowiednio zapisanego celu. Jeśli celem jest doprowadzenie do zatrudnienia beneficjentów warunkowane ukończeniem szkolenia, to wówczas wskaźnikiem rezultatu będzie liczba (proporcja) beneficjentów, którym udało się to osiągnąć. Rezultatem może być także samo uzyskanie certyfikatu, np. prawa jazdy, jeśli zostanie uznane za docelową zmianę po stronie beneficjentów (produktem będzie w tym przypadku dostarczenie odpowiedniego kursu i jazd szkoleniowych, co można nawet uzupełnić kryterium zaliczonego egzaminu wewnętrznego). Certyfikat oznacza bowiem wtedy zewnętrznie potwierdzoną zmianę w kwalifikacjach, która, przynosząc istotne uprawnienia, sama w sobie zmienia sytuację beneficjenta. To, czy ograniczymy się do tak rozumianego rezultatu, zależy od sposobu sformułowania celu. Rezultat zawsze oznacza ujętą na poziomie celu, oczekiwaną zmianę po stronie beneficjentów wywołaną interwencją.

**Rezultaty na poziomie beneficjentów ostatecznych mają przyczynić się do wpływu programu na jakieś cechy całej zbiorowości.** Przypuśćmy, że projekt, w którym zdobywano kwalifikacje zawodowe, realizowano w ramach programu operacyjnego Kapitał Ludzki. Celami strategicznymi, będącymi punktem odniesienia programu, były np. ustalenia strategii lizbońskiej – zwiększanie poziomu zatrudnienia osób w wieku produkcyjnym. Wówczas chcielibyśmy wiedzieć, czy produkt w postaci szkolenia przyczynił się do realizacji (1) celu ogólnego (wyrażanego w kategoriach wpływu) za pośrednictwem zmiany w sytuacji czy cechach beneficjentów ujmowanych jako (2) rezultaty. W naszym prostym przykładzie rezultatem może być fakt, iż dana osoba była w stanie uzyskać certyfikat oraz że nowa umiejętność przyczyniła się do znalezienia przez tę osobę pracy.



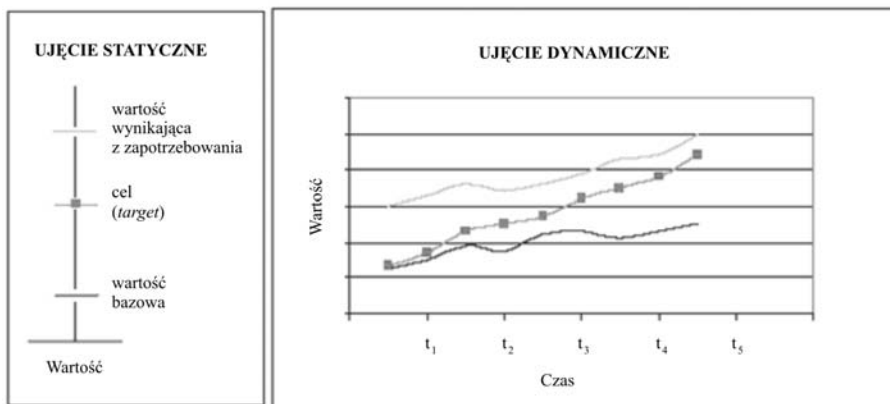
Rysunek 2. Cele interwencji (*objectives*): ogólne, szczegółowe i operacyjne wyrażone w kategoriach wpływu, rezultatów i produktów

Źródło: Opracowanie własne.

Odróżnienie (2) rezultatów od (1) wpływu pozwala (przy pomocy ustalonych wskaźników) prześledzić oddziaływanie zmian dotyczących (2) beneficjentów na zmianę (1) ogólnej sytuacji społecznej. Może się bowiem zdarzyć, że nasi beneficjenci zostaną przeszkoleni (produkt), uzyskają w wyniku egzaminu certyfikaty, a w konsekwencji – zatrudnienie (rezultat), ale ich zatrudnienie spowoduje zwolnienie innych osób, a nawet większą ilość zwolnień z pracy niż przyjęć. W takiej sytuacji cele ogólne (wpływ) NIE zostały osiągnięte i, mimo sukcesów na niższych poziomach (produktu, rezultatu), programu nie można uznać za udany z punktu widzenia celów ogólnych. W efekcie naruszenia warunków konkurencji rynkowej jedna grupa uzyska przywileje (kwalifikacje bez ponoszenia pełnych kosztów i zatrudnienie) kosztem innej grupy, przy braku efektu społecznego. Tego rodzaju posunięcie da się czasami wyjaśnić pewnymi względami społecznymi, ale nie jest to łatwe. Ponadto, zawsze użycie środków na jeden cel wyklucza wykorzystanie ich na inny. Oczywiście, przedstawiony przykład jest bardzo dużym uproszczeniem, pozwala jednak prześledzić logikę hierarchii celów, która staje się podstawą ewaluacji *ex-post*. Warto dodać, że pierwszym wyzwaniem pozostaje na obecnym etapie ciągle jeszcze właściwy pomiar rezultatów i ustalenie stopnia, w jakim są one efektem interwencji.

**Problem: czy do wszystkich poziomów celów interwencji (*objectives*) określać konkretną wartość do osiągnięcia (*targets*)?**

Wróćmy do zasadniczego pytania, zasygnalizowanego już wcześniej: na których poziomach programu interwencji trzeba ustalać cele poprzez wartości wskaźników do osiągnięcia (*targets*) i do kogo miałyby należeć to zadanie? Ustalając cel, zadajemy sobie pytanie, jaki jest obecny stan danego problemu (diagnoza sytuacji) i co chcemy osiągnąć (stan przyszły, czyli cel). Stawiając sobie cele, chcemy zmienić sytuację. Ewaluacja ma nam odpowiedzieć, co zadziałało, które cele (rezultaty i wpływ) zrealizowaliśmy, które cele (produkty) przyczyniły się do osiągnięcia rezultatów i wpływu. Jednak, aby można było tego dokonać, cele winny być „skwantyfikowane”: musimy umieć je zmierzyć, musimy zidentyfikować zmienną (cechę/właściwość/charakterystykę, która może dla różnych obiektów analizy przyjmować co najmniej dwie różne wartości) informującą nas o tym, na jakim etapie w danym momencie czasowym jesteśmy. Zmianę społeczną, jakiej chcemy dokonać przez osiągnięcie celu, można wyrazić tylko kierunkowo, jednak powinno się uściślić jej zakres ustalając wartości docelowe wskaźników (*target*)<sup>8</sup>.



Rysunek 3. Koncepcja statyczna i dynamiczna formułowania celów poprzez wskaźniki

Źródło: Opracowanie własne, niewielka modyfikacja na podstawie: European Commission, *Indicators For Monitoring And Evaluation...*, op.cit., s. 15.

Rys. 3 ukazuje dwa podejścia do określania celów poprzez wartości docelowe wskaźnika (*targets*). Wartość docelową (*target*) można ustalić w odnie-

<sup>8</sup> Na przykład, zależy nam na zmniejszeniu poziomu ubóstwa wśród dzieci. Taki cel nie jest jeszcze skwantyfikowany, musimy bowiem wybrać wskaźnik ubóstwa najlepiej oddający nasze rozumienie ubóstwa. Może to być np. tzw. ubóstwo absolutne. Wówczas osiągnięcie celu będzie oznaczało zbliżenie wartości wskaźnika ubóstwa absolutnego wśród dzieci (która wyjściowo była wyższa) do wartości dla całego społeczeństwa (*benchmarking*) albo do określonej wartości (np. z 20% do 0%), gdy np. przyjmujemy założenie, że ubóstwo wśród dzieci w ogóle nie powinno mieć miejsca.



sieniu do wartości bazowych, uzyskanych w oparciu o (1) ekstrapolację istniejących trendów, albo (2) przyjmując jakąś wartość wzorcową (*benchmark*)<sup>9</sup>. W pierwszym przypadku wykorzystujemy koncepcję dynamicznego ujęcia zaprezentowanego na rys. 3. Wartość docelową zakładamy na ogół jako poprawę (w sensie normatywnym) wartości w stosunku do bazowej prognozy (czy ma to być wartość większa czy mniejsza, zależy od normatywnej interpretacji wskaźnika). Natomiast, w przypadku ustalania wartości wzorcowej, stosuje się rozmaite podejścia, od opartych o kryteria normatywne, po standardy zaczerpnięte z wzorcowych grup odniesienia lub wyznaczone kryteria statystyczne (np. średnią dla szerszego układu społecznego). Oczywiście, kwestia wyznaczania ściśle skwantyfikowanych wartości docelowych jako kryterium osiągnięcia celu programu jest kwestią dyskusyjną, szczególnie jeśli chodzi o wskaźniki wpływu, a nawet rezultatu, w związku z ogólniejszym zagadnieniem przewidywalności zjawisk społecznych. Na pewno nadają się do tego produkty i dlatego kontrole wykonania szczególnie skrupulatnie zajmują się właśnie wskaźnikami produktów. Rezultaty są wynikiem powiązania wielu okoliczności i dlatego ich wartość docelowa będzie zawsze tylko pewnym przybliżeniem tego, co faktycznie zostanie osiągnięte. Ewaluacja powinna uwzględniać ten fakt i poszukiwać odpowiedzi na temat przyczyn odchyień w pomiarach, a nie tylko konstatować fakt rozbieżności (w którąkolwiek stronę). Dla poziomu wartości docelowych wskaźników wpływu, zwłaszcza makroekonomicznych, duże znaczenie mają rozmaite zależności bilansowe, które trzeba brać pod uwagę. Także tu nie należy się spodziewać możliwości bezbłędnego przewidywania przyszłości po interwencji.

### **Wskaźniki w pomiarze użyteczności programu** – rola diagnozy społecznej

Istotnym kryterium w ewaluacji *ex-post* jest użyteczność programu (projektu), czyli stopień, w jakim program zaspokaja istotne potrzeby grup docelowych lub przyczynia się do rozwiązania problemów, niezależnie od tego, czy były one zasygnalizowane przez cele programu. Aby móc ustalić użyteczność programu konieczne jest odniesienie go do szerszego kontekstu społecznego. Taki kontekst powinien być dany już w diagnozie społecznej, która stanowi integralną część programu. W diagnozie społecznej posługujemy się wskaźnikami społecznymi do opisu sytuacji i zidentyfikowania problemów. Diagnoza powinna posługiwać się wyraźnymi ocenami opisywanych zjawisk: oczekujemy w niej określenia, czego jest za dużo, czego jest za mało, a czego w sam raz. W diagnozie należy także ujmować potencjalne konsekwencje podjęcia i niepodjęcia określonych działań. Opis rzeczywistości w części dia-

<sup>9</sup> European Commission, *Indicators For Monitoring And Evaluation...*, op.cit., s. 13.

gnostycznej jest dokonywany także w postaci wskaźników, gdyż tylko przy pomocy wskaźników można w kategoriach ilościowych scharakteryzować analizowane zjawiska. Wskaźniki te nie odnoszą się wprost do celów interwencji, dlatego określamy je mianem wskaźników kontekstowych. Trafny dobór wskaźników kontekstowych do opisu rzeczywistości społecznej stwarza możliwość dostrzeżenia istotnych problemów społecznych, które mogą być zasygnalizowane przez program, a także ograniczeń, wobec których może stanąć wdrażanie.

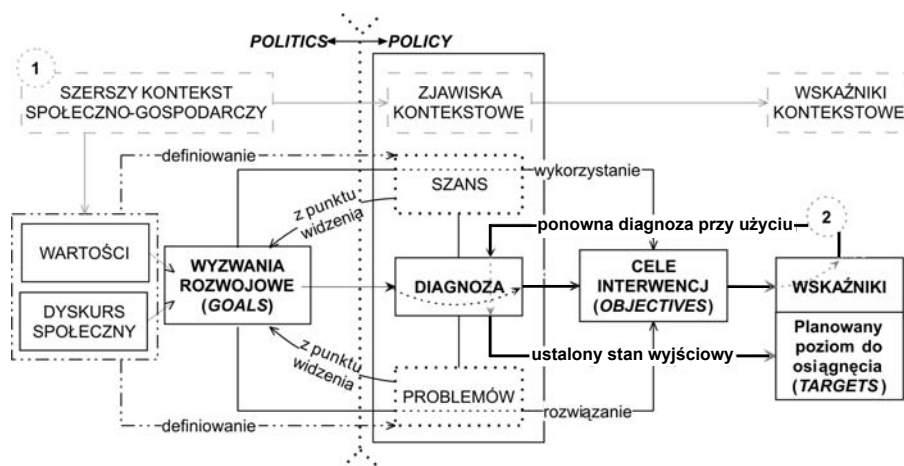
Identyfikacji problemów służy – oprócz oceny względem wartości wzorcowych ustalonych teoretycznie lub politycznie – porównanie wartości zmiennych kontekstowych: (1) w danym kraju (regionie) lub w różnych krajach (regionach) w przeciągu kilku lat, czyli analiza szeregów czasowych, (2) w danym kraju (regionie) do średniej/przeciętnej dla różnych krajów (regionów) (3) oraz w danym kraju (regionie) w podziale na wyróżnione segmenty (np. ze względu na płeć czy wiek), co służy m.in. identyfikacji grup ryzyka.

W ewaluacji *ex-post* ustalenie wartości wskaźników kontekstowych po interwencji i porównanie ich z wartościami wyjściowymi daje możliwość oceny użyteczności programu. Warto też, w toku wdrażania programu, monitorować pewne aspekty sytuacji społecznej, które mogą wpływać na przebieg i efekty programu, jak np.: migracje, przemiany demograficzne itp.

Wskazanie przy okazji diagnozy na wskaźniki kontekstowe nie powinno odwracać naszej uwagi od wskaźników związanych bezpośrednio z celami programu, zwłaszcza od wskaźników wpływu. Diagnoza musi powstawać w toku procesu iteracyjnego: w oparciu o debatę wokół podstawowych kwestii społecznych i wstępną wizję strategicznych celów określonej polityki przystępujemy do konstrukcji ram diagnozy (w tym wskaźników), wykonujemy diagnozę, korygujemy definicję sytuacji i koncepcje celów, wracamy do diagnozy w celu dokonania stosownych uzupełnień itd. Ostatecznie interwencje publiczne są pewnymi wizjami dotyczącymi wagi określonych problemów społecznych i koncepcjami ich skutecznego rozwiązania. Ich kreatywny charakter powoduje, że, z jednej strony, ich kształt nie wynika nigdy samistnie z diagnozy, z drugiej zaś – w związku z niepewnością – wymagają, przynajmniej w warunkach demokracji, przyjęcia odpowiedzialności wobec zaufania społecznego. Diagnoza społeczna powinna służyć temu, by zarówno odpowiedzialność wobec społeczeństwa, jak i zaufanie społeczne sprostały wyzwaniom niepewności.

Zatem, obok analizy przy pomocy (1) „wskaźników kontekstowych” (zob. kolor pomarańczowy na rys. 4), należy po sformułowaniu celów interwencji i określeniu wskaźników tych celów (2) ponownie dokonać diagnozy, tym razem przy użyciu tych zmiennych, które będą wykorzystywane we właściwej ewaluacji. Pomaga to zweryfikować wskaźniki przyjęte dla celów, lepiej zrozumieć same wskaźniki oraz strukturę celów (wyrażoną teraz w postaci mierzalnej, przy pomocy wskaźników; zob. pogrubiona linia na rys. 4).

