TEORIA I PRAKTYKA EWALUACJI INTERWENCJI PUBLICZNYCH

PODRĘCZNIK AKADEMICKI

KOŹMIŃSKI Ewaluacja programów UE

TEORIA I PRAKTYKA EWALUACJI INTERWENCJI PUBLICZNYCH

PODRĘCZNIK AKADEMICKI

POD REDAKCJĄ NAUKOWĄ
Karola Olejniczaka, Marka Kozaka, Bartosza Ledziona

WYDAWNICTWA AKADEMICKIE I PROFESJONALNE

Wydawca
Bożena Kućmierowska

Redakcja
Anna Goryńska

Recenzja
dr hab. Witold T. Bielecki

Projekt okładki, stron tytułowych, opracowanie typograficzne
Jacek Staszewski

Zdjęcie na okładce
Aleksandar Bracinac, Dreamstime.com

Publikacja została sfinansowana ze środków Unii Europejskiej w ramach Projektu Środki Przejściowe PL2005/017-488.01.02.06 „Wzmocnienie zdolności ewaluacyjnych w sektorze publicznym”.
Za jej treść odpowiadają Autorzy. Poglądy w niej wyrażone nie odzwierciedlają w żadnym razie oficjalnego stanowiska Unii Europejskiej.

Copyright © by Wydawnictwa Akademickie i Profesjonalne
Copyright © by Akademia Leona Koźmińskiego
Warszawa 2008

ISBN 978–83–61408–01–7
ISBN 978–83–89437–89–1

Wydawnictwa Akademickie i Profesjonalne Spółka z o.o.
Grupa Kapitałowa WSiP S.A.
00–696 Warszawa, ul. J. Pankiewicza 3
www.waip.com.pl

Akademia Leona Koźmińskiego
03–301 Warszawa, ul. Jagiellońska 57/59
www.kozminski.edu.pl

SPIS TREŚCI

Veronica Gaffey, Antonella Schulte-Braucks SŁOWO WSTĘPNE **9**
Karol Olejniczak, Marek Kozak, Bartosz Ledzion WPROWADZENIE **11**

CZĘŚĆ PIERWSZA

ZAGADNIENIA OGÓLNE **15**

1.1. **Karol Olejniczak** Wprowadzenie do zagadnień ewaluacji **17**
1.2. **Marek Kozak** Polityka spójności jako przedmiot ewaluacji **44**
1.3. **John Bachtler** Ewaluacja regionalnej polityki w Europie: kultura, zaangażowanie i potencjał **60**
1.4. **Piotr Żuber, Stanisław Bienias** System ewaluacji w Polsce – dotychczasowe doświadczenia i wyzwania na przyszłość **84**

CZĘŚĆ DRUGA

ZARYS METODOLOGII I ORGANIZACJI BADAŃ **99**

2.1. **Karol Olejniczak** Ewaluacja jako proces badawczy **101**
2.2. **Jarosław Górniak, Karolina Keler** Rola systemów wskaźników w ewaluacji **109**
2.3. **Sławomir Mandes** Metody jakościowe w ewaluacji **129**
2.4. **Łukasz Widła** Zastosowanie metod ilościowych w ewaluacji **158**
2.5. **Krzysztof Piech** Ewaluacja oddziaływania funduszy unijnych – teoria i praktyka modelowania makroekonomicznego **179**
2.6. **Agnieszka Haber** Zarządzanie badaniem ewaluacyjnym – perspektywa zlecającego **212**
2.7. **Bartosz Ledzion** Zarządzanie kontraktem ewaluacyjnym – perspektywa wykonawcy **231**

CZĘŚĆ TRZECIA

BADANIE EWALUACYJNE – POMOC W PLANOWANIU **249**

3.1. **Bartosz Ledzion** Wprowadzenie. **251**

3.2. **Paweł Kościelecki** Ewaluacja *ex ante* na poziomie projektu. Studium przypadku projektu Muzeum Regionalnego w Stalowej Woli. **253**

3.3. **Bartosz Ledzion, Elżbieta Kozłowska** Studium przypadku Ewaluacja *ex ante* Programu Operacyjnego Infrastruktura i Środowisko 2007–2013 ..... **274**

3.4. **Adam Płoszaj** Ewaluacja *ex ante* RPO Lubuskie 2007–2013 ..... **284**

3.5. **Agnieszka Sternik** Ewaluacja *ex ante* Programu Operacyjnego „Zasoby Ludzkie i Zatrudnienie 2007–2013 w Czechach". **302**

CZĘŚĆ CZWARTA

BADANIE EWALUACYJNE – POMOC W ZARZĄDZANIU I BUDOWANIU PARTNERSTWA ... **311**

4.1. **Bartosz Ledzion** Wprowadzenie. **313**

4.2. **Wojciech Pander, Elżbieta Kozłowska** Ewaluacja bieżąca Programu Środki Przejściowe (*Transition Facility*) w Polsce **315**

4.3. **Anna Domaradzka** Ewaluacja *on-going* II priorytetu ZPORR – identyfikacja barier w systemie wdrażania **331**

4.4. **Maciej Szałaj, Bartosz Ledzion** Ewaluacja bieżąca sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego. **349**

4.5. **Karol Olejniczak** Ewaluacja bieżąca programów INTERREG IIIA Austria – nowe kraje członkowskie. **361**

CZĘŚĆ PIĄTA

BADANIE EWALUACYJNE – W POSZUKIWANIU EFEKTÓW I SZERSZYCH WNIOSKÓW **381**

5.1. **Bartosz Ledzion** Wprowadzenie. **383**

5.2. **Magdalena Kuśmierz** Ewaluacja *ex-post* projektu systemowego KPRM w zakresie szkoleń, realizowanego w ramach Działania 2.4. SPO RZL. **385**

5.3. **Martin Ferry, Karol Olejniczak** Ewaluacja *ex post* PHARE-CBC Polska-Niemcy 1999–2001 **402**

5.4. **Agnieszka Haber, Jacek Pokorski** Ewaluacja *ex post* programu Phare 2003 SSG komponent MSP **419**

5.5. **Grzegorz Gorzelak, Adam Płoszaj** Ewaluacja z długą perspektywą – przykład Programu Inicjatyw Lokalnych **438**

5.6. **Martin Ferry** Studium przypadku z Wielkiej Brytanii: Średniookresowa Ewaluacja Specjalnego Programu Przejściowego *Highlands & Islands* 2000–2006 ..... **450**

CZĘŚĆ SZÓSTA

STRATEGICZNE OBSZARY TEMATYCZNE BADAŃ EWALUACYJNYCH **465**

6.1. **Bartosz Ledzion** Wprowadzenie **467**

6.2. **Michał Wolański, Elżbieta Kozłowska** Rozwój i modernizacja infrastruktury **468**

6.3. **Michał Miedziński** Wybrane zagadnienia ewaluacji polityki innowacyjnej **480**

6.4. **Agnieszka Sternik** Ewaluacja interwencji w obszarze rozwoju zasobów ludzkich **499**

6.5. **Grzegorz Gorzelak, Dominika Wojtowicz** Ocena wpływu NPR i NSRO na rozwój regionalny i przestrzenny oraz zwiększenie spójności regionalnej i przestrzennej **517**

Marek Kozak, Karol Olejniczak, Bartosz Ledzion PODSUMOWANIE: PERSPEKTYWY I DYLEMATY EWALUACJI **531**

INFORMACJA O AUTORACH **535**

SŁOWO WSTĘPNE

Ewaluacja odgrywa znaczącą rolę w planowaniu i zarządzaniu programami w obszarze rozwoju społeczno-ekonomicznego. Pozwala gromadzić wiedzę i dostarcza informacji na temat efektów oraz szerszego oddziaływania polityki prowadzonej w tym zakresie. Jej zadaniem jest także wzmacnianie odpowiedzialności za wykorzystanie środków publicznych, wyjaśnianie mechanizmów działania programów oraz upowszechnianie wiedzy na temat poziomu ich skuteczności i efektywności. W polityce spójności Unii Europejskiej ewaluacja wspiera strategiczny rozwój tej polityki oraz wywiera pozytywny wpływ na jakość programów. Dodatkowo, ewaluacja to narzędzie, które Komisja Europejska wykorzystuje w celu skuteczniejszego informowania społeczeństwa na temat „wartości dodanej" Unii Europejskiej.

Ewaluacja to dynamiczny proces, który podlega zmianom w odpowiedzi na wyzwania wynikające z realizacji strategii i wdrażania programów. W polityce spójności Unii Europejskiej, dotychczasowe podejście do ewaluacji, bazujące na sztywnym kalendarzu szeroko zakrojonych badań średniookresowych, zostało zastąpione w nowym okresie programowania 2007–2013 ewaluacjami bieżącymi, dostosowanymi do potrzeb. Kraje członkowskie mogą obecnie definiować zakres ewaluacji, decydować o najbardziej odpowiednim, z ich punktu widzenia, terminie przeprowadzenia badania oraz ulepszać wykorzystywane metody badawcze. Większa odpowiedzialność państw członkowskich w zakresie ewaluacji umożliwia jej lepszą integrację z cyklem wdrażania programów oraz wykorzystanie jej wyników jako jednego z kluczowych narzędzi zarządzania wsparciem.

Oddawany do Państwa rąk podręcznik *Teoria i praktyka ewaluacji interwencji publicznych* daje wspaniałą możliwość zgłębienia praktyki i warsztatu ewaluacji. Wierzymy, że publikacja ta wesprze dalszy rozwój potencjału polskiej administracji, a także wzmocni rolę ewaluacji w polityce spójności Unii Europejskiej.