W ewaluacji *ex-post* powracamy do wskaźników kontekstowych i wskaźników celu wykorzystywanych w fazie diagnozy.



Rysunek 4. Związek między procesem tworzenia polityk publicznych a wskaźnikami

Źródło: Opracowanie własne.

## Ocena przyjętych wskaźników

W toku dobierania wskaźników, a także już po włączeniu ich w proponowany zestaw, obejmujący wszystkie kluczowe aspekty programu, musimy poddać je ocenie. Zwykle dokonuje się tego na etapie ewaluacji *ex-ante*, a weryfikuje w toku pierwszej ewaluacji śródkresowej. Przy ocenie posługujemy się szeregiem kryteriów, które jednocześnie pełnią rolę przewodnika na etapie ustanawiania wskaźników. Ze względu na to, że wskaźniki powinny zostać przyjęte na etapie konstruowania programu, ewaluator *ex-post* powinien mieć w zasadzie do dyspozycji komplet wskaźników, którymi będzie się posługiwał – będzie tak oczywiście w przypadku programu dobrze skonstruowanego. Jeśli ewaluator ma ocenić program, który nie został zaopatrzony we właściwie skonstruowany zestaw wskaźników, musi – w uzgodnieniu z jednostką zlecającą ewaluację – samodzielnie ustalić wskaźniki i w ten sposób „zrekonstruować” operacyjny sens programu.

Kluczowym zagadnieniem przesadzającym o wartości wskaźnika jest oczywiście trafność, czyli zdolność wskaźnika do dostarczania miarodajnej wiedzy o tym właśnie aspekcie zagadnienia, dla którego został ustanowiony. Umiejętne uchwycenie istoty rzeczy nie wyczerpuje jednak kryterium adekwatności wskaźnika. Już w przypadku określenia celów interwencji wspominaliśmy o normatywnym odniesieniu wskaźników używanych dla reprezentacji zmian, które w ramach prowadzenia polityk muszą być monitorowane.

Cele interwencji *explicite* lub *implicite* wyrażają założenia dotyczące wartości określonego zestawu i hierarchii celów. Wskaźniki winny odzwierciedlać ową normatywną wizję ze względu na to, że są jedynie narzędziem pomiaru. Niemniej jednak cele interwencji zyskują swe ostateczne znaczenie poprzez ustalenie wskaźników pozwalających owe cele sprecyzować, a nawet skwantyfikować, stąd problem uwzględnienia wartości dotyczy nie tylko formułowania celów, ale i wskaźników.

Normatywna interpretacja wskaźników oznacza stwierdzenie, czy jest „dobrze”, czy „źle”, i czy – w wymiarze czasowym – zmierzamy w „pożądanym” kierunku. Wskaźniki społeczne muszą mieć jasne (zrozumiałe w sposób intuicyjny) i akceptowalne (przez ogół społeczeństwa) konotacje normatywne<sup>10</sup>. Nie są „zawieszane w próżni”, zwykle odnoszą się do całego zespołu założeń i wartości wyrażanych poprzez ideologie czy doktryny; owe zespoły mogą mieć natomiast charakter eklektyczny. Dobrze, by założenia aksjologiczno-normatywne zostały wyrażone wprost przy ustanawianiu wskaźników, gdyż ułatwia to poprawny wybór wskaźników, ogranicza manipulację oraz zwiększa obiektywizm i transparentność ocen.

Uwagi te dotyczą zwłaszcza wskaźników wpływu związanych z najbardziej ogólnym poziomem celów; warunkują one cele (wskaźniki) niższego rzędu. Wskaźniki wchodzące w skład całego zestawu wskaźników społecznych powinny być równomiernie rozłożone pomiędzy wszystkie istotne aspekty ważne dla społeczeństwa (dotyczące warunków rozwoju lub problemów społecznych), a tak skomponowany zestaw powinien być zrozumiały dla członków społeczeństwa, którego sytuacji dotyczy<sup>11</sup>.

Istotną kwestią jest też liczebność zestawu wskaźników: im jest ich więcej, tym lepiej odzwierciedlają badane zjawisko, ale też, z drugiej strony, im mamy ich mniej, tym bardziej przejrzyste dają wyniki. „Nie możemy oczekiwać [od wskaźników], że będą w zupełności oddawały stan społeczeństwa. [...] Natura [...] wnioskowań [na ich podstawie] zależeć będzie od decyzji dotyczących uwzględnianych definicji oraz od danych. Różne wskaźniki uwydatniają inne cechy problemów społecznych i sugerują odmienne priorytety dla interwencji politycznych”<sup>12</sup>. Nie wolno nam zaniedbywać tego, w jaki sposób ideologie, wyznawane przez badacza wartości, wartości ogólnospołeczne czy układ polityczny wpływają na rezultaty badań; należy ujawniać te wartości w prezentacji wskaźników, nie powinny one jednak wpływać na samo ustalanie stanu faktycznego.

Oceny poszczególnych wskaźników i kompozycji ich pełnego zestawu dokonuje się na ogół na podstawie szerszego zestawu kryteriów. Wskaźniki źle ocenione w oparciu o takie kryteria powinny być skorygowane lub zastąpione

<sup>10</sup> Zob. Atkinson T., Cantillon B., Marlier E., Nolan B., *Social Indicators. The EU Social Inclusion*, Oxford University Press, New York 2002, s. 19, 21.

<sup>11</sup> Zob. *ibidem*, s. 24–25.

<sup>12</sup> *Ibidem*, s. 19–20.

innymi. Kryteriami oceny posługujemy się też faktycznie w toku całej pracy nad wskaźnikami, toteż ocena ma raczej charakter ciągłego przedsięwzięcia, a nie jednorazowej akcji. W zasadzie, dobrze jest myśleć o kryteriach już w trakcie ustanawiania wskaźników. Przechodząc do charakterystyki przykładowych zestawów kryteriów jakości wskaźników, podkreślamy raz jeszcze, że priorytetem jest zawsze kryterium trafności: odpowiedniości teoretycznej i normatywnej względem ustanowionych celów. Kryterium trafności występuje w zestawach kryteriów oceny wskaźników, choć czasami w nieco zawoalowanej postaci, jak np.: „effective indicators require a clear conceptual basis” („skuteczne wskaźniki wymagają jasnej bazy pojęciowej”)<sup>13</sup>.

Kryteria jakości wskaźników dzielimy na takie, które dotyczą pojedynczego wskaźnika i na dotyczące całego systemu wskaźników. Przykład zestawu kryteriów oceny, zaczerpnięty z opracowania Rady Europy poświęconego spójności społecznej zawiera tab. 1<sup>14</sup>. W ramach przedstawionego schematu przyjęto cztery aspekty oceny jakości wskaźników polityk prowadzonych w analizowanej domenie publicznej:

- czy dają to, czego się od nich oczekuje (czy są adekwatne do założonych celów),
- jaki jest ich specyficzny (szczególny) wkład,
- czy są łatwe w użyciu (z punktu widzenia użytkownika),
- czy są kosztowne/trudne do uzyskania (koszty uzyskania).

Tabela 1. Przegląd kryteriów jakości

	<b>Adekwatność do założonych celów</b>	<b>Szczególny wkład</b>	<b>Łatwość użycia z punktu widzenia użytkownika</b>	<b>Koszty uzyskania</b>
Pytania	odpowiednie do potrzeb informacyjnych	spójny	klarowne	(nie dotyczy)
Zestaw pytań	kompletny i zbilansowany			
Wskaźniki	reprezentatywne (wobec wskaźnika lub zestawu)	jednoznaczne	o jasnej i akceptowalnej interpretacji	pozyskanie informacji nie może być zbyt kosztowne
Dane	wiarygodne, rzetelne	istotne	niepoddające się manipulacji, porównywalne	dostępne regularnie

Źródło: Opracowanie własne na podstawie: Council of Europe, *Concerted Development...*, op.cit.

Wymienione wyżej aspekty zostały skrzyżowane z czterema elementami składowymi systemu oceny polityk spójności:

<sup>13</sup> Cobb C.W., Rixford C., *Lessons Learned From The History Of Social Indicators*, Redefining Progress, San Francisco 1998, [www.rprogress.org/newpubs/1998/SocIndHist.pdf](http://www.rprogress.org/newpubs/1998/SocIndHist.pdf), s. 16–17.

<sup>14</sup> Council of Europe, *Concerted Development of Social Cohesion Indicators: Methodological Guide*, Council of Europe Publishing, Strasbourg 2005, [www.coe.int/T/E/Social\\_cohesion/Social\\_Policies/GUIDE\\_EN.pdf](http://www.coe.int/T/E/Social_cohesion/Social_Policies/GUIDE_EN.pdf), s. 86.

- poszczególnymi pytaniami, które artykułują potrzeby w zakresie wiedzy,
- zestawem pytań jako całością; od jego kompozycji zależy objęcie analizą wszystkich istotnych aspektów badanej domeny,
- wskaźnikami, które mają dostarczyć odpowiedzi na pytania,
- danymi, które są podstawą obliczenia wartości wskaźników.

Schemat ten słusznie zwraca uwagę na to, że wskaźniki są pochodną stawianych pytań. Jest to podejście szczególnie adekwatne w warunkach konstruowania diagnozy społecznej. Będzie przydatne także w sytuacji (ciągle bardzo częstej), gdy ewaluator *ex-post* nie dysponuje właściwym zestawem wskaźników zawartych w programie lub z różnych powodów musi wyjść poza ten zestaw na zlecenie podmiotu zamawiającego ewaluację. Wtedy jego praca w znacznym stopniu będzie przypominała pracę diagnosty. W sytuacji, gdy program jest zaopatrzone w sensowny zestaw wskaźników, kryteria dotyczące pytań mogą zostać użyte do sprawdzenia, czy przy ustaleniu wskaźników zadbano o to, by odzwierciedlały one wszystkie istotne aspekty ustanowionych celów i pozwalały dokonać oceny ze względu na wszystkie przyjęte kryteria ewaluacji. Kryteria zawarte w poszczególnych komórkach tabeli mają sens heurystyczny i mogą być zaopatrzone albo w dychotomiczną skalę tak/nie, informującą o osiągnięciu pod danym względem poziomu aspiracji, bądź też rozwinięte w szacunkową skalę ocen.

Przedstawiony wyżej przykład kryteriów oceny jest jednym z wielu. Często, przez analogię do wcześniej omówionych kryteriów *SMART*, dla oceny celów programu stosuje się kryteria *SMART* dla wskaźników. Akronim *SMART* jest przy tym rozwijany w rozmaity sposób, stąd też trudno uznać to za jakąś powszechnie uzgodnioną metodykę. Na przykład, litera A oznacza raz *attainable* (jak niżej), innym razem zaś *agreed*, czyli uzgodnione (co mówi o konieczności uzyskania konsensusu w zakresie sposobu pomiaru efektów pomiędzy zaangażowanymi aktorami). Poza tym, samo określenie „mierzalny” bardziej nadaje się do oceny, czy mamy w ogóle do czynienia ze wskaźnikiem, niż do oceny jego jakości: mierzalność jest definicyjną własnością wskaźników – nie ma wskaźników niemierzalnych. Biorąc jednak pod uwagę fakt, że w rozmaitych dokumentach programowych, strategiach czy projektach ciągle jeszcze pojawiają się „niby-wskaźniki”, które nie mogą pełnić roli wskaźników, takie kryterium okazuje się potrzebne. Cały zespół kryteriów ma sens heurystyczny, a akronim ułatwia zapamiętanie tego, co trzeba wziąć pod uwagę zarówno ustalając, jak i oceniając wskaźniki. Kryteria *SMART* mają dość uniwersalny charakter i stosują się do wskaźników z różnego poziomu logiki programowania.

A oto przykład jednego z rozwinięć akronimu *SMART*, według którego dobry wskaźnik powinien być<sup>15</sup>:

---

<sup>15</sup> Sandhu-Rojon R., *Selecting Indicators for Impact Evaluation*, UNDP, Paris 2003, [www.pnud.ne/RENSE/Biblioth%E8que/PNUD14.pdf](http://www.pnud.ne/RENSE/Biblioth%E8que/PNUD14.pdf), s. 11–12.

- 1) **Specific** – konkretny, dokładny:
  - Czy wiadomo dokładnie, co będzie mierzone? Czy został określony odpowiedni poziom dezagregacji?
  - Czy wskaźnik oddaje istotę pożądanego rezultatu? (w innym punkcie dokładnie to samo)
  - Czy pozwala uchwycić różnice pomiędzy obszarami lub kategoriami ludzi?
  - Czy wskaźnik jest na tyle dokładny, by zmierzyć postęp w kierunku rezultatu?
- 2) **Measurable** – mierzalny:
  - Czy jesteśmy w stanie potwierdzić zachodzące zmiany w sposób obiektywny?
  - Czy wskaźnik będzie w stanie ukazać pożądane zmiany?
  - Czy jest niezawodną i jasną miarą rezultatów?
  - Czy jest na tyle czuły, by ukazać zmiany polityki czy programów?
  - Czy decydenci są zgodni co do tego, co dokładnie należy mierzyć?
- 3) **Attainable** – osiągalny:
  - Jakie zmiany są oczekiwane jako rezultaty udzielonej pomocy?
  - Czy rezultaty są realistyczne? Niezbędne jest wiarygodne połączenie pomiędzy produktami (*outputs*) a rezultatami (*outcomes*).
- 4) **Relevant** – istotny, odpowiedni:
  - Czy wskaźnik oddaje istotę pożądanego rezultatu?
  - Czy jest odpowiedni wobec planowanych produktów i rezultatów? Aby określić znaczenie wskaźników, decydenci mogą zidentyfikować grupę docelową, jej potrzeby, oczekiwania i kryteria.
  - Czy wskaźnik jest w sposób przekonujący powiązany ze sferą działalności?
- 5) **Trackable** – rejestrowalny, możliwy do ustalenia:
  - Czy rzeczywiście dostęp do danych zapewnią rozsądne nakłady i czy nie wiąże się on z większymi trudnościami?
  - Czy znamy źródła danych? Decydenci winni ustalić realistyczne zasady, mechanizmy i obowiązki w zakresie gromadzenia danych.
  - Czy wskaźnik monitoruje realizację planu?

Mimo heurystycznej użyteczności kryteriów *SMART* optujemy za nieco precyzyjniej skonstruowaną siatką kryteriów oceny wskaźników, czyli taką, która precyzyjniej i bardziej systematycznie nawiązuje do metodologicznych i pragmatycznych aspektów ich stosowania. Taki katalog kryteriów oceny wskaźników został zawarty w raporcie opracowanym przez zespół kierowany przez prof. Atkinsona dla potrzeb *EU Social Protection Committee*<sup>16</sup>. Skatalogowano w nim zestaw cech, które powinny mieć dobre wskaźniki i zestawy wskaźników dla określonej domeny polityk publicznych. Autorzy

<sup>16</sup> Zespół pracował w składzie: Tony Atkinson, Bea Cantillon, Eric Marlier i Brian Nolan.

podkreślają, że na poziomie strategii i programów kluczowe znaczenie ma uchwycenie, przy pomocy wskaźników, wpływu interwencji publicznej na sytuację społeczną, a nie nakładów czy działań ucieleśniających się w produktach. Szczegółowo rozważane przez autorów cechy dobrych wskaźników wpływu polityki publicznej na obszar życia społecznego doprowadziły do ich „skatalogowania”, którego syntezę prezentujemy poniżej<sup>17</sup>:

- 1) Wskaźniki powinny ujmować istotę problemu oraz posiadać jasną oraz normatywnie akceptowalną interpretację:
  - zadaniem wskaźników nie jest podsumowanie całej rzeczywistości, ale uchwycenie istoty problemu,
  - odbiorca powinien być pewien, jak ma oceniać kierunek zmian reprezentowanych przez dany wskaźnik,
  - wskaźniki powinny być uznawane za sensowne przez opinię publiczną (postulowana partycypacja przy ustanawianiu wskaźników),
  - wskaźniki powinny charakteryzować się intuicyjną trafnością,
  - wskaźniki powinny korespondować z konkretnymi celami polityk i programów,
  - wskaźniki powinny mieć postać pozwalającą na ustanawianie planowanych wartości docelowych i ocenę stopnia ich realizacji.
- 2) Wskaźniki powinny być odporne na zniekształcenia i sprawdzone pod względem statystycznym:
  - wskaźniki powinny być oparte na wiarygodnych danych, które nie zostały poddane arbitralnym korektom,
  - jeśli dane pochodzą z sondaży, to tylko z reprezentatywnych i spełniających najwyższe standardy metodologicznie pod względem doboru próby, konstrukcji kwestionariuszy, realizacji badań, obróbki danych i analizy wyników,
  - należy dążyć do minimalizacji błędu badania, a zwłaszcza do unikania obciążenia wyników błędem systematycznym,
  - powinno się unikać wskaźników, które podlegają silnemu wpływowi nieprzewidywalnych (np. pogoda) i niewyjaśnialnych czynników,
  - należy zachować ostrożność wobec wskaźników, których wartość w znacznym stopniu nie zależy od polityk publicznych, lecz od czynników niezależnych, np. cyklu koniunkturalnego.
- 3) Wskaźnik powinien odzwierciedlać oddziaływanie interwencji publicznych, ale nie może być przedmiotem łatwych manipulacji politycznych:
  - wskaźnik powinien zmieniać wartość, jeśli interwencja przynosi efekty,
  - wskaźniki powinny dać się powiązać z działaniami,
  - wskaźniki nie powinny pozwalać na manifestowanie pozornych efektów (np. proporcję podmiotów objętych świadczeniem można łatwo zwiększyć przyznając na szeroką skalę minimalne świadczenia).

<sup>17</sup> Atkinson T., Cantillon B., Marlier E., Nolan B., *Social Indicators...*, op.cit.



- 4) Wskaźnik powinien być aktualny oraz umożliwiać niezbędne korekty:
  - podstawowe dane makroekonomiczne są na ogół dość aktualne, natomiast inne dane często nie spełniają postulatu wystarczającej aktualności,
  - bardzo ważne jest zbudowanie odpowiedniego potencjału statystycznego w kraju, który będzie zdolny do dostarczania na czas właściwych informacji,
  - jeśli następuje zmiana w definicji celów polityk lub postęp w rozumieniu sensu zjawisk, powinno to znaleźć odzwierciedlenie w ustalaniu wskaźników.
- 5) Pomiar wskaźnika nie powinien być zbyt dużym obciążeniem dla państwa, dla krajowych przedsiębiorstw ani dla obywateli – dlatego potrzebna jest kreatywność w używaniu danych administracyjnych i już dostępnych danych statystycznych.

Ponadto Atkinson i jego współpracownicy sformułowali zasady dotyczące ustalenia zestawu wskaźników:

- 6) Zestaw wskaźników powinien być rozsądny pod względem liczby obejmowanych przez siebie wymiarów, ale musi ujmować możliwie wszystkie kluczowe obszary problemowe danego zagadnienia społecznego.
- 7) Wskaźniki powinny się wzajemnie uzupełniać, a waga każdego wskaźnika w zestawie powinna być jasno określona, pozwalając na odróżnienie wskaźników o kluczowym znaczeniu od tych, które są pomocnicze.
- 8) Zestaw wskaźników powinien być jak najbardziej przejrzysty, zrozumiały i dostępny dla obywateli UE, tak aby opinia publiczna była w stanie zrozumieć znaczenie wskaźników.

Ten zbiór kryteriów jest lepiej dostosowany do domeny polityk publicznych od prezentowanego wcześniej zestawu kryteriów *SMART*. Łatwo jednak zauważymy, że kluczowymi wymiarami oceny są tu nadal klasyczne, metodologiczne kryteria trafności i rzetelności pomiaru. Wskaźniki celów polityk publicznych muszą być nade wszystko poprawnymi wskaźnikami w ogólnometodologicznym znaczeniu tego słowa, o czym często się zapomina. W konsekwencji nie pamięta się też o tym, że wskaźniki ustanawiane są w procesie operacjonalizacji pojęć, czyli koncepcyjnego poszukiwania sposobu wyrażenia badanych pojęć (w tym celów) w postaci ich mierzalnych przejawów. Uświadomienie sobie tego faktu otwiera drzwi do obszernego dorobku metodologii badań społecznych. Jedną z podstawowych konsekwencji takiego wskaźnikowego „oświecenia” powinno być odejście od „dobierania” wskaźników do celów z istniejących kompilacji na rzecz konceptualizacji wskaźników w procesie ustanawiania jasnych i mierzalnych celów działania.

## Zakończenie

Z powyższych rozważań wynika, że wskaźniki są jedynie sposobem – bezpośrednim lub pośrednim – ustalenia stopnia realizacji poszczegól-

nych wymiarów celu realizowanego przedsięwzięcia. Problem z ewaluacją, a nawet szerzej: z całym procesem projektowania i zarządzania programami publicznymi, nie jest rezultatem ubóstwa informacyjnego zarzucanego wskaźnikom ani tego, że nie są one w stanie oddać dobrze jakości programów. To raczej skutek ich oderwania od istoty programu. W mniejszym stopniu problem ten dotyczy pomiaru nakładów, gdyż tu rządzą zasady rachunkowości oraz dobrze opanowana sztuka sprawozdawczości z podejmowanych działań. W większym stopniu natomiast odnosi się to do pomiaru wyników, na czym się skupiliśmy.

Można zgodzić się z wieloma zastrzeżeniami co do pomiaru wyników, formułowanymi przez część specjalistów od ewaluacji, a zwłaszcza przez reprezentantów szkoły konstruktywistycznej, stawiającej na ewaluację jako interwencję społeczną i praktyczną zmianę organizacyjną. Rzeczywiście, często trudno jest dobrze uchwycić wszystkie istotne aspekty ewaluowanego programu przy pomocy wskaźników ilościowych. Dlatego trzeba być otwartym na multiplizm – wykorzystywanie wielu metod badawczych w celu uzyskania wszechstronnego i pogłębionego oglądu ewaluowanych programów<sup>18</sup>. Niemniej krytyka pomiaru wyników przy pomocy wskaźników jako istotnego składnika ewaluacji, spotykana w literaturze i wśród praktyków, nie powinna odwracać uwagi od istoty problemu – niskiej jakości projektowania polityk i programów publicznych. W Polsce konieczne jest położenie nacisku na pomiar wyników, a zwłaszcza na formułowanie celów strategii, programów i projektów w sposób jasny, możliwy do ujęcia przy pomocy zestawu wskaźników, pozwalających ustalić, czy lub w jakim stopniu cel programu został osiągnięty. To dyscyplinuje tworzących projekty i decydentów, pozwala ocenić racjonalność alokacji zasobów do programu i porównać efekty z innymi projektami.

Dobrze zdefiniowane i wyrażone poprzez wskaźniki w postaci mierzalnej cele strategii programów i projektów są w praktyce warunkiem przeprowadzenia sensownej ewaluacji *ex-post*, gdyż jednym z jej podstawowych zadań jest ocena tego, czy i jak zrealizowane zostały cele programu. Oczywiście, powinna ona sięgać dalej: ustalać związki pomiędzy działaniami a wynikami, sprawdzać, czy leżące u podstaw programu założenia teoretyczne (czy *quasi*-teoretyczne) okazały się adekwatne, dostarczać wiedzy o przyczynach niepowodzeń, które dzięki temu mogą okazać się nauką na przyszłość, wartą wydanych pieniędzy. Ewaluacja powinna stymulować zmiany w politykach publicznych, zarządzaniu i alokacji środków. Tym jednak, co porządna ewaluacja (w tym *ex-ante*) może spowodować względnie szybko, jest poprawa standardów projektowania przedsięwzięć publicznych, w szczególności precyzji formułowania celów. Wiemy na ten temat bardzo dużo, a mimo to powstają i są wdrażane poważne programy bez poważnej matrycy logicznej projektu,

<sup>18</sup> Por. Cook T.D., „Postpositivist critical multiplism”, w: R.L. Shotland, M.M. Mark (red.), *Social Science and Social Policy*, s. 21–62, Sage Publications, Thousand Oaks 1985.

będącej techniką porządkującą myślenie o celach, wskaźnikach i monitoringu, a w konsekwencji – ewaluacji. Ten stan rzeczy trzeba zmienić.

Zwykle kojarzy się programy i projekty publiczne, a także pomiar wyników i ewaluację, z działaniami alokacyjnymi. Nie traktuje się natomiast analogicznie działań regulacyjnych – tworzenia ustaw, rozporządzeń i innych aktów prawnych – choć ich skutki są na ogół wymierne. Powinno stać się zasadą, że wprowadzanie regulacji prawnych jest traktowane jako projekt, który także powinien mieć jasno zdefiniowane cele i system pomiaru efektów.

Pomiar osiągnięć polityk publicznych nie stanowi łatwego zadania. Najprościej jest określić oczekiwane produkty programu czy projektu, gdyż są one bezpośrednim, namacalnym lub przynajmniej opisywalnym wynikiem podejmowanych działań. Trudniej jest mierzyć wpływ interwencji na rozwiązanie problemów społecznych, którym interwencja ta jest dedykowana.

Trwająca debata nad wskaźnikami z Laeken pokazuje, że stosowanie tych zasad nie jest łatwe i, nade wszystko, że dobór wskaźników stanowi funkcję definicji celu polityk publicznych, a więc wybór *sensu stricto* polityczny, choć podlegający metodologicznym rygorom. Dlatego dobór wskaźników, zwłaszcza tych określających ostateczne wpływy programu, nie może być wyizolowany z politycznej debaty nad podstawowymi, strategicznymi celami rozwojowymi. Mówiąc krótko i dosadnie – **wyobcowanie wskaźników z polityki jest źródłem impotencji pomiaru wyników i ewaluacji.**

Na rozmaitych szczeblach administracji państwowej i samorządowej powstają diagnozy sytuacji dla potrzeb strategii rozwoju. Ich autorzy zasypują decydentów danymi statystycznymi, czyli wartościami wskaźników, które nijak się nie mają do ostatecznie przyjmowanych, *explicite* bądź *implicite*, celów działania. Dlatego często diagnozy te nie mają większego znaczenia i traktowane są jako swoista dekoracja. Trzeba podjąć trud opracowania i przyjęcia zestawu wskaźników dla poszczególnych obszarów polityk publicznych (częściowo zostaną przyjęte rozwiązania wypracowane w Unii Europejskiej, ale nie zaspokoją one wszystkich potrzeb). Trzeba także zwiększyć zaangażowanie Głównego Urzędu Statystycznego w gromadzenie potrzebnych danych i przetwarzanie ich we wskaźniki przyjęte na potrzeby polityk publicznych, wskaźniki na tyle aktualne, by miały praktyczne znaczenie dla podejmowania decyzji. Alternatywą – nie zawsze dostępną – jest zamówienie odpowiednich badań w renomowanych ośrodkach naukowych czy badawczych.

Podkreślając rolę wskaźników nie tylko w ewaluacji *ex-post*, lecz w całym procesie planowania i realizacji polityk publicznych, nie można zapominać o tym, że, choć cele trzeba formułować w odniesieniu do wskaźników czy nawet – poprzez wskaźniki, decydenci „powinni mieć na uwadze to, iż zadaniem programu jest przede wszystkim osiągnięcie celów, a nie konkretnych wartości wskaźników”<sup>19</sup>. Ustalając cele, należy kierować się zasadą: celem

<sup>19</sup> ILO, *IPEC Guidelines. Identifying & Using Indicators Of Achievement*, ILO 2003, s. 3, [http://www.ilo.org/public/english/standards/ipecc/themes/timebound/downloads/pp5\\_5en.pdf](http://www.ilo.org/public/english/standards/ipecc/themes/timebound/downloads/pp5_5en.pdf).

jest to, co jesteśmy w stanie osiągnąć i mamy ambicję osiągnąć, a nie – przyjmując postawę zachowawczą – to, co wiemy, że i tak osiągniemy.

---

**Jarosław Górniak** – doktor habilitowany, profesor Uniwersytetu Jagiellońskiego, kierownik Zakładu Socjologii Gospodarki i Edukacji Instytutu Socjologii UJ, prodziekan Wydziału Filozoficznego UJ, profesor w Wyższej Szkole Europejskiej im. Ks. J. Tischnera w Krakowie. Socjolog i ekonomista, specjalista w zakresie socjologii gospodarki i organizacji, metod badań i analizy danych, metodologii ewaluacji i analizy polityk publicznych. Ekspert Zespołu koordynacyjnego ds. ewaluacji *ex-ante* dokumentów unijnych na lata 2007–2013 w Ministerstwie Rozwoju Regionalnego. Konsultant w zakresie programowania i ewaluacji, a także w zakresie systemów analizy danych w zarządzaniu i marketingu.

**Karolina Keler** – absolwentka socjologii Uniwersytetu Jagiellońskiego (specjalności: praca socjalna oraz socjologia gospodarki i badania rynku), doktorantka w Zakładzie Socjologii Gospodarki i Edukacji Instytutu Socjologii UJ, wykładowca w Wyższej Szkole Europejskiej im. Ks. J. Tischnera, prowadzi zajęcia z obszaru badań społecznych.

Brała udział m.in. w pilotażowej ewaluacji śródkresowej Działania 1.2 i 1.3 Sektorowego Programu Operacyjnego Rozwoju Zasobów Ludzkich 2004–2006 oraz w projekcie badawczym *Małopolski system analiz i programowania polityk rynku pracy*, poświęconym budowie systemu wskaźników wykluczenia społecznego i rynku pracy dla Małopolski, realizowanym w ramach Zintegrowanego Programu Rozwoju Regionalnego. Ponadto, w latach 2005–2007 brała udział w trzech edycjach międzynarodowego kursu: „Social Exclusion in Europe”. Obecnie koordynuje badania poświęcone systemowemu wykorzystaniu danych zbieranych przez Miejski Ośrodek Pomocy Społecznej w Krakowie do celów diagnostycznych i ewaluacyjnych.