Veronica Gaffey
Naczelnik Wydziału do Spraw Ewaluacji,
Dyrekcja Generalna ds. Polityki Regionalnej,
Komisja Europejska

Antonella Schulte-Braucks
Naczelnik Wydziału do Spraw Ewaluacji i Oceny Skutków Regulacji,
Dyrekcja Generalna ds. Zatrudnienia, Spraw Społecznych i Równych Szans,
Komisja Europejska

Karol Olejniczak
Marek Kozak
Bartosz Ledzion

WPROWADZENIE

Niniejsza publikacja jest pierwszą w Polsce próbą ujęcia tematyki ewaluacji interwencji publicznych współfinansowanych ze środków Unii Europejskiej w formie podręcznika akademickiego.

Głównym celem, jaki przyświecał całemu zespołowi autorskiemu, było pokazanie ewaluacji w sposób całościowy – jako sztuki zadawania pytań i uczenia się instytucji sektora publicznego. Ten proces uczenia się dzięki badaniom ewaluacyjnym jest jednym z kluczowych elementów racjonalizowania i podnoszenia skuteczności programów publicznych – rdzenia współczesnego zarządzania publicznego. Wierzymy, że uczenie się – na bazie własnych i cudzych doświadczeń – oraz wyciąganie wniosków z podjętych działań prowadzą wprost do trwałego rozwoju społeczno-gospodarczego, a w konsekwencji – podnoszenia jakości życia.

Publikację kierujemy do osób zainteresowanych współczesną teorią i praktyką ewaluacji programów rozwoju społeczno-gospodarczego – zarówno pracowników administracji publicznej, naukowców, prywatnych konsultantów wykonujących badania ewaluacyjne, jak i osób zainteresowanych wkroczeniem w bogatą tematykę badań interwencji publicznych – studentów administracji publicznej, socjologii, regionalistyki.

Staraliśmy się wyjść poza utarte schematy poradników wykonywania ewaluacji „krok po kroku" i nadać naszej publikacji charakter bardziej akademicki. Odwołujemy się więc do szerszych teorii i paradygmatów współczesnych nauk społecznych oraz prezentujemy najnowsze teorie, podejścia, innowacje w polu badań ewaluacyjnych, które pojawiają się w bogatej międzynarodowej literaturze przedmiotu. Nasze refleksje nad użytecznością ewaluacji osadziliśmy jednak głęboko w praktyce programów funduszy strukturalnych, odnosząc nakreślone modele teoretyczne do praktycznych studiów przypadków (głównie krajowych).

Publikacja jest więc też pierwszą ustrukturalizowaną analizą polskich doświadczeń z konkretnych ewaluacji.

Podręcznik został podzielony na sześć części.

Celem pierwszej części jest wprowadzenie Czytelników do zagadnień ewaluacji. W rozdziale 1.1, autorstwa Karola Olejniczaka, przedstawione zostały definicje, typy i funkcje ewaluacji w programach sektora publicznego. Zasygnalizowano także pewne napięcia, wpisane w zróżnicowane role, jakie ewaluacja może odegrać w sektorze publicznym. W rozdziale tym znalazł się również opis modelu ewaluacji jako procesu produkcji i wykorzystania wiedzy. Model ten jest rdzeniem niniejszej książki. Posłużył jako punkt odniesienia w kolejnych rozdziałach, zaś dla rozdziałów zawierających studia przypadków stał się ich dokładną „mapą drogową", wytyczającą kolejne elementy analizy danego przypadku.

W rozdziale 1.2 Marek Kozak przedstawia szersze tło zagadnień ewaluacji programów polityki spójności Unii Europejskiej, jako szczególnie istotnej w dzisiejszej polityce rozwoju Unii i Polski. Polityka ta nie ma charakteru samoistnego: jest częścią i narzędziem osiągania ogólnych celów rozwojowych Unii. Bez znajomości jej założeń, celów, głównych instrumentów i doświadczeń, trudno sobie wyobrazić poprawną ocenę długofalowych skutków tej polityki, zwłaszcza badania jej odpowiedniości wobec zmierzeń oraz współczesnego paradygmatu rozwoju. Rozdział ten przedstawia zwięźle ewolucję polityki spójności, ukazując ją na tle problemów rozwojowych, właściwych różnym etapom jej rozwoju oraz powiązań z innymi politykami. Uwzględnia też relacje między polityką spójności a Unią Gospodarczą i Walutową oraz kolejnymi poszerzeniami Unii.

Rozdział 1.3, autorstwa Johna Bachtlera, jest syntezą europejskich doświadczeń w polu ewaluacji programów rozwoju regionalego. Autor wprowadza Czytelników w genezę badań ewaluacyjnych oraz omawia – odwołując się do bogatej literatury i badań empirycznych – główne czynniki ich rozwoju w Unii Europejskiej. Dodatkową wartością rozdziału jest przybliżenie polskim czytelnikom zróżnicowanych praktyk ewaluacji w siedmiu wybranych krajach UE.