## Bibliografia

- Atkinson T., Cantillon B., Marlier E., Nolan B., *Social Indicators. The EU Social Inclusion*, Oxford University Press, New York 2002.
- Atkinson A.B., Marlier E., Nolan B., „Indicators and targets for social inclusion in the UE”, *Journal of Common Market Studies* 2004, vol. 42, nr 1, s. 47–75.
- Birckmayer J.D., Weiss C.H., „Theory-based evaluation in practice. What do we learn?”, *Evaluation Review* 2000, vol. 24, nr 4, s. 407–431.
- Council of Europe, *Concerted Development of Social Cohesion Indicators: Methodological Guide*, Council of Europe Publishing, Strasbourg 2005, [www.coe.int/T/E/Social\\_cohesion/Social\\_Policies/GUIDE\\_EN.pdf](http://www.coe.int/T/E/Social_cohesion/Social_Policies/GUIDE_EN.pdf).
- Cobb C.W., Rixford C., *Lessons Learned From The History Of Social Indicators*, Redefining Progress, San Francisco 1998, [www.rprogress.org/newpubs/1998/SocIndHist.pdf](http://www.rprogress.org/newpubs/1998/SocIndHist.pdf).
- Cook T.D., (1985) „Postpositivist critical multiplism”, w: R.L. Shotland, M.M. Mark (red.), *Social Science and Social Policy*, s. 21–62, Sage Publications, Thousand Oaks 1985.
- European Commission, *Indicators For Monitoring And Evaluation: A Practical Guide*, The New Programming Period, 2007–2013: Methodological Working Papers, Draft Working

- Paper [X], DG Regional Policy 2006, [http://ec.europa.eu/regional\\_policy/sources/docof- fic/working/doc/indic\\_sf2007-2013\\_en.pdf](http://ec.europa.eu/regional_policy/sources/docof- fic/working/doc/indic_sf2007-2013_en.pdf).
- ILO, *IPEC Guidelines. Identifying & Using Indicators Of Achievement*, ILO 2003, s. 3, [http:// www.ilo.org/public/english/standards/ipecc/themes/timebound/downloads/pp5\\_5en.pdf](http:// www.ilo.org/public/english/standards/ipecc/themes/timebound/downloads/pp5_5en.pdf).
- Nowak S., *Metodologia badań społecznych*, Państwowe Wydawnictwo Naukowe, Warszawa 1985.
- Sandhu-Rojon R., *Selecting Indicators for Impact Evaluation*, UNDP, Paris 2003, [www.pnud. ne/RENSE/Biblioth%E8que/PNUD14.pdf](http://www.pnud. ne/RENSE/Biblioth%E8que/PNUD14.pdf).



Roman Konarski, Michał Kotnarowski

## Zastosowanie metody *propensity score matching* w ewaluacji *ex-post*

### Wprowadzenie

Jednym z ważnych celów ewaluacji *ex-post* jest oszacowanie efektu przyczynowego podjętej interwencji. Z metodologicznego punktu widzenia prawdziwy eksperyment byłby najbardziej odpowiednim planem badawczym, jaki można by wykorzystać w tym celu. Jednak w większości przypadków badań prowadzonych w naukach społecznych zastosowanie planu eksperymentalnego nie jest możliwe. Eksperymenty społeczne są zbyt kosztowne w realizacji oraz nieraz wymagają nieetycznego przymusu partycypacji w grupach porównawczych. Z jednej strony, osoby biorące udział w badaniu mogą nie wyrażać zgody na poddanie się manipulacji eksperymentalnej. Z drugiej zaś – brak takiej manipulacji może być nieetyczny, szczególnie gdy byłaby ona korzystna. Na przykład, wystąpiłby znaczący problem etyczny, jeżeli badacz przypisywałby uczestników w sposób losowy do grupy ujętej programem kształcenia i grupy kontrolnej (pozbawionej takiego programu) w celu oszacowania efektu programu.

Wykorzystanie danych *quasi*-eksperymentalnych lub nieeksperymentalnych w celu wnioskowania przyczynowego wiąże się z wystąpieniem znaczących problemów<sup>1</sup>. Badania korelacyjne charakteryzują się brakiem losowej selekcji (randomizacji) do poziomów zmiennej egzogenicznej (niezależnej), co nie pozwala na przyjęcie założenia równoważnych grup porównawczych. W ostatnich trzydziestu latach podejście stanu kontrfaktycznego, rozwijane przez statystyków<sup>2</sup> oraz ekonometryków<sup>3</sup> stało się dominujące w analizie

<sup>1</sup> Holland P.W., „Statistics and causal inference”, *Journal of the American Statistical Association* 1986, nr 81(396), s. 945–960; LaLonde R.J., „Evaluating the econometric evaluations of training programs with experimental data”, *American Economic Review* 1986, nr 76, s. 604–620.

<sup>2</sup> Rubin D.B., „Estimating causal effects of treatments in randomized and nonrandomized studies”, *Journal of Educational Psychology* 1974, nr 66, s. 688–701; Rubin D.B., „Assignment to treatment group on the basis of a covariate”, *Journal of Educational Statistics* 1977, nr 2, s. 1–26; Rubin D.B., „Bayesian inference for causal effects: The role of randomization”, *Annals of Statistics* 1978, nr 6, s. 34–58; Rosenbaum P.R., Rubin D.B., „The central role of the propensity score in observational studies for causal effects”, *Biometrika* 1983, nr 70(1), s. 41–55.

<sup>3</sup> Heckman J.J., Hotz V.J., „Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training”, *Journal of the American Statistical Association* 1989, nr 84, s. 862–880.

przyczynowej w badaniach ewaluacyjnych. Blisko powiązana z podejściem stanu kontrfaktycznego jest metoda *propensity score matching* (PSM), rozwinięta ponad dwadzieścia lat temu przez Rosenbauma i Rubina<sup>4</sup>.

W tym rozdziale prezentujemy zarys podejścia stanu kontrfaktycznego oraz ściśle z nim związanej metody PSM (*propensity score matching*) w analizie przyczynowej skuteczności interwencji. Rozdział składa się sześciu części. Część druga stanowi wprowadzenie do podejścia stanów kontrfaktycznych. W części trzeciej przedstawione zostało zastosowanie podejścia PSM do selekcji grupy kontrolnej w badaniach ewaluacyjnych. W części czwartej prezentujemy przykład zastosowania metody PSM w analizie efektu netto komponentu regionalnego programu Phare 2002 SSG RZL. W części piątej natomiast podajemy wyniki oszacowania efektu netto dla tego programu. Część szóstą podsumowuje możliwości wykorzystania metody PSM w badaniach ewaluacyjnych.

## Stany kontrfaktyczne

Podejście stanu kontrfaktycznego pozwala na konceptualne ujęcie problemu wnioskowania o wielkości efektu przyczynowego w sytuacjach nieeksperymentalnych<sup>5</sup>. Przyjmijmy, że  $Y$  jest miarą wyniku interwencji, której efektywność stanowi przedmiot naszej oceny. Interwencja może przyjąć formę programu kształcenia lub pomocy uczestnikom programu (beneficjentom) w znalezieniu zatrudnienia. Przykładami wyniku programu mogą być zarobki, status zatrudnienia lub osiągnięcia edukacyjne beneficjentów. Dalej przyjmijmy, że, chociaż określona osoba mogłaby zostać przypisana *a priori* do jednego z dwóch alternatywnych stanów, to dana osoba może być poddana tylko jednemu z tych dwóch warunków. Każdy z owych dwóch alternatywnych stanów jest charakteryzowany przez swoisty zbiór warunków, mający różny wpływ na wynik interwencji  $Y$ . Dla potrzeb obecnej prezentacji będziemy określali te dwa stany jako warunek interwencji i warunek kontrolny. Warunek interwencji to udział w programie, a warunek kontrolny to brak udziału w programie.

W typowym badaniu ewaluacyjnym określona grupa beneficjentów bierze udział w programie (interwencji), inna zaś, z reguły większa grupa osób nie bierze w nim udziału. Brak udziału w programie określamy mianem warunku kontrolnego. Fundamentalna różnica pomiędzy badaniem eksperymentalnym a badaniem ewaluacyjnym polega na procesie selekcji do warunku interwen-

<sup>4</sup> Rosenbaum P.R., Rubin D.B., „The central role...”, op.cit.

<sup>5</sup> Por. Holland P.W., „Statistics...”, op.cit. W literaturze traktującej o podejściu stanów kontrfaktycznych istnieje relatywnie duża różnorodność w stosowanym zapisie. W obecnej prezentacji podejścia stanów kontrfaktycznych opieramy się na zapisie rozwiniętym przez Winshipa i Morgana, zob.: Winship C., Morgan S.L., „The estimation of causal effects from observational data”, *Annual review of Sociology* 1999, nr 25, s. 659–706.



cji i warunku kontrolnego. W przypadku eksperymentu taka selekcja stanowi wynik randomizacji. W przypadku badania efektu programu udział w programie jest wynikiem procesu autoselekcji, wynikającego z cech (obserwowalnych i nieobserwowalnych) beneficjentów programu. Kluczowym założeniem podejścia efektu kontrfaktycznego jest to, że osoby przypisane do stanu interwencji (programu) i warunku kontrolnego (nie-uczestniczenia w programie) posiadają potencjalne wyniki w obu stanach, chociaż w rzeczywistości są obserwowane tylko w jednym z dwóch warunków. W efekcie osoby badane mają zaobserwowane wyniki w warunku, w którym są obserwowane i potencjalne wyniki w warunku, któremu nie są poddane.

Ujęcie wyników oraz stanów kontrfaktycznych podsumowano w tab. 1, w której każda osoba  $i$  posiada wynik prawdziwy  $Y_i$  jako rezultat poddania jej dwóm alternatywnym warunkom, określanym jako warunek interwencji (t) i warunek kontrolny (k). Wynik  $Y_i$  można określić na skali interwałowej (np. poziom zarobków) lub dyskretnej (np. status zatrudnienia). Wyniki  $Y_i^t$  i  $Y_i^k$  to potencjalne wyniki osoby  $i$  zaobserwowane jako wynik poddania tej osoby warunkowi interwencji i warunkowi kontrolnemu. W ujęciu stanów kontrfaktycznych zakładamy, że oba potencjalne wyniki istnieją (są możliwe) w teorii, ale tylko jeden potencjalny wynik,  $Y_i^t$  lub  $Y_i^k$ , może być zaobserwowany na poziomie każdej indywidualnej osoby  $i$ .

Tabela 1. Podsumowanie ujęcia stanów kontrfaktycznych

		Stan (warunek)	
		Interwencji (t)	Kontrolny (k)
Grupa	Interwencji (T) ( $i \in T$ )	wynik zaobserwowany $Y_i^t$	wynik (nieobserwowany) kontrfaktyczny $Y_i^k$
	Kontrolna (K) ( $i \in K$ )	wynik (nieobserwowany) kontrfaktyczny $Y_i^t$	wynik zaobserwowany $Y_i^k$

Efekt przyczynowy interwencji (programu) na wynik każdej osoby  $i$  jest zdefiniowany jako różnica pomiędzy potencjalnymi wynikami zaobserwowanymi w stanie interwencji oraz w stanie kontrolnym. Ponieważ oba wyniki  $Y_i^t$  i  $Y_i^k$  są teoretycznie możliwe, to możemy, przynajmniej w teorii, zdefiniować indywidualny efekt przyczynowy programu jako

$$\delta_i = Y_i^t - Y_i^k.$$

Zauważmy jednak, że nie jest możliwe zaobserwowanie obu wyników  $Y_i^t$  i  $Y_i^k$  na poziomie indywidualnej osoby  $i$ . W konsekwencji nie możemy bezpośrednio obliczyć efektów przyczynowych interwencji na poziomie indywidualnych uczestników programu.

Problem obliczenia wyniku przyczynowego  $\delta_i$  wynika z faktu, iż wyniki  $Y_i^t$  i  $Y_i^k$  mogą być zaobserwowane jedynie dla wzajemnie wykluczają-

cych się podgrup w populacji: grupy interwencji (T) i grupy kontrolnej (K). Aby zaobserwować wynik  $Y$ , każda osoba musi być poddana procesowi selekcji do jednego z dwóch warunków. Osoby wyselekcjonowane do grupy interwencji ( $i \in T$ ) posiadają wartość zmiennej selekcji  $T_i = 1$ . Natomiast osoby wyselekcjonowane do grupy kontrolnej ( $i \in K$ ) posiadają wartość zmiennej selekcji  $T_i = 0$ . W efekcie zaobserwowany wynik osoby  $Y_i = Y_i^t$ , gdy  $T_i = 1$  i  $Y_i = Y_i^k$ , gdy  $T_i = 0$ .

W podejściu stanów kontrfaktycznych przyjmujemy, że chociaż nie jest możliwe zaobserwowanie indywidualnego efektu przyczynowego interwencji  $\delta_i$ , to można zdefiniować i oszacować średni efekt przyczynowy na poziomie populacji jako

$$\bar{\delta} = \bar{Y}^t - \bar{Y}^k,$$

gdzie  $\bar{Y}^t$  to oczekiwana (średnia) populacyjna wartość wszystkich indywidualnych wyników  $Y_{i \in T}^t$  osób w warunku interwencji, a  $\bar{Y}^k$  to oczekiwana (średnia) populacyjna wartość wszystkich indywidualnych wyników  $Y_{i \in K}^k$  osób w warunku kontrolnym. Ponieważ  $\bar{Y}^t$  i  $\bar{Y}^k$  nie są obserwowalne (są brakujące) dla wzajemnie wykluczających się podzbiorów populacji  $\bar{Y}^t$  i  $\bar{Y}^k$ , nie mogą zostać obliczone dla całej populacji. Możliwe jest jednak oszacowanie wielkości  $\bar{Y}^t$  i  $\bar{Y}^k$ , jeżeli przyjmiemy założenie, iż można zrównać grupę interwencji z grupą kontrolną w taki sposób, że spełnione będą następujące warunki:  $\bar{Y}_{i \in T}^t = \bar{Y}_{i \in K}^t$  oraz  $\bar{Y}_{i \in K}^k = \bar{Y}_{i \in T}^k$ .

Jeżeli przyjmie się powyższe założenia, spójny estymator średniego efektu interwencji jest różnicą pomiędzy oszacowanymi średnimi wynikami w grupie uczestników i grupie nie-uczestników interwencji:

$$\hat{\delta} = \hat{Y}_{i \in T}^t - \hat{Y}_{j \in K}^k.$$

W konsekwencji proces selekcji osób do grupy interwencji i grupy kontrolnej determinuje stopień obciążenia, z jakim estymator  $\hat{\delta}$  szacuje średni efekt programu  $\bar{\delta}$ . Metoda PSM próbuje wyeliminować obciążenie wynikające z nieodłącznych różnic pomiędzy grupą interwencji a grupą kontrolną.

Ujęcie stanów kontrfaktycznych w analizie efektu przyczynowego przyjmuje krytyczne założenie, znane pod nazwą *stable unit-treatment assumption* (SUTVA)<sup>6</sup>. SUTVA określa, że efekt interwencji dla określonego uczestnika programu nie zależy od uczestnictwa innych osób w programie. Założenie to nie jest weryfikowalne empirycznie. Przykładem jego naruszenia może być sytuacja, w której w wyniku uczestnictwa licznej grupy osób w programie aktywizacji bezrobotnych następuje zmiana w lokalnym rynku pracy. Jak zauważa Winship i Morgan<sup>7</sup> w takiej sytuacji prostota podejścia kontrfaktycznego w analizie efektu przyczynowego zanika.

<sup>6</sup> Rosenbaum P.R., Rubin D.B., „The central role...”, op.cit.

<sup>7</sup> Winship C., Morgan S.L., „The estimation of causal effects...”, op.cit.