Ostatni w tej części rozdział 1.4, autorstwa Piotra Żubera oraz Stanisława Bieniasa, wprowadza nas w zagadnienia kultury ewaluacji w Polsce. Autorzy omówili w nim zarówno genezę, jak i obecną krajową praktykę ewaluacji, zwracając uwagę, że w Polsce dominującym polem praktyki ewaluacji są fundusze strukturalne i polityka spójności. W odniesieniu do tego właśnie pola – programów polityki spójności Unii Europejskiej – przeprowadzili szczegółową, spójną analizę systemu i jego obecnego potencjału, nakreślili także perspektywy i przyszłe wyzwania.

Celem części drugiej podręcznika jest zarysowanie podstaw metodologicznych i organizacyjnych badań ewaluacyjnych. Rozdział 2.1, autorstwa Karola Olejniczaka, przedstawia główne elementy ewaluacji jako procesu badawczego. Rozdział 2.2. – autorstwa Jarosława Górniaka

i Karoliny Keler – omawia zagadnienie kluczowe dla skutecznej analizy efektów programów publicznych – wskaźniki. W rozdziale nakreślono najważniejsze pojęcia oraz kwestie fundamentalne dla budowy systemów wskaźników. Omówiono także zagadnienia użycia gotowych systemów wskaźników w ewaluacji.

Kolejne trzy rozdziały dotyczą zagadnień metodologicznych. Nie mają one jednak na celu szczegółowego omówienia całości bogatej praktyki metod badań społecznych (co ze względu na formę i objętość publikacji byłoby niemożliwe), lecz przybliżenie najważniejszych pojęć, paradygmatów, a także pokazanie zalet i ograniczeń poszczególnych warsztatów metodologicznych w odniesieniu do ewaluacji. I tak, rozdziały 2.3 (Sławomira Mandesa) i 2.4 (Łukasza Widły) omawiają funkcje, zalety i ograniczenia dwóch odrębnych, choć wzajemnie uzupełniających się warsztatów ewaluatora – metod jakościowych oraz metod ilościowych. Natomiast w rozdziale 2.5 Krzysztof Piech przybliża Czytelnikom w sposób spójny i przejrzysty trudną i szeroką tematykę wykorzystania modeli ekonometrycznych, odwołując się do dotychczasowej praktyki i perspektyw użycia tego narzędzia w Polsce.

Ostatnie dwa rozdziały części drugiej podręcznika są ciekawym dwugłosem pokazującym praktykę zrządzania kontraktem ewaluacyjnyh z perspektywy zlecającego (rozdział 2.6 autorstwa Agnieszki Haber) oraz wykonawcy (rozdział 2.7 autorstwa Bartosza Ledziona).

Kolejne trzy części książki przedstawiają studia przypadków przebiegu i użyteczności konretnych badań ewaluacyjnych. Pogrupowaliśmy je według trzech modelowych funkcji ewaluacji. I tak, część trzecia zawiera studia ewaluacji, których głównym celem było podniesienie jakości planów programów publicznych. W części czwartej przedstawiamy cztery przykłady ewaluacji nakierowanych na pomoc w bieżącym zarządzaniu i budowaniu partnerstwa w ramach funkcjonujących programów. Natomiast w części piątej przedstawiamy historie badań, w których starano się oszacować szersze efekty i oddziaływanie interwencji publicznych.

Chcemy podkreślić, że prezentowane przypadki nie są „najlepszymi praktykami". Kluczem wyboru była dla nas tematyka studiów (przedstawienie jak najszerszej gamy tematów interwencji publicznych, ich formy i skali oraz różnych okresów wdrażania), a także ich wartość jako ilustracji procesu uczenia się instytucji i jednostek. Często proces ten był trudny, nie zawsze całkowicie udany, a na pewno zawsze wymagający wyborów i zaangażowania tak ze strony zlecających, jak i ewaluatorów. Aby studia przypadków były dla Czytelników przejrzyste, prezentujemy je, odnosząc się do konkretnego modelu analitycznego (przedstawionego w pierwszej części książki). Dzięki temu można prześledzić historie wyzwań uczenia się z doświadczeń ewaluacji w sposób spójny i ustrukturalizowany, a także rozwijać własne porównania i analogie między poszczególnymi przypadkami.

Część szósta podręcznika zawiera teksty o szerszym, eseistycznym charakterze. Sygnalizują one ograniczenia bieżących praktyk ewaluacyjnych, a także zarysowują wyzwania filozoficzne, koncepcyjne i metodologiczne w wybranych, głównych polach tematycznych ewaluacji polityki spójności.

Mamy nadzieję, że teksty będą przyczynkiem do dyskusji nad przyszłymi kierunkami rozwoju paradygmatów badań ewaluacyjnych.

W zakończeniu książki staramy się zarysować główne wyzwania i perspektywy ewaluacji programów publicznych w Polsce na najbliższe lata. Przygotowując wnioski, opieraliśmy się zarówno na refleksjach zawartych w przekrojowych rozdziałach książki, doświadczeniach z przedstawianych studiów przypadków, jak i na obserwacjach z praktyki szerokiego zespołu autorów.

Praca powstała dzięki wytężonej pracy zespołu złożonego z ekspertów o uznanych osiągnięciach w zakresie zarówno teorii, jak i metodologii i praktyki ewaluacji, o doświadczeniach zdobytych nie tylko w Polsce, ale również za granicą, na różnych poziomach badań ewaluacyjnych. Mamy nadzieję, że niniejszą publikację – wspólne dzieło 27 autorów, uznają Państwo za wartościowe źródło informacji i inspiracji – zarówno do dalszych badań, jak i praktycznych pomysłów na realizowane przez Państwa ewaluacje programów publicznych.

Mamy także nadzieję, że mimo niewątpliwych ograniczeń niniejszej książki spotka się ona z zainteresowaniem i życzliwością. Będziemy wdzięczni za wszelkie spostrzeżenia i komentarze, które postaramy się – wedle naszych możliwości – uwzględnić w przyszłych edycjach podręcznika.

CZĘŚĆ PIERWSZA

ZAGADNIENIA OGÓLNE

ROZDZIAŁ 1.1.

Karol Olejniczak

WPROWADZENIE DO ZAGADNIEŃ EWALUACJI

1. WSTĘP

Ewaluacja zaczęła się rozwijać już w latach 30. XX wieku w USA, w polu edukacji (Guba, Lincoln 1989: 27; Worthen 1990)[1]. Przełomem były jednak lata 50. i 60., gdy wraz z rosnącą liczbą, skalą i poszerzaniem zakresu interwencji publicznych, ewaluację zaczęto wykorzystywać na innych, niż edukacja, polach (np. programów ochrony zdrowia, rozwoju regionalnego, infrastruktury, rozwoju zasobów ludzkich, programów socjalnych).

Ponad trzy dekady praktyki ewaluacji interwencji publicznych (rozwijanej głównie w administracjach krajów wysoko rozwiniętych) zaowocowały bogatym i zróżnicowanym zbiorem interpretacji, definicji, typologii i praktycznych podejść[2].

---

[1] Te edukacyjne korzenie ewaluacji widać również, mimo że dopiero pół wieku później (w latach 90.), także w polskim przypadku. Pierwszy polscy autorzy piszący o ewaluacji umieszczali ją właśnie w kontekście programów edukacyjnych (porównaj: Korporowicz 1997).

[2] Najbardziej dynamicznie rozwijała się ewaluacja w Stanach Zjednoczonych i Kanadzie. W Europie silną kulturą ewaluacji charakteryzują się Wielka Brytania, Szwecja i Niemcy. Szerzej o potencjale ewaluacyjnym poszczególnych krajów UE traktuje rozdział 1.3 autorstwa Johna Bachtlera.

Niniejszy rozdział jest próbą usystematyzowania głównych pojęć i koncepcji, do których będziemy się odwoływać w dalszej części książki. W rozdziale odnosimy się do czterech kluczowych kwestii. Pierwszą jest termin „ewaluacja". Przedstawimy główne definicje spotykane zarówno w literaturze tematu, jak i w praktyce organizacji międzynarodowych. Na tej podstawie proponujemy definicję ewaluacji wykorzystaną w dalszych rozdziałach książki.

Drugą kwestią, której poświęcamy uwagę, jest przegląd typologii ewaluacji, uwzględniający kilka perspektyw – perspektywę przedmiotu, celu badania, momentu oraz sposobu organizacji badania. Tutaj też omawiamy miejsce ewaluacji w cyklu wdrażania interwencji publicznych.

Trzecim elementem naszej analizy są funkcje, jakie ewaluacja może spełniać w sektorze publicznym. Wprowadzamy Czytelnika w praktyczne wybory (podejścia badawczego i role uczestników), wynikające z orientacji konkretnego badania na daną funkcję, którą ma spełnić.

Wreszcie na zakończenie rozdziału koncentrujemy się na motywie przewodnim niniejszej książki – ewaluacji jako procesie uczenia się. Proponujemy model analityczny, pozwalający śledzić i weryfikować skuteczność procesu produkcji i wykorzystania wiedzy w prowadzonych ewaluacjach. Model ten wykorzystujemy następnie w rozdziałach poświęconych studiom przypadków (rozdziały 5.2–7.4.).

2. DEFINICJE EWALUACJI

Termin ewaluacja (ang. *evaluation*, franc. *évaluation*) oznacza „oszacowanie wartości" (Oxford University Press 2002). W praktyce polskiej administracji połowy lat 90. termin ten jednak zniekształcono, tłumacząc go jako „ocena". Tymczasem ewaluacja to proces, w którym „ocena" jest tylko jednym z elementów. Ponadto, jak zauważa Stern (2005a: 23), ewaluacja to także tworząca się dyscyplina, dziedzina wiedzy, łącząca nauki stosowane z praktyką polityk publicznych[3].