## Metoda PSM

Metoda PSM (*propensity score matching*) oferuje szczególnie efektywny sposób korekty obciążenia szacunków efektu netto spowodowanego selekcją. Celem metody PSM jest zrównanie grupy interwencji i grupy kontrolnej na cechach mierzalnych zawartych w wektorze  $X$ . Dopasowanie grupy kontrolnej do grupy interwencji osiąga się poprzez selekcję grupy kontrolnej z większej puli osób nie poddanych interwencji, przeprowadzoną w taki sposób, że rozkład cech zmiennych mierzalnych w  $X$  jest maksymalnie zbliżony (zbalansowany) w obu grupach. Najbardziej atrakcyjną właściwością metody PSM to możliwość zastąpienia zbioru obciążających korelatów pojedynczą wartością, którą jest oszacowany *propensity score*.

Prawdopodobnie najbardziej oczywistym sposobem zrównania obu grup na zmiennych mierzalnych w  $X$  byłoby znalezienie dla każdej osoby  $i$  w grupie interwencji osoby  $j$  w puli osób nie poddanych interwencji z dokładnie takim samym wektorem zmiennych mierzalnych. Jednak znalezienie dokładnych dopasowań może okazać się niemożliwe, szczególnie gdy wszystkie lub niektóre ze zmiennych obserwowalnych są określone są na skali ciągłej. Kompromisem mogłoby być skategoryzowanie zmiennych w  $X$  i powiązanie wszystkich osób z grupy interwencji i grupy kontrolnej, które znajdują się w tej samej komórce powstałej tabeli krzyżowej. Niestety takie podejście także nie jest praktyczne. Jeżeli skategoryzujemy 20 zmiennych tylko do dwóch kategorii, to potencjalna grupa kontrolna będzie rozłożona pomiędzy  $2^{20} = 1048576$  kategorii dopasowania. Z tak dużą liczbą możliwych kategorii możemy posiadać w grupie interwencji wiele jednostek, dla których nie będzie można znaleźć dokładnie dopasowanej jednostki w puli potencjalnych jednostek kontrolnych.

Niemożność dopasowania przypadku kontrolnego do każdego przypadku poddanego interwencji określa się mianem problemu *common support*.

Dlatego w praktyce rezygnujemy z dokładnego dopasowania i wybieramy do grupy kontrolnej osoby, które są w określonym stopniu „podobne” do osób poddanych interwencji. Przenikliwość Rosenbauma i Rubina<sup>8</sup> polegała na obserwacji, że wymiarowość problemu selekcji grupy kontrolnej może być istotnie zredukowana poprzez zrównanie grupy kontrolnej z grupą poddaną interwencji na podstawie *propensity score*, który możemy nieformalnie zdefiniować jako skłonności do partycypacji w warunku interwencji. *Propensity score* jest skalarem otrzymanym przez modelowanie prawdopodobieństwa selekcji do warunku interwencji jako funkcji obserwowalnych korelatów (cech) zawartych w wektorze  $X$ . W efekcie *propensity score* określa prawdopodobieństwo selekcji do warunku interwencji, warunkowo na wartościach zmiennych obserwowalnych w  $X$ <sup>9</sup>. Innymi słowy, *propensity score* dla osoby  $i$

<sup>8</sup> Rosenbaum P.R., Rubin D.B., „The central role...”, op.cit.

<sup>9</sup> Ibidem.

jest prawdopodobieństwem, że osoba  $i$ , charakteryzowana zbiorem wartości zmiennych obserwowalnych  $X_i$ , będzie przypisana do grupy interwencji

$$P(X_i) = \Pr(T = 1 | X_i),$$

gdzie  $T = \{0, 1\}$  jest wskaźnikiem udziału w grupie interwencji,  $X_i$  to  $q$ -wymiarowy wektor wartości cech zaobserwowany przed interwencją i  $0 < P(X_i) < 1$ .

Jeżeli udział w programie jest w pełni funkcją zaobserwowanych cech w  $X$  (selekcja odbywa się jedynie na podstawie zaobserwowanych cech), to warunkowo na zmiennych w  $X$  udział w programie jest losowy z punktu widzenia wyniku  $Y$ . To założenie nazywa się *conditional independence assumption* (CIA).

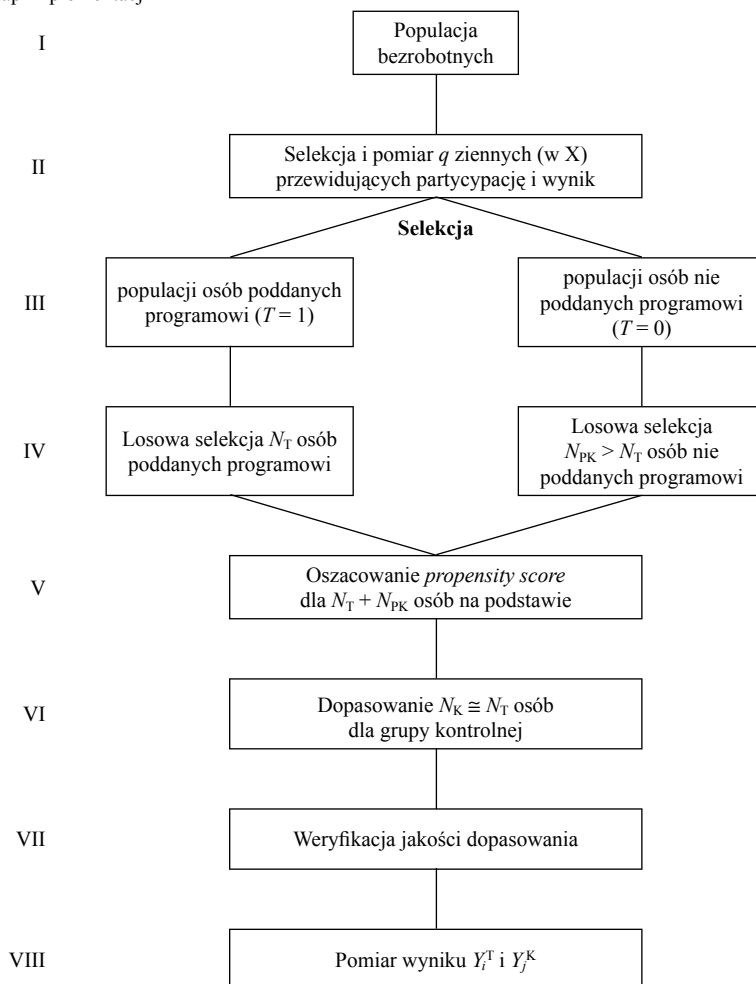
CIA wypływa z właściwości balansującej, jaką posiada *propensity score*. Rosenbaum i Rubin<sup>10</sup> demonstrują analitycznie, że jeżeli grupa interwencji i grupa kontrolna posiadają podobne rozkłady swoich *propensity scores*, to będą (zbalansowane) posiadały także podobne rozkłady zmiennych mierzalnych w  $X$ . W efekcie właściwość balansująca naśladuje randomizację w planie eksperymentalnym. Spełnienie zarówno założenia SUTVA, jak i założenia CIA pozwala na oszacowanie wyniku kontrfaktycznego dla grupy poddanej interwencji. Ponadto, analogicznie do możliwości empirycznej weryfikacji założenia SUTVA, empiryczna weryfikacja założenia CIA nie jest w pełni możliwa. Spełnienie założenia CIA zależy głównie od identyfikacji i kontroli wszystkich zmiennych determinujących partycypację i wynik interwencji.

Implementacja metody PSM jest zasadniczo procedurą dwuetapową, w której najpierw szacujemy *propensity scores*, a następnie dobieramy grupę kontrolną na podstawie wcześniej oszacowanych *propensity scores*. Jednak dla klarowności niniejszej prezentacji metody PSM podzielimy implementację tej metody na osiem odrębnych etapów. Etapy te są podsumowane na rys. 1. Implementacja metody PSM rozpoczyna się od zdefiniowania populacji osób, dla których pragniemy oszacować efekt interwencji. Potem odbywa się selekcja i pomiar zmiennych determinujących proces selekcji oraz wynik programu. Jest konieczne, aby pomiar tych zmiennych odbył się przed naturalnym procesem selekcji osób do grupy poddanej warunkowi interwencji. Przyjmujemy, że proces ten nie jest losowy i odpowiada za mierzalne (i niemierzalne) różnice pomiędzy grupą beneficjentów programu a grupą nie poddaną oddziaływaniu programu. Konsekwencje tego właśnie procesu selekcji pragniemy osłabić poprzez zastosowanie procedury PSM. Celem etapu czwartego jest selekcja reprezentatywnej próby  $N_T$  z populacji wszystkich osób poddanych interwencji oraz relatywnie większej próby  $N_{PK}$  z puli kontrolnej osób nie poddanych interwencji. Z puli osób nie poddanych interwencji dobiera się grupę kontrolną w dalszym etapie implementacji metody PSM. Etap piąty

<sup>10</sup> Ibidem. Zob. też: Rosenbaum P.R., „Constructing matched sets of strata”, w: *Observational Studies*, s. 200–224, Springer-Verlag, New York 1995.

polega na oszacowaniu *propensity scores* dla  $N_T + N_{PK}$  osób wcześniej wyselekcjonowanych do grupy interwencji i puli kontrolnej. W etapie szóstym następuje selekcja dopasowanej (równoważnej) grupy  $N_K$  jednostek kontrolnych. W etapie siódmym weryfikujemy jakość dopasowania grupy kontrolnej. Zazwyczaj ostatnim etapem implementacji metody PSM jest ocena efektu netto jako różnicy pomiędzy średnim wynikiem w grupie poddanej programowi a średnim wynikiem wyselekcjonowanej grupy kontrfaktycznej.

Etap implementacji



Rysunek 1. Podsumowanie etapów implementacji metody PSM

W dalszej części rozdziału szczegółowo omówimy cztery etapy implementacji metody PSM, które są szczególnie krytyczne dla trafności oszacowanego efektu interwencji. Owe cztery etapy to: selekcja i pomiar korelatów

partycypacji oraz wyniku programu, szacowanie wyników skłonności do partycypacji, selekcja grupy kontrolnej oraz ocena jakości dopasowanej grupy kontrolnej.

### Selekcja i pomiar korelatów partycypacji

Selekcja obciążających korelatów szacowanego efektu netto jest jednym z krytycznych etapów implementacji metody PSM. Dwa podstawowe kryteria selekcji tych zmiennych to przypuszczalny wpływ na decyzję uczestnictwa w programie i wynik programu oraz czas pomiaru w stosunku do czasu podjętej interwencji. Po pierwsze, skuteczność zrównania grupy kontrolnej z grupą poddaną procesowi interwencji wiąże się bezpośrednio z selekcją odpowiednich zmiennych. Cechy determinujące decyzję uczestnictwa w programie oraz wynik programu mogą być obserwowalne lub nieobserwowalne. W sytuacji idealnej wszystkie zmienne wpływające na udział w programie i jego wynik powinny być zawarte w modelu szacującym *propensity scores*<sup>11</sup>. Zmienne nie posiadające wpływu ani na udział w programie, ani na wynik są w oczywisty sposób nieistotne dla różnic między grupą poddaną interwencji a pulą kontrolną. Nie ma także potrzeby kontrolowania zmiennych wpływających jedynie na udział w programie, gdyż wynik programu nie zależy od nich. Analogicznie, jeżeli zmienne wpływają jedynie na wynik programu, to nie ma potrzeby kontrolowania tych zmiennych, gdyż ich rozkłady nie będą różniły się między grupą poddaną programowi a pulą kontrolną. W efekcie zmienne wymagające kontroli to zmienne determinujące udział w programie oraz wynik. Rozkłady takich zmiennych będą różniły się między grupą uczestników programu i grupą nie-uczestników programu, a jednocześnie w istotny sposób wpłyną na zaobserwowany wynik w obu grupach. Selekcja zmiennych powinna być oparta na rozważaniach teoretycznych oraz poprzednich badaniach empirycznych. Ważne jest, aby ostrożnie rozważyć wszystkie czynniki mające wpływ na selekcję do programu oraz wynik programu i wybrać maksymalną liczbę zmiennych mierzalnych. Wyselekcjonowane korelaty są podstawą oszacowania *propensity scores* i od jakości tych cech zależy stopień spełnienia założenia CIA. Heckman, Smith i Clements<sup>12</sup> wykazali, że ominięcie ważnych zmiennych może w znacznym stopniu obciążyć szacunki efektu netto.

Po drugie, pomiar wartości wyselekcjonowanych zmiennych powinien nastąpić przed rozpoczęciem programu będącego przedmiotem ewaluacji. Należy unikać zwłaszcza zmiennych mierzonych już po rozpoczęciu pro-

<sup>11</sup> Rosenbaum P.R., „The consequences of adjustment for a concomitant variable that has been affected by the treatment”, *Journal of the Royal Statistical Society* 1984, seria A, nr 147, s. 656–666.

<sup>12</sup> Heckman J.J., Smith J., Clements N., „Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts”, *Review of Economic Studies* 1997, nr 64, s. 487–537.

gramu, których wartości mogły się zmienić jako efekt podjętego programu<sup>13</sup>. Rosenbaum<sup>14</sup> oraz Heckman, LaLonde i Smith<sup>15</sup> pokazują, dlaczego zawarcie takiej zmiennej w modelu partycypacji może prowadzić do znacznego obciążenia szacowanego efektu programu. Aby to zrozumieć, rozważmy cechę poziomu motywacji. Poziom motywacji osoby bezrobotnej jest prawdopodobnie jednym z najważniejszych czynników wpływających na przystąpienie do programu aktywizacji bezrobotnych oraz na wynik programu, który może być zdefiniowany jako znalezienie zatrudnienia. Jeżeli poziom motywacji uczestników programu zmienia się pod wpływem uczestnictwa w programie, my zaś zrównamy grupę kontrolną do poziomu motywacji osób poddanych programowi w chwili jego zakończenia, skutkiem takiego zrównania może być zamaskowanie (obciążenie) szacowanego efektu programu. Dlatego Rosenbaum<sup>16</sup> utrzymuje, że najprostszym i najbardziej bezpiecznym rozwiązaniem jest unikanie zmiennych, których pomiar nastąpił już po rozpoczęciu interwencji, nawet, jeżeli dana zmienna stanowi kluczowy determinant decyzji uczestnictwa w programie i jego wyniku.

### Szacowanie wyników skłonności do partycypacji

W praktyce wyniki skłonności do partycypacji (*propensity scores*)  $P(X_i)$  osób w grupie interwencji i osób w puli kontrolnej nie są znane i muszą być oszacowane. Chociaż istnieją różnorodne metody szacowania  $P(X_i)$ , zdecydowanie najszerszej wykorzystywanym modelem statystycznym jest model regresji logistycznej. Na przykład Rosenbaum i Rubin<sup>17</sup> oraz Rosenbaum<sup>18</sup>, prezentując metodę PSM, wykorzystują model regresji logistycznej.

Szacowanie  $\Pr(T_i = 1 | X_i)$  osiąga się za pomocą następującego modelu regresji logistycznej:

$$\Pr(T_i = 1 | X_i) = \frac{\exp\left[\beta_0 + \sum_{r=1}^q \beta_r X_{ir}\right]}{1 + \exp\left[\beta_0 + \sum_{r=1}^q \beta_r X_{ir}\right]},$$

gdzie  $\beta_0$  to stała regresji, a  $\beta_r$  ( $r = 1, 2, \dots, q$ ) to współczynniki regresji dla każdego z  $q$  predyktorów wskaźnika udziału w grupie interwencji. Szacowany

<sup>13</sup> Rosenbaum P.R., „The consequences...”, op.cit.