[3] Błąd w tłumaczeniu popełniony w połowie lat 90. (w ramach grantu Banku Światowego na przygotowanie systemu ewaluacji w Polsce), powielono i przeniesiono do języka dokumentów prawnych (np. ustawa o Narodowym Planie Rozwoju 2004–06; zamówienia przetargowe organizowane przez ministerstwa, tłumaczenia Rozporządzenia UE na temat funduszy strukturalnych itp.). Tymczasem w codziennej praktyce – również instytucji rządowych – używa się oryginalnego terminu „ewaluacja" (np. konferencje ewaluacyjne organizowane przez PARP i MRR, poradniki ewaluacji, roczne plany ewaluacji przedstawiane przez ministerstwa i wdrażane przez ich jednostki ewaluacyjne). W rezultacie w Polsce wciąż utrzymuje się zamieszanie terminologiczne. Nadzieje na uporządkowanie tego stanu rzeczy są nikłe – w tłumaczeniach rozporządzeń UE na lata 2007–2013 beztrosko skopiowano, idąc za latami 2004–2006, tłumaczenie terminu *evaluation* jako „ocena", a nie „ewaluacja".

Jako punkt wyjścia dla naszych rozważań proponujemy więc przyjąć, iż „ewaluacja" to **zróżnicowany zestaw metod i działań zorientowany na krytyczną refleksję nad wartością i jakością interwencji publicznych – zarówno procesów ich wdrażania, jak i ich efektów**.

W literaturze przedmiotu silnie eksponuje się pięć specyficznych kwestii dotyczących pojęcia ewaluacji.

Po pierwsze, jest to analityczny i usystematyzowany (ang. *systematic*) charakter ewaluacji. „Krytyczna refleksja" oznacza oparcie się na zestawie podejść i metod badawczych z nauk społecznych, takich jak stosowanie podejścia naukowego, analizy opartej na danych, przestrzeganie rygoru badawczego etc. (Langbein, Felbinger 2006: 33; Patton 2004; Rossi et al. 2004; Scriven 1991: 131). Aby ocenić wartość programów publicznych, najpierw trzeba zrozumieć i wyjaśnić mechanizmy leżące u ich podstaw oraz możliwe alternatywy. Według Sterna ten „wyjaśniający" aspekt ewaluacji sprawia, że przenika się ona z głównym nurtem nauk społecznych (Stern 2005a: 22).

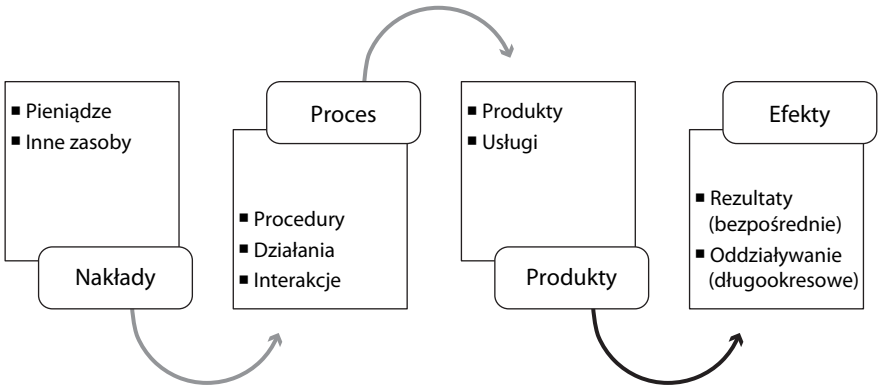
Elementem drugim, bezpośrednio związanym z pierwszą kwestią, jest wychodzenie poza kanon jednej metodologii i łączenie różnych narzędzi oraz źródeł. Ten specyficzny „eklektyzm" metodologiczny (określany zwykle terminem *mix-method designs*) jest szczególnie podkreślany w najnowszej literaturze i praktyce ewaluacji. Ewaluacja nie jest więc związana czy zdominowana przez wyłącznie jedną filozofię i jedno podejście badawcze, lecz wzajemnie uzupełnia i wykorzystuje potencjał różnych pól i technik. Wybór jest podyktowany zwykle bieżącymi potrzebami i przedmiotem badania, a nie ustalony *a priori* (Greene et al. 1989; Prime Minister's Strategy Unit 2003: 3–4).

Po trzecie, coraz częściej uwypuklana jest też systemowość ewaluacyjnej refleksji (po części wynikająca z przestrzegania wiarygodnego warsztatu badawczego). Oznacza to zarówno całościowe podejście – uwzględniające różne perspektywy i aspekty analizowanej interwencji – jak i cykliczność oraz systematyczność analizy (podejmowanie jej na różnych etapach cyklu wdrażania interwencji; szerzej o tym w kolejnym podpunkcie).