<sup>14</sup> Ibidem.

<sup>15</sup> Heckman J.J., LaLonde R.J., Smith J.A., „The economics and econometrics of active labor market programs”, w: O.C. Ashenfelter, D. Card (red.), *Handbook of Labor Economics*, vol. 3, s. 1865–2097, Elsevier, Amsterdam 1999.

<sup>16</sup> Rosenbaum P.R., „The consequences...”, op.cit., s. 663.

<sup>17</sup> Rosenbaum P.R., Rubin D.B., „Constructing a control group using multivariate matched sampling methods that incorporate the propensity score”, *The American Statistician* 1985, nr 39(1), s. 33–38.

<sup>18</sup> Rosenbaum P.R., „Constructing matched of sets of strata”, w: *Observational Studies*, s. 200–224, Springer-Verlag, New York 1995, rozdz. 9.

model regresji może zawierać efekty główne oraz interakcje i/lub efekty kwadratowe<sup>19</sup>.

Należy rozważyć dwa zagadnienia związane ze specyfikacją szacowanego modelu partycypacji. Po pierwsze, musimy zastanowić się nad tym, które z wcześniej wyselekcjonowanych determinantów partycypacji i wyniku trzeba zawrzeć w ostatecznym modelu regresji. Rubin i Thomas<sup>20</sup> radzą, by nie wykluczać z modelu zmiennych, których współczynniki nie są statystycznie istotne. Autorzy ci<sup>21</sup> utrzymują, że zawarcie w modelu zmiennych nieistotnych statystycznie może poprawić dopasowanie grupy kontrolnej i tylko w minimalny sposób komplikuje oszacowanie indywidualnych skłonności do partycypacji (*propensity scores*). Również Zhao<sup>22</sup> potwierdził w serii badań symulacyjnych, iż zawarcie zmiennych nieistotnych dla predykcji partycypacji nie obciąża szacunków indywidualnych *propensity scores*.

Po drugie, trzeba rozważyć, czy testowany model powinien być „addytywny” i zawierać tylko efekty główne, czy też powinien zawierać efekty interakcji oraz/lub efekty kwadratowe (lub wyższego rzędu). Chociaż Rosenbaum i Rubin<sup>23</sup> szacują model zawierający efekty kwadratowe, to nie spotyka się często w literaturze innych przykładów wykorzystania modelu bardziej złożonego od addytywnego. Przeciwno rutynowemu wykorzystaniu modeli zawierających interakcje oraz/lub efekty wyższego rzędu może świadczyć fakt, że rośnie prawdopodobieństwo wystąpienia problemów numerycznych w szacowaniu takich modeli<sup>24</sup>. Ponadto Zhao<sup>25</sup> wykazuje, że zawarcie statystycznie nieistotnych efektów wyższego rzędu może doprowadzić do obciążonych szacunków skłonności do partycypacji i, w efekcie, obciążonych szacunków efektu netto podjętej interwencji.

## Selekcja grupy kontrolnej

W literaturze statystycznej opisano różnorodne metody dopasowania grupy kontrolnej, bazujące na wcześniej oszacowanych *propensity scores*<sup>26</sup>.

<sup>19</sup> Zainteresowany czytelnik znajdzie szczegóły szacowania modelu regresji logistycznej w klasycznych publikacjach, takich jak (kolejność rosnąca technicznego zaawansowania prezentacji): Menard S., *Applied Logistic Regression Analysis, Second Edition*, Sage Publications, Thousand Oaks 2002; Hosmer D.W., Lemeshow S., *Applied Logistic Regression, Second Edition*, Wiley, New York 2000; Cox D.R., Snell E.J., *Analysis of Binary Data, Second Edition*, Chapman and Hall, New York 1989.

<sup>20</sup> Rubin D.B., Thomas N., „Matching using estimated propensity scores: relating theory to practice”, *Biometrics* 1996, nr 52, s. 253.

<sup>21</sup> Ibidem.

<sup>22</sup> Zhao Z., *Sensitivity of Propensity Score Methods to the Specifications*, Discussion Paper No. 1873, Institute for the Study of Labor, Bonn 2005.

<sup>23</sup> Rosenbaum P.R., Rubin D.B., „The central role...”, op.cit.

<sup>24</sup> Zob. Hosmer D.W., Lemeshow S., *Applied Logistic Regression...*, op.cit., rozdz. 4.5.

<sup>25</sup> Zhao Z., *Sensitivity...*, op.cit.

<sup>26</sup> Zob. Rosenbaum P.R., Rubin D.B., „Constructing a control group...”, op.cit.; Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator: Evidence from evaluating



Rosenbaum i Rubin<sup>27</sup> przystępnie objaśniają implementację metody najbliższego sąsiada (*nearest neighbor matching*), bazującej na odległościach Mahalanobisa, zawierających skłonność do partycypacji (*Mahalanobis metric matching including the propensity score*), oraz implementację zmodyfikowanej metody caliper (*nearest available Mahalanobis metric matching within calipers defined by the propensity score*). Heckman i jego współpracownicy<sup>28</sup> prezentują metodę kernel (*kernel matching*).

**Metoda najbliższego sąsiada** (*nearest neighbor*) jest najczęściej wykorzystywaną, a jednocześnie najbardziej intuicyjną metodą selekcji grupy kontrolnej<sup>29</sup>. W tradycyjnej wersji tego algorytmu najpierw odbywa się losowe uporządkowanie przypadków poddanych interwencji, a następnie selekcjonuje się dla każdej osoby w grupie interwencji osobę kontrolną z najbardziej zbliżonym *propensity score*. Taki algorytm można stosować ze zwracaniem lub bez zwracania. W wersji ze zwracaniem wyselekcjonowany przypadek kontrolny powraca do puli kontrolnej i może być ponownie dopasowany do innego przypadku poddanego interwencji. W wersji bez zwracania przypadek kontrolny jest usunięty z puli kontrolnej po tym, jak zostaje dopasowany do przypadku poddanego interwencji.

Metoda najbliższego sąsiada pozwala na selekcję jednej jednostki ( $k_i = 1$ ) lub kilku jednostek (np.  $k_i = 2$ ) kontrolnych dla każdego przypadku w grupie interwencji. Każdej osobie w grupie interwencji należy przypisać tę samą liczbę osób z puli kontrolnej. Dopasowanie kilku przypadków kontrolnych zwiększa liczebność grupy kontrolnej, a tym samym redukuje błąd standardowy szacowanego efektu interwencji. Może jednak także zwiększyć obciążenie szacunku efektu interwencji, gdyż zwiększa prawdopodobieństwo wystąpienia słabych (dalekich) dopasowań. Metoda najbliższego sąsiada efektywnie przewycięża problem *common support*, ponieważ każda osoba w grupie interwencji zawsze otrzyma dopasowaną osobę z puli kontrolnej. Wadą tej metody selekcji grupy kontrolnej jest, jak już wspomnieliśmy, możliwość wystąpienia słabych dopasowań, przynajmniej dla niektórych przypadków w grupie interwencji.

**Metoda caliper** niejako modyfikuje metodę najbliższego sąsiada poprzez tworzenie dopasowania tylko w granicach określonej tolerancji (caliper)  $c$ . Im mniejsza wielkość tolerancji, tym bliższe dopasowanie. Rosenbaum i Rubin<sup>30</sup> rekomendują ustalenie wielkości tolerancji  $c$  jako frakcji rozproszenia roz-

---

a job training programme”, *Review of Economic Studies* 1997, nr 64, s. 605–654; Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator”, *Review of Economic Studies* 1998, nr 65, s. 261–294.

<sup>27</sup> Rosenbaum P.R., Rubin D.B., „Constructing a control group...”, op.cit.

<sup>28</sup> Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme”, op.cit.; Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator”, op.cit.

<sup>29</sup> Rosenbaum P.R., Rubin D.B., „Constructing a control group...”, op.cit.

<sup>30</sup> Ibidem.

kładu *propensity scores*, na przykład,  $c = 0.25s$ , gdzie  $s = \sqrt{(s_T^2 + s_{PK}^2)}/2$  jest uśrednionym odchyleniem standardowym rozkładu *propensity scores* w grupie interwencji ( $s_T^2$ ) i puli kontrolnej ( $s_{PK}^2$ ). Z drugiej strony, Dehejia i Wahba<sup>31</sup> ustalają wielkość tolerancji jako  $c = 0,0001$  lub  $c = 0,00001$ . Metoda caliper efektywnie zapobiega wystąpieniu słabych dopasowań, ale może spowodować, że niektóre osoby w grupie interwencji nie otrzymają jednostek kontrolnych, co oznacza wystąpienie problemu *common support*. Tak jak w przypadku metody najbliższego sąsiada, algorytm caliper można stosować w wersji ze zwracaniem lub bez zwracania oraz w wersji selekcji jednej jednostki ( $k_i = 1$ ) lub kilku jednostek (np.,  $k_i = 2$ ) kontrolnych dla każdego przypadku w grupie interwencji.

**Metodę kernel**, będącą modyfikacją metody caliper rozwinęli Heckman i jego współpracownicy<sup>32</sup>. W metodzie tej każdy uczestnik grupy interwencji otrzymuje wszystkie jednostki z puli kontrolnej w odległości określonej wielkością tolerancji  $c$ . Wyselekcjonowane jednostki kontrolne są ważone odwrotnie proporcjonalnie do jakości dopasowania (odległości) do swojego przypadku w grupie interwencji. Wagi  $w_{ij}$  są zdefiniowane jako

$$w_{ij} = \frac{K\left(\frac{P(X_i) - P(X_j)}{c}\right)}{\sum_j K\left(\frac{P(X_i) - P(X_j)}{c}\right)},$$

gdzie  $K(\cdot)$  jest funkcją kernel z rozkładu normalnego (Gaussa). Suma wszystkich wag jednostek kontrolnych  $j$  równa jest 1, w ramach każdej jednostki w grupie interwencji  $i$   $\sum_j w_{ij} = 1$ .

### Weryfikacja jakości dopasowania grupy kontrfaktycznej

Weryfikacja jakości dopasowania grupy kontrolnej polega w głównej mierze na określaniu podobieństwa rozkładu *propensity scores* oraz rozkładów predyktorów partycypacji, wykorzystanych w modelu regresji w grupie poddanej interwencji i w grupie kontrolnej (kontrfaktycznej). Brak różnicy w rozkładach między tymi grupami świadczy o jakości wyselekcjonowanej grupy kontrolnej.

<sup>31</sup> Dehejia R.H., Wahba S., „Propensity score-matching methods for nonexperimental causal studies”, *The Review of Economics and Statistics* 2002, nr 84(1), s. 151–161.

<sup>32</sup> Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme”, op.cit.; Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator”, op.cit.

Rosenbaum i Rubin<sup>33</sup> podają sposób na określenie skuteczności zastosowanego algorytmu selekcji grupy kontrolnej. W tej procedurze najpierw określamy rozmiar braku balansu pomiędzy grupą poddaną interwencji a pulą kontrolną dla wszystkich korelatów partycypacji w  $X$ . W tym celu obliczamy wystandaryzowaną procentową różnicę pomiędzy średnimi określonej zmiennej predykcijnej dla grupy poddanej interwencji i osób w puli kontrolnej:  $100(\bar{X}_T - \bar{X}_{PK}) / \sqrt{(s_T^2 + s_{PK}^2)}/2$ , gdzie  $\bar{X}_T$  i  $\bar{X}_{PK}$  są średnimi w grupie poddanej interwencji i puli potencjalnych jednostek kontrolnych (nie poddanych interwencji), a  $s_T^2$  i  $s_{PK}^2$  to analogiczne szacunki wariancji. W podobny sposób możemy obliczyć wystandaryzowaną procentową różnicę pomiędzy średnimi określonej zmiennej predykcijnej dla grupy poddanej interwencji i grupy kontrolnej:  $100(\bar{X}_T - \bar{X}_K) / \sqrt{(s_T^2 + s_{PK}^2)}/2$ , gdzie  $\bar{X}_T$  i  $\bar{X}_K$  to średnie w grupie poddanej interwencji i grupie kontrolnej. Wariancje  $s_T^2$  i  $s_{PK}^2$  zostały zdefiniowane wcześniej. Porównanie tych dwóch wystandaryzowanych procentowych różnic, otrzymanych przed zastosowaniem metody PSM i po jej zastosowaniu, pokazuje skuteczność wykorzystanej procedury PSM w redukcji braku dopasowania pomiędzy grupą interwencji a jednostkami kontrolnymi, dla każdej zmiennej obserwowalnej w  $X$ .

Należy jednak zaznaczyć, że opisana powyżej metoda weryfikacji jakości grupy kontrolnej w żaden sposób nie weryfikuje stopnia spełnienia ani założenia SUTVA, ani założenia CIA. Oba założenia nie są weryfikowalne empirycznie. Tylko odpowiednia wiedza teoretyczna, wsparta badaniami empirycznymi odnośnie procesu prowadzącego do selekcji i wyniku interwencji, pozwala na określenie naruszenia tych założeń.

### **Analiza efektu netto w ewaluacji *ex-post* komponentu regionalnego programu Phare 2002 SSG RZL<sup>34</sup>**

Celem prezentowanego badania było określenie wielkości efektu netto dla Podprojektu 1 komponentu regionalnego programu Phare 2002 Spójność Społeczna i Gospodarcza (SSG) – Rozwój Zasobów Ludzkich (RZL). Grupę interwencji w Podprojekcie 1 programu Phare 2002 SSG RZL stanowiły osoby bezrobotne lub zagrożone bezrobociem, którym świadczone różnego rodzaju usługi rynku pracy. Podprojekt 1 miał na celu pomoc beneficjentom w znalezieniu zatrudnienia. W zaprezentowanej analizie efektu netto wykorzystujemy dane dotyczące beneficjentów, którzy w momencie przystępowania

<sup>33</sup> Rosenbaum P.R., Rubin D.B., „Constructing a control group...”, op.cit.

<sup>34</sup> PBS DGA, *Ewaluacja ex-post komponentu regionalnego programu Phare 2002 Spójność Społeczna i Gospodarcza – Komponent Rozwój Zasobów Ludzkich*, Raport z badania przeprowadzonego na zlecenie Polskiej Agencji Rozwoju Przedsiębiorczości, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2006, <http://www.parp.gov.pl/index/index/110>.

do programu byli zarejestrowani jako bezrobotni w Powiatowych Urzędach Pracy (PUP). Analizowane dane pochodzą z systemu ewidencji osób bezrobotnych PULS<sup>35</sup>. Dane w systemie PULS gromadzone są przez Powiatowe Urzędy Pracy i zawierają informacje o zarejestrowanych osobach bezrobotnych. Dotyczą one historii zatrudnienia bezrobotnych, korzystania z różnego rodzaju usług rynku pracy, a także zawierają pewne socjo-demograficzne charakterystyki zarejestrowanych.