Po czwarte, podkreśla się, że ewaluacja idzie dalej niż badania naukowe, oceniając zarówno jakość danej interwencji (ang. *merit – quality*), jak i jej wartość (ang. *worth – value*) (Davidson 2005: 1–2; Paddock 1998). To wartościowanie jest prowadzone w odniesieniu do konkretnych kryteriów i standardów (Stern 2005a) które, co jednak należy podkreślić, mogą być nieco odmiennie definiowane i interpretowane w poszczególnych badaniach.

Wreszcie ostatnim elementem podkreślanym w definicjach ewaluacji jest to, że badania ewaluacyjne mogą obejmować zarówno wstępną ocenę struktury interwencji (czyli logiczną weryfikację pomysłu na program),

ocenę bieżącego procesu wdrażania danego programu (procedur, działań), jak i oszacowane jego rzeczywistych efektów, a więc empiryczną weryfikację relacji przyczynowo-skutkowej między działaniami a efektami (Bingham, Felbinger 2002: 3; Stufflebeam, Shinkfield 2007: 8; Weiss 1998: 4)[4]. Tym samym, ewaluacja weryfikuje relacje modelu logicznego, będącego rdzeniem współczesnych programów publicznych. Przedstawiamy go na rysunku 1 (szerzej o tym w rozdziale 2.2 Jarosława Górniaka i Karoliny Keler)[5].



Rysunek 1. Model logiczny interwencji publicznych

Źródło: Frechtling 2007: 22; Keehley, Abercrombie 2008: 32.

W praktyce organizacji międzynarodowych i instytucji sektora publicznego ewaluacja jest definiowana przez pryzmat jej praktycznego celu – poprawy jakości i wartości interwencji publicznych (polityk, programów, projektów). Z reguły pojawia się więc jako element efektywnego zarządzania sektorem publicznym czy wprowadzanego ostatnio terminu „dobrego rządzenia" (ang. *good governance*).

---

[4] Użyte tu przez nas pojęcie efektów (ang. *outcomes*) obejmuje zarówno rezultaty (ang. *results*), czyli efekty bezpośrednio wynikające z interwencji i pojawiające się tuż po jej zakończeniu oraz odnoszące się bezpośrednio do adresatów pomocy, jak i oddziaływanie (ang. *impact*) – zmiany odczuwalne dla szerszej populacji, wynikające z interwencji i możliwe do zaobserwowania w dłuższej perspektywie czasowej.

[5] Przedstawiony przez nas model jest jedną z uproszczonych interpretacji. W programach UE stosuje się zwykle następujący model: Potrzeby wynikające z kontekstu → Cele → Nakłady → Działania → Produkty → Rezultaty → Oddziaływanie (European Commission 2001; 2006: 7). Natomiast w literaturze spoza pola funduszy strukturalnych popularne są modele: CMO – Context → Mechanisms → Outcomes (Pedersen, Rieper 2008) czy CIPP – Context → Input → Process → Product (Mathison 2005: 60–65). Wszystkie opierają się jednak na tym samym rdzeniu – ciągu logicznym od nakładów przez działania (interwencję), po jej efekt.

OECD proponuje bardzo rozbudowaną definicję. Zawiera ona zarówno dyskutowane już powyżej elementy (usystematyzowanie, obiektywizm i wiarygodność), jak i cele ewaluacji (oszacowanie wartości, jakości, uczenie się), zróżnicowane przedmioty ewaluacji (programy lub projekty, procesy, efekty), kryteria oceny oraz sposoby wykorzystania wniosków (OECD, 2002).

Bank Światowy definiuje ewaluację jako: „użyteczne, wiarygodne i przejrzyste badania, które opierając się na obiektywnych podstawach oceniają wdrażanie i efekty polityk, programów, projektów. Tym samym przyczyniają się do wzmacniania odpowiedzialności działań [Banku], oraz pomagają wyciągać wnioski i doskonalić przyszłe interwencje". Ta definicja, podobnie jak poprzednia, wyraźnie zarysowuje funkcję rozliczeniowo-podsumowującą oraz funkcję uczenia się z ewaluacji (World Bank, 2008).

W dokumentach Unii Europejskiej ewaluacja jest definiowana dużo bardziej lakonicznie: jako „osąd wartości interwencji publicznej w odniesieniu do konkretnych kryteriów i standardów" (European Commission, 1999b: vol. 6), bądź też jako ocena działań publicznych w zależności od ich rezultatów, oddziaływania i potrzeb, które mają zaspokoić (European Commission, 2000).