Prezentowana ewaluacja programu Phare została przeprowadzona osobno w każdym z 16 województw. Szczegóły selekcji próby beneficjentów programu oraz puli nie-uczestników programu w każdym województwie zainteresowany czytelnik znajdzie we właściwym raporcie<sup>36</sup>. Na potrzeby tego artykułu skupiamy się jedynie na etapach badania bezpośrednio powiązanych z implementacją metody PSM. Dla lepszej ekonomii prezentacji, przedstawione wyniki ewaluacji posegregowano na poziome całego kraju.

### Selekcja i pomiar korelatów partycypacji

W selekcji dostępnych korelatów partycypacji wykorzystaliśmy dwa najważniejsze kryteria. Po pierwsze, wyselekcjonowane zmienne powinny posiadać potencjalny związek z decyzją udziału w programie i wynikiem programu (znalezienia zatrudnienia). Po drugie, pomiar wyselekcjonowanych zmiennych powinien być dokonany przed przystąpieniem do programu. Jeżeli pomiar danej zmiennej nastąpił po rozpoczęciu programu, to wartości tej zmiennej nie mogły ulec zmianie w czasie. Listę zmiennych wyselekcjonowanych jako predykatory szacowanych skłonności do partycypacji (*propensity scores*) przedstawiono w tab. 2. Zawiera ona też wykaz średnich oraz wystandaryzowanych różnic procentowych pomiędzy średnimi dla grupy  $N_T = 4\ 818$  uczestników programu<sup>37</sup> i puli  $N_{PK} = 823\ 476$  osób nie uczestniczących w programie. Można zauważyć, że grupa osób poddanych programowi w znacznym stopniu różni się od pozostałych osób bezrobotnych. Na przykład, osoby uczestniczące w programie są młodsze, lepiej wykształcone i pozostawały bezrobotne przed przystąpieniem do programu. Celem zastosowanej procedury PSM było wyselekcjonowanie grupy kontrolnej zdecydowanie bardziej podobnej do grupy uczestników programu niż zaobserwowana pula osób nie poddanych programowi.

---

<sup>35</sup> SI PULS jest systemem informatycznym wspomagającym pracę Powiatowych Urzędów Pracy

<sup>36</sup> PBS DGA, *Ewaluacja ex-post komponentu regionalnego...*, op.cit.

<sup>37</sup> Łączna liczba beneficjentów Podprojektu 1 komponentu regionalnego programu Phare 2002 Spójność Społeczna i Gospodarcza – Rozwój Zasobów Ludzkich wyniosła 13 005 osób. Ze względu na praktycznych oraz organizacyjnych nie można było przeprowadzić analiz na danych pochodzących ze wszystkich PUP-ów. W związku z tym dokonano warstwowego doboru próby PUP-ów. Szczegóły tego doboru opisano w raporcie: PBS DGA, *Ewaluacja ex-post komponentu regionalnego...*, op.cit.

Tabela 2. Obciążenie zmiennych obserwowalnych: średnie w grupie uczestników programu i puli kontrolnej oraz standaryzowana różnica (w %)

Zmienna	Opis cechy	Uczestnicy programu	Nie-uczestnicy	% obciążenia
V01	Płeć (kobieta)	50,17	52,35	-4,4
V02	Wiek (w chwili rozpoczęcia programu)	30,96	35,17	-40,3
V03	Stan cywilny (w związku)	40,54	50,95	-21,0
V04	Samotne wychowywanie dzieci	2,89	3,00	-0,7
V05	Liczba dzieci	0,46	0,62	-15,7
V06	Niepełnosprawność (niepełnosprawny)	3,13	2,52	3,7
V09	Wykształcenie			
	1 Brak lub niepełne podstawowe	0,12	0,40	-5,4
	2 Podstawowe	15,01	27,81	-31,6
	3 Gimnazjalne	0,66	0,31	5,1
	4 Zasadnicze zawodowe	25,16	31,47	-14,1
	5 Średnie	40,76	28,27	26,5
	6 Pomaturalne/policealne	6,00	3,68	10,8
	7 Wyższe (w tym licencjat)	12,29	8,06	14,0
V11	Zawód			
	0 Siły zbrojne	0,48	0,19	5,0
	1 Przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy	0,77	0,77	-0,1
	2 Specjaliści	7,18	6,07	4,5
	3 Technicy i średni personel	15,11	10,85	12,7
	4 Pracownicy biurowi	8,12	6,29	7,1
	5 Pracownicy usług osobistych i sprzedawcy	14,78	15,79	-2,8
	6 Rolnicy, ogrodnicy, leśnicy i rybacy	0,95	2,06	-9,1

Tabela 2 – cd.

Zmienna	Opis cechy	Uczestnicy programu	Nie-uczestnicy	% obciążenia
	7 Robotnicy przemysłowi i rzemieślnicy	22,40	24,54	-5,1
	8 Operatorzy i monterzy maszyn i urządzeń	5,54	6,39	-3,6
	9 Pracownicy przy pracach prostych	11,44	15,95	-13,2
	10 Brak zawodu	8,47	7,75	2,6
	99 Brak danych	4,77	3,34	7,3
V14	Posiadanie prawa jazdy kat. B	12,08	9,52	8,3
V16	Liczba szkoleń w ciągu roku przed Phare	0,04	0,01	14,7
V17	Liczba dni spędzonych na szkoleniach	1,33	0,59	10,9
V13	Liczba propozycji pracy w ciągu roku przed Phare	0,46	0,39	4,3
V18	Udział w stażu przed rozpoczęciem programu			
	0 Nie brał udziału w stażu	85,08	83,69	3,8
	1 Brał udział później niż rok i w ostatnim roku przed programem	1,54	1,73	-1,5
	2 Brał udział później niż rok przed programem	8,01	8,30	-1,1
	3 Brał udział w stażu w ostatnim roku przed programem	5,38	6,28	-3,8
V07	Staż pracy w momencie przystąpienia do Phare	2035,79	2487,75	-15,1
V10	Liczba dni na bezrobociu przed udziałem w Phare	1087,22	1438,69	-31,3
V12	Liczba wykonywanych zawodów	2,33	2,22	6,3
V15	Liczba dni przepracowanych w ramach staży przed Phare absolwentskich, prac interwencyjnych, robót publicznych	3,31	6,16	-10,1
V22	Liczba dni najdłuższego ciągłego okresu bezrobocia bezrobocia długoterminowego	287,31	337,32	-18,0
V21	Praca lub staż +/- 7 dni od programu	2,74	2,50	1,5
V19	Liczba dni na zasiłku przed udziałem w programie	362,49	356,70	1,9

Tabela 2 – cd.

V08 Odsetek stawień się na wezwanie z urzędu					
0%		7,10	8,07	-3,7	
0–50%		5,77	5,25	2,3	
51–99%		9,36	8,02	4,8	
100%		34,45	30,01	9,5	
Nie dotyczy		43,32	48,65	-10,7	
V20 Prawo do zasiłku w momencie przystępowania do Phare					
0 Brak prawa do zasiłku		32,69	44,82	-25,1	
1 Miał prawo do zasiłku i w trakcie <1 miesiąca		5,54	5,12	1,9	
2 Miał prawo do zasiłku i w trakcie 1–3 miesięcy		16,06	9,28	20,5	
3 Miał prawo do zasiłku i w trakcie >3 miesięcy		45,70	40,78	10,0	

## Szacowanie wyników skłonności do partycypacji

Szacowanie indywidualnych wyników skłonności do partycypacji (*propensity scores*), dla  $N_T = 4818$  uczestników programu i puli  $N_{PK} = 823476$  osób nie uczestniczących w programie zostało przeprowadzone za pomocą modelu regresji logistycznej, niezależnie (osobno) w każdym województwie. W szacowanym modelu regresji zmienną zależną była dychotomiczna zmienna Phare (1 = udział w programie, 0 = brak udziału w programie). Zmiennymi niezależnymi były 22 zmienne zdefiniowane w tab. 2. Testowany model był modelem addytywnym i nie zawierał ani efektów interakcji, ani efektów wyższego rzędu. Ponadto, zgodnie z zaleceniami Rubina i Thomasa<sup>38</sup> zmienne statystycznie nieistotne nie były usuwane z modelu. Efekt końcowy tego etapu implementacji metody PSM stanowiły szacunki *propensity scores* dla wszystkich uczestników i wszystkich nie-uczestników programu Phare.

## Selekcja grupy kontrolnej

Dobór grupy kontrolnej z puli nie-uczestników programu został przeprowadzony osobno w każdym województwie. W selekcji grupy kontrolnej zastosowaliśmy metodę najbliższego sąsiada. Pozwoliła ona na relatywną jednoznaczność definicji jednostek kontrolnych (osoba z najbliższymi *propensity scores*) oraz relatywną prostotę zastosowania. Jednocześnie metoda ta daje grupę kontrolną o zbliżonych właściwościach do grup kontrolnych wyselekcjonowanych za pomocą bardziej złożonych algorytmów<sup>39</sup>. Relatywna prostota podejścia najbliższego sąsiada polega na braku konieczności definiowania maksymalnej tolerancji odległości (jak w metodzie caliper) oraz braku konieczności zastosowania złożonego ważenia jednostek kontrolnych (jak w przypadku metody kernel).

Celem zastosowanego algorytmu selekcji był dobór jednej jednostki kontrolnej dla każdego beneficjenta programu. Ponieważ liczebność puli osób nie poddanych programowi, z której selekcjonowaliśmy grupę kontrolną była relatywnie duża (kilkadziesiąt tysięcy w województwie), zastosowaliśmy metodę selekcji bez zwracania. Pomimo swojej konceptualnej jednoznaczności oraz prostoty implementacja metody najbliższego sąsiada niesie z sobą swoje „komplikacje” wynikające z faktu, że dopasowanie osoby kontrolnej do danej osoby z grupy programowej nie zawsze wynosi jeden-do-jednego. Z reguły dopasowanie jeden-do-jednego było możliwe. W przypadku, gdy dla jednej osoby w grupie programowej wystąpiło kilka potencjalnych osób (z takim samym *propensity score*) w puli kontrolnej (dopasowanie jeden-do-

<sup>38</sup> Rubin D.B., Thomas N., „Matching using estimated propensity scores: relating theory to practice”, *Biometrics* 1996, nr 52, s. 249–264.

<sup>39</sup> Zob. np.: Rosenbaum P.R., Rubin D.B., „Constructing a control group using multivariate matched sampling methods that incorporate the propensity score”, *The American Statistician* 1985, nr 39(1), s. 33–38.

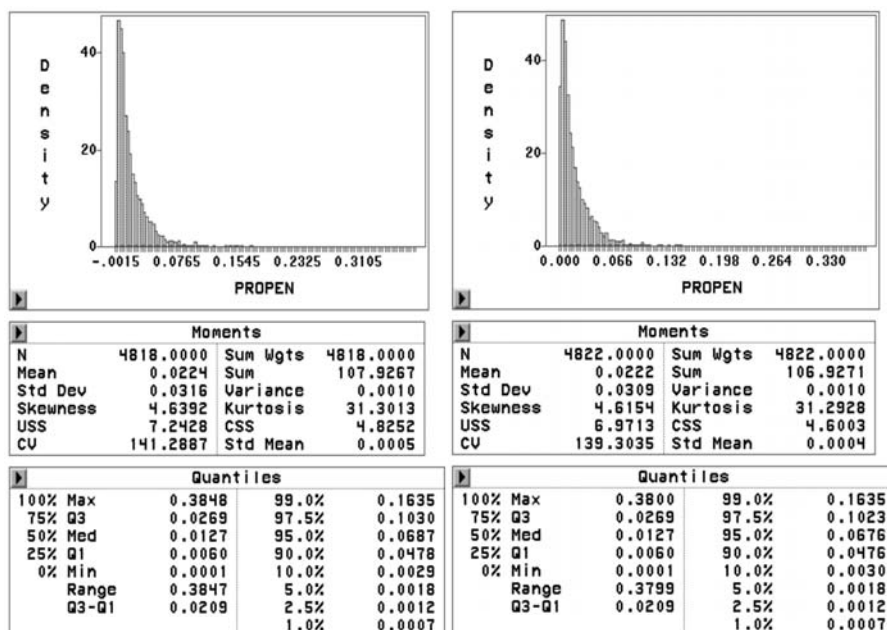


-wielu) zaimplementowany przez nas algorytm selekcjonował jedną osobę kontrolną w sposób losowy. W przypadku, gdy jedna osoba z puli kontrolnej była najbliższym sąsiadem dla kilku osób z grupy programowej (dopasowanie wiele-do-jednego) algorytm selekcjonował kolejnych najbliższych sąsiadów dla każdej jednostki w grupie programowej z identycznym *propensity score*. Sytuacja ta stawała się numerycznie złożona, kiedy wiele osób w grupie programowej posiadało po kilku tych samych, równie dobrych najbliższych sąsiadów (dopasowanie wiele-do-wielu). Wtedy selekcja dokładnej liczby grupy kontrolnej mogła być niemożliwa. Z tego powodu liczebność grupy kontrolnej ( $N_K = 4822$ ) nieznacznie różni się od liczebności grupy poddanej programowi ( $N_T = 4818$ ).

Jakość zastosowanego algorytmu selekcji można zweryfikować poprzez porównanie rozkładu szacowanych wyników skłonności do partycypacji (*propensity scores*) w grupie poddanej programowi i wyselekcjonowanej grupie kontrolnej. Rozkłady indywidualnych wyników skłonności do partycypacji oraz podstawowe statystyki opisujące cechy tych rozkładów w obu grupach przedstawia rys. 2. Jak możemy zauważyć, rozkłady oszacowanych *propensity scores* w obu grupach są bardzo zbliżone z punktu widzenia nie tylko pierwszego momentu statystycznego (średniej), lecz także wyższych (wariancji/skośności/kurtozy) momentów statystycznych.

(A) Uczestnicy programu

(B) Grupa kontrolna



Rysunek 2. Rozkład *propensity scores* w grupie uczestników programu (panel A) i w dopasowanej grupie kontrolnej (panel B)

### **Weryfikacja jakości dopasowania grupy kontrolnej**

Skuteczność zastosowanej metody PSM w dopasowaniu grupy kontrolnej do grupy beneficjentów programu Phare może być oceniona poprzez analizę redukcji obciążenia zmiennych mierzalnych. Tab. 3 prezentuje średnie poziomy oraz wystandaryzowaną procentową różnicę między średnimi dla grupy uczestników programu Phare i dopasowanej grupy kontrolnej. Zauważmy, że wystandaryzowana różnica w tab. 3 posiada ten sam mianownik, co wystandaryzowana różnica w tab. 2. Porównanie poziomu obciążenia zmiennych mierzalnych w tab. 3 (po dopasowaniu) z poziomem obciążenia w tab. 2 (przed dopasowaniem) pokazuje korzyści, jakie wynikają z implementacji metody PSM. Podczas gdy przed zastosowaniem procedury PSM prawie połowa wystandaryzowanych różnic przekroczyła poziom 20% (zob. tab. 2), po zastosowaniu procedury dopasowania żadna z wystandaryzowanych różnic nie przekroczyła poziomu 9% (zob. tab. 3). Jak widzimy w tab. 3, zastosowana metoda PSM okazała się nadzwyczaj skuteczna w zrównaniu grupy kontrolnej w cechach, w których można było zaobserwować największe różnice (np. wiek, stan cywilny, wykształcenie, liczba dni bezrobocia przed programem) między beneficjentami a pulą osób nie uczestniczących w programie Phare. Zwróćmy jednak uwagę, że chociaż redukcja obciążenia na większości cech jest znaczna, to obciążenie w niektórych kategoriach wykształcenia (V07) pozostaje na niezadowalającym poziomie. Kolejne statystyczne korekty rozkładu wykształcenia mogą być wymagane w celu dalszej redukcji możliwego obciążenia szacunku efektu netto.

Tabela 3. Obciążenie zmiennych obserwowalnych: średnie w grupie uczestników programu i dopasowanej grupie kontrolnej oraz standardyzowana różnica (w %)

Zmienna	Opis cechy	Uczestnicy programu	Dopasowana grupa kontrolna	% obciążenie
V01	Płeć (kobieta)	50,17	50,37	-0,4
V02	Wiek (w chwili rozpoczęcia programu)	30,96	31,11	-1,5
V03	Stan cywilny (w związku)	40,54	40,79	-0,5
V04	Samotne wychowywanie dzieci	2,89	2,68	1,2
V05	Liczba dzieci	0,46	0,46	0,6
V06	Niepełnosprawność (niepełnosprawny)	3,13	3,30	-1,0
V09	Wykształcenie			
	1 Brak lub niepełne podstawowe	0,12	0,29	-3,2
	2 Podstawowe	15,01	14,41	1,5
	3 Ginnazjalne	0,66	0,41	3,6
	4 Zasadnicze zawodowe	25,16	28,37	-7,2
	5 Średnie	40,76	36,64	8,7
	6 Pomaturalne/policealne	6,00	5,04	4,5
	7 Wyższe (w tym licencjat)	12,29	14,83	-8,4
V11	Zawód			
	0 Siły zbrojne	0,48	0,29	3,2
	1 Przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy	0,77	0,73	0,5
	2 Specjaliści	7,18	7,01	0,7
	3 Technicy i średni personel	15,11	15,41	-0,9
	4 Pracownicy biurowi	8,12	8,38	-1,0
	5 Pracownicy usług osobistych i sprzedawcy	14,78	14,66	0,3

Tabela 3 – cd.

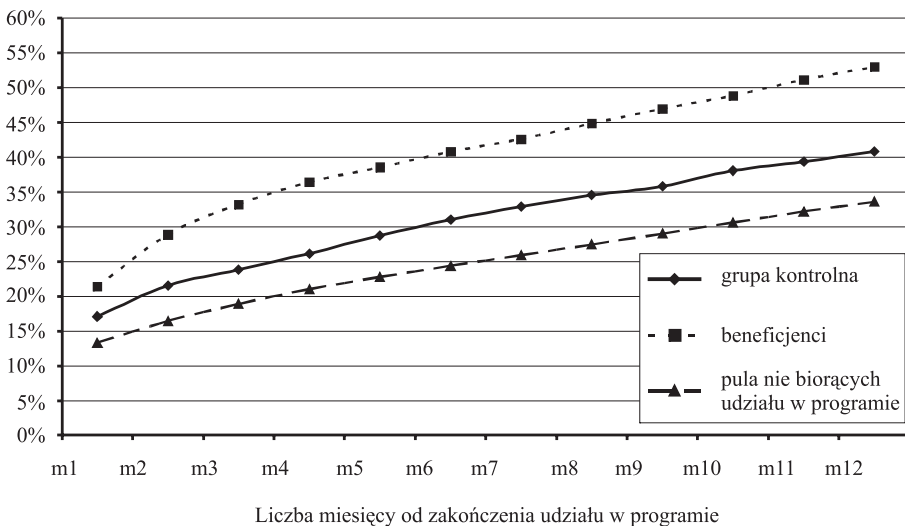
Zmienna	Opis cechy	Uczestnicy programu	Dopasowana grupa kontrolna	% obciążenie
	6 Rolnicy, ogrodnicy, leśnicy i rybacy	0,95	0,89	0,5
	7 Robotnicy przemysłowi i rzemieślnicy	22,40	22,48	-0,2
	8 Operatorzy i monterzy maszyn i urządzeń	5,54	5,23	1,3
	9 Pracownicy przy pracach prostych	11,44	11,26	0,5
	10 Brak zawodu	8,47	9,15	-2,5
	99 Brak danych	4,77	4,52	1,3
V14	Posiadanie prawa jazdy kat. B	12,08	11,41	2,2
V16	Liczba szkoleń w ciągu roku przed Phare	0,04	0,03	1,3
V17	Liczba dni spędzonych na szkoleniach	1,33	1,25	1,1
V13	Liczba propozycji pracy w ciągu roku przed Phare	0,46	0,44	1,1
V18	Udział w stażu przed rozpoczęciem programu			
	0 Nie brał udziału w stażu	85,08	84,78	0,8
	1 Brał udział później niż rok i w ostatnim roku przed programem	1,54	1,18	2,8
	2 Brał udział później niż rok przed programem	8,01	8,52	-1,9
	3 Brał udział w stażu w ostatnim roku przed programem	5,38	5,52	-0,6
V07	Staż pracy w momencie przystąpienia do Phare	2035,79	2043,63	-0,3
V10	Liczba dni na bezrobociu przed udziałem w Phare	1087,22	1094,36	-0,6
V12	Liczba wykonywanych zawodów	2,33	2,34	-0,5
V15	Liczba dni pracodawanych w ramach staży przed Phare absolwentkich, prac interwencyjnych, robót publicznych	3,31	3,26	0,2
V22	Liczba dni najdłuższego ciągłego okresu bezrobocia bezrobotnego	287,31	296,47	-3,3
V21	Praca lub staż +/- 7 dni od programu	2,74	3,17	-2,7
V19	Liczba dni na zasłuku przed udziałem w programie	362,49	369,52	-2,3

Tabela 3 – cd.

V08 Odsetek stawień się na wezwanie z urzędu					
	0%		7,10	7,84	-2,8
	0-50%		5,77	5,85	-0,3
	51-99%		9,36	9,42	-0,2
	100%		34,45	35,44	-2,1
	Nie dotyczy		43,32	41,46	3,7
V20 Prawo do zasiłku w momencie przystępowania do Phare					
	0 Brak prawa do zasiłku		32,69	31,81	1,8
	1 Miał prawo do zasiłku i w trakcie <1 miesiąca		5,54	5,89	-1,5
	2 Miał prawo do zasiłku i w trakcie 1-3 miesięcy		16,06	16,01	0,2
	3 Miał prawo do zasiłku i w trakcie >3 miesięcy		45,70	46,29	-1,2
Propensity score			2,24	2,22	1,0

## Efekt netto komponentu regionalnego programu Phare 2002 SSG RZL

Wykres odsetka zatrudnionych w grupie beneficjentów Podprojektu 1 Phare 2002 SSG RZL, w grupie kontrolnej oraz w puli nie biorącej udziału w programie przedstawia rys. 3<sup>40</sup>. Przy porównaniu odsetków zatrudnionych wśród beneficjentów i w grupie kontrolnej można zaobserwować, że już w pierwszym miesiącu od zakończenia udziału w programie beneficjenci podejmowali pracę statystycznie istotnie częściej niż osoby należące do grupy kontrolnej. Odsetek zatrudnionych wśród beneficjentów po miesiącu od zakończenia udziału w programie wyniósł 21,4%, podczas gdy analogiczny odsetek dla grupy kontrolnej wyniósł 17,1%. W kolejnych miesiącach różnica między odsetkami zatrudnionych w obu wyróżnionych grupach powiększała się i była wciąż istotna statystycznie. Po sześciu miesiącach od zakończenia udziału w programie odsetek zatrudnionych w grupie beneficjentów wynosił 40,8%, podczas gdy w grupie kontrolnej – 31,0%. Rok po zakończeniu programu pracowało 53,0% beneficjentów i 40,8% osób z grupy kontrolnej. Po roku od zakończenia udziału w programie odsetek zatrudnionych beneficjentów był więc 1,29 raza większy od analogicznego odsetka w grupie kontrolnej, co oznacza, że dwa lata po zakończeniu programu udział w nim zwiększał szansę beneficjenta na znalezienie pracy o 29%<sup>41</sup>.



Rysunek 3. Odsetki zatrudnionych wśród beneficjentów i w grupie kontrolnej – Podprojekt 1 Phare 2002 SSG RZL

<sup>40</sup> Na podstawie: PBS DGA, *Ewaluacja ex-post komponentu regionalnego...*, op.cit.

<sup>41</sup> Ibidem.

Jak widzimy na rys. 3, różnica pomiędzy wyselekcjonowaną grupą kontrolną (kontrafaktyczną) i pulą osób nie biorących udziału w programie Phare wynosi około 5%, bez względu na czas, który upłynął od zakończenia programu. Różnica ta podkreśla dwa aspekty podjętej przez nas ewaluacji programu Phare. Po pierwsze, przybliża potencjalne obciążenie szacunku efektu programu, gdyby taki szacunek był oparty o proste porównanie beneficjentów z nie-beneficjentami programu, a tym samym pokazuje korzyści wypływające z zastosowanej procedury PSM.

Po drugie, różnica między grupą kontrolną (kontrafaktyczną) i całą pulą osób nie biorących udziału w programie sygnalizuje, że najprawdopodobniej oszacowany efekt netto jako różnica między beneficjentami i grupą kontrolną stanowi jedynie górny limit efektu tego programu. Niestety, dolnego limitu tego efektu nie znamy ponieważ nie wiemy jaki byłby efekt netto gdyby założenie CIA było w pełni spełnione i grupa kontrolna była w pełni dopasowana do grupy beneficjentów na wszystkich zmiennych (mierzalnych i niemierzalnych) istotnych dla procesu selekcji i wyniku programu.

## Podsumowanie

W tym rozdziale zaprezentowaliśmy potencjał metody PSM w ewaluacji *ex-post*. Metoda PSM pozwala na selekcję grupy porównawczej z większej puli potencjalnych jednostek kontrolnych na podstawie oszacowanych indywidualnych wyników skłonności do partycypacji (*propensity score*). Gdy grupa beneficjentów programu znacznie różni się od puli osób nie uczestniczących w programie, metoda PSM oferuje efektywny sposób korekty obciążenia szacunku efektu netto. Jakość tak skorygowanego efektu netto zależy od jakości dopasowania grupy kontrolnej oraz od stopnia spełnienia założeń leżących u podłoża metody PSM. Tylko pełne zrozumienie procesu selekcji do programu oraz wyniku programu pozwala na trafne zdefiniowanie ważnych cech mierzalnych, które mogą być następnie wykorzystane w modelach szacujących *propensity scores*. Oczywiście dane dotyczące tych czynników muszą być także dostępne dla ewaluatorów programów interwencji.

---

**Roman Konarski** – absolwent York University (Toronto, Kanada), gdzie otrzymał tytuł doktora psychologii ze specjalizacją w psychometrii i statystyce. Wykładał statystykę w Queen's University oraz University of New Brunswick. Obecnie jest adiunktem w Instytucie Psychologii Uniwersytetu Gdańskiego oraz głównym statystykiem w PBS DGA. Jego akademickie zainteresowania koncentrują się wokół zastosowań modelu równań strukturalnych do problemów pomiaru psychologicznego. W PBS DGA zajmuje się metodologią i statystycznymi zagadnieniami w badaniach rynkowych. Jest autorem wielu prac z zakresu psychometrii i statystyki.

**Michał Kotnarowski** – skończył studia w Instytucie Socjologii UW. Pracuje w PBS DGA (dawniej Pracownia Badań Społecznych w Sopocie) na stanowisku konsultanta-kierownika projektów badawczych. Posiada wieloletnie doświadczenie w prowadzeniu projektów z zakresu badań opinii publicznej na bieżące tematy społeczno-polityczne (w tym pomiar preferencji parlamentarnych i prezydenckich), badań o charakterze etnograficznym, ewaluacji projektów finansowanych z funduszy unijnych.

## Bibliografia

- Cox D.R., Snell E.J., *Analysis of Binary Data, Second Edition*, Chapman and Hall, New York 1989.
- Dehejia R.H., Wahba S., „Propensity score-matching methods for nonexperimental causal studies”, *The Review of Economics and Statistics* 2002, nr 84(1), s. 151–161.
- Heckman J.J., Hotz V.J., „Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training”, *Journal of the American Statistical Association* 1989, nr 84, s. 862–880.
- Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme”, *Review of Economic Studies* 1997, nr 64, s. 605–654.
- Heckman J.J., Smith J., Clements N., „Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts”, *Review of Economic Studies* 1997, nr 64, s. 487–537.
- Heckman J.J., Ichimura H., Todd P., „Matching as an econometric evaluation estimator”, *Review of Economic Studies* 1998, nr 65, s. 261–294.
- Heckman J.J., LaLonde R.J., Smith J.A., „The economics and econometrics of active labor market programs”, w: O.C. Ashenfelter, D. Card (red.), *Handbook of Labor Economics*, vol. 3, s. 1865–2097, Elsevier, Amsterdam 1999.
- Holland P.W., „Statistics and causal inference”, *Journal of the American Statistical Association* 1986, nr 81(396), s. 945–960.
- Hosmer D.W., Lemeshow S., *Applied Logistic Regression, Second Edition*, Wiley, New York 2000.
- LaLonde R.J., „Evaluating the econometric evaluations of training programs with experimental data”, *American Economic Review* 1986, nr 76, s. 604–620.
- Menard S., *Applied Logistic Regression Analysis, Second Edition*, Sage Publications, Thousand Oaks 2002.
- PBS DGA, *Ewaluacja ex-post komponentu regionalnego programu Phare 2002 Spójność Społeczna i Gospodarcza – Komponent Rozwój Zasobów Ludzkich*, Raport z badania przeprowadzonego na zlecenie Polskiej Agencji Rozwoju Przedsiębiorczości, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2006, <http://www.parp.gov.pl/index/index/110>.
- Rosenbaum P.R., „The consequences of adjustment for a concomitant variable that has been affected by the treatment”, *Journal of the Royal Statistical Society* 1984, seria A, nr 147, s. 656–666.
- Rosenbaum P.R., „Constructing matched sets of strata”, w: *Observational Studies*, s. 200–224, Springer-Verlag, New York 1995.
- Rosenbaum P.R., Rubin D.B., „The central role of the propensity score in observational studies for casual effects”, *Biometrika* 1983, nr 70(1), s. 41–55.
- Rosenbaum P.R., Rubin D.B., „Constructing a control group using multivariate matched sampling methods that incorporate the propensity score”, *The American Statistician* 1985, nr 39(1), s. 33–38.



- Rubin D.B., „Estimating causal effects of treatments in randomized and nonrandomized studies”, *Journal of Educational Psychology* 1974, nr 66, s. 688–701.
- Rubin D.B., „Assignment to treatment group on the basis of a covariate”, *Journal of Educational Statistics* 1977, nr 2, s. 1–26.
- Rubin D.B., „Bayesian inference for causal effects: The role of randomization”, *Annals of Statistics* 1978, nr 6, s. 34–58.
- Rubin D.B., Thomas N., „Matching using estimated propensity scores: relating theory to practice”, *Biometrics* 1996, nr 52, s. 249–264.
- Winship C., Morgan S.L., „The estimation of causal effects from observational data”, *Annual review of Sociology* 1999, nr 25, s. 659–706.
- Zhao Z., *Sensitivity of Propensity Score Methods to the Specifications*, Discussion Paper No. 1873, Institute for the Study of Labor, Bonn 2005.