Rozumienie ewaluacji w instytucjach Unii Europejskiej zmieniało się w kolejnych okresach programowania budżetowego. Dobrze ilustrują to kolejne rozporządzenia dotyczące polityki spójności. Początkowo ewaluacja nakierowana była na pomoc w planowaniu (używano nawet odrębnego terminu: ang. *appraisal*) oraz przedstawieniu oddziaływania programów (Council of The European Economic Community, 1993: art. 26). W kolejnym okresie (2000–2006) silnie podkreślono natomiast element bieżącego nadzoru i podnoszenia efektywności, łącząc wyniki ewaluacji z impulsami finansowymi – ewaluacja średniookresowa miała decydować o rozdysponowaniu przez Komisję rezerw finansowych (Council of The European Union, 1999: Art. 40–44). W nowym okresie programowania (2007–13) UE zdefiniowała ewaluację programów polityki spójności jako badania, które „mają na celu poprawę jakości, skuteczności i spójności pomocy funduszy oraz strategii i realizacji programów operacyjnych w odniesieniu do konkretnych problemów strukturalnych dotykających dane państwa członkowskie i regiony" (Rozporządzenie Rady WE, 2006: art. 47.1). Wyraźnie oddzielono też studia o charakterze strategicznym (których celem jest ocena ewolucji programu) od badań o charakterze operacyjnym (wspierających procesy i monitoring).

Część źródeł precyzuje pojęcie ewaluacji, zestawiając je z kontrolą, monitoringiem, audytem czy badaniami naukowymi (European Commission, 2004a: 10–11; Kierzkowski 2002: 15–17). Szczegółowe różnice między badaniem *stricte* naukowym, ewaluacją i innymi formami

dyscyplinowania sektora publicznego (monitoringiem, audytem) zostały przeanalizowane przez Olejniczaka (2007: 363–365). W tym miejscu warto uwypuklić jedynie trzy kwestie.

Po pierwsze, kluczowym elementem, który odróżnia ewaluację od badań *stricte* naukowych jest właśnie połączenie empirii i diagnozy (elementu wyjaśniającego) z charakterem normatywnym (osądem o wartości danej interwencji). Ewaluacja idzie więc krok dalej niż badania naukowe i zawsze podejmuje ocenę na podstawie sumiennie zebranych (choć często ograniczonych) dowodów.

Po drugie, ocena ta jest jednak dużo bardziej „miękka" niż w przypadku audytu czy kontroli. Nie wiąże się też z negatywnymi konsekwencjami i wyciąganiem odpowiedzialności dla osób i instytucji będących przedmiotem badania. Co więcej, w odróżnieniu od audytu, ewaluacja ma silny element interaktywności i współpracy między badanymi a ewaluatorami. Sama ocena przy niektórych podejściach ewaluacyjnych jest zresztą prowadzona wspólnie z interesariuszami.

Po trzecie, rdzeniem ewaluacji jest jej utylitaryzm. Przeprowadzona diagnoza sytuacji i ocena danego programu są więc podstawą do postawienia rekomendacji (będących integralną częścią badania ewaluacyjnego). Ewaluacja ma za zadanie zdiagnozować czy ocenić sytuację, a następnie wspomagać konkretnych odbiorców w rozwiązaniu problemów (Stufflebeam 2001: 11), wywołaniu pozytywnych zmian i poprawie jakości tak bieżącej, jak i przyszłych, analogicznych interwencji (Weiss 1998: 5). Ewaluacja jest więc pomyślana jako proces uczenia się ludzi i instytucji. Od połowy lat 80. za standard uznano, że dla efektywności tegoż procesu kluczowe są interakcje i negocjacje (Greene 2003: 592).

Podsumowując nasze rozważania na temat pojęcia ewaluacji i łącząc definicje zaproponowane w omawianych już źródłach, a także proponowane przez innych autorów (Forunier 2005; Górniak 2007: 11; Mathison 2005: 139–140; Olejniczak 2008: 21; Patton 2004; Rossi et al. 2004; Weiss 1998), proponujemy jej poniższą definicję.

Ewaluacja to systematyczne badanie, prowadzone z użyciem zróżnicowanych metod, złożone ze zbierania danych, analizy, oceny oraz informowania o wynikach. Jego celem jest oszacowanie (w odniesieniu do jasno sformułowanych kryteriów) jakości i wartości procesu oraz efektów wdrażania interwencji publicznych.

Elementy tej definicji będą rozwijane i dogłębnie omawiane na przykładach w kolejnych rozdziałach tej książki. W tym miejscu pozostaje nam jednak sprecyzować jeden kluczowy element. Chodzi o kryteria oceny.

Zdecydowaliśmy się przyjąć w tym miejscu perspektywę praktyki Unii Europejskiej, ponieważ nasze dalsze analizy (w tym studia przypadków)

odnoszą się do interwencji współfinansowanych z funduszy strukturalnych. W Unii Europejskiej wypracowany został standard kilku podstawowych kryteriów ewaluacji. Ich podsumowanie przedstawiamy w tabeli 1.

Do każdego z tych kryteriów możemy przypisać szereg normatywnych pytań ewaluacyjnych. Taką przykładową mapę pytań przedstawiliśmy na rysunku 2.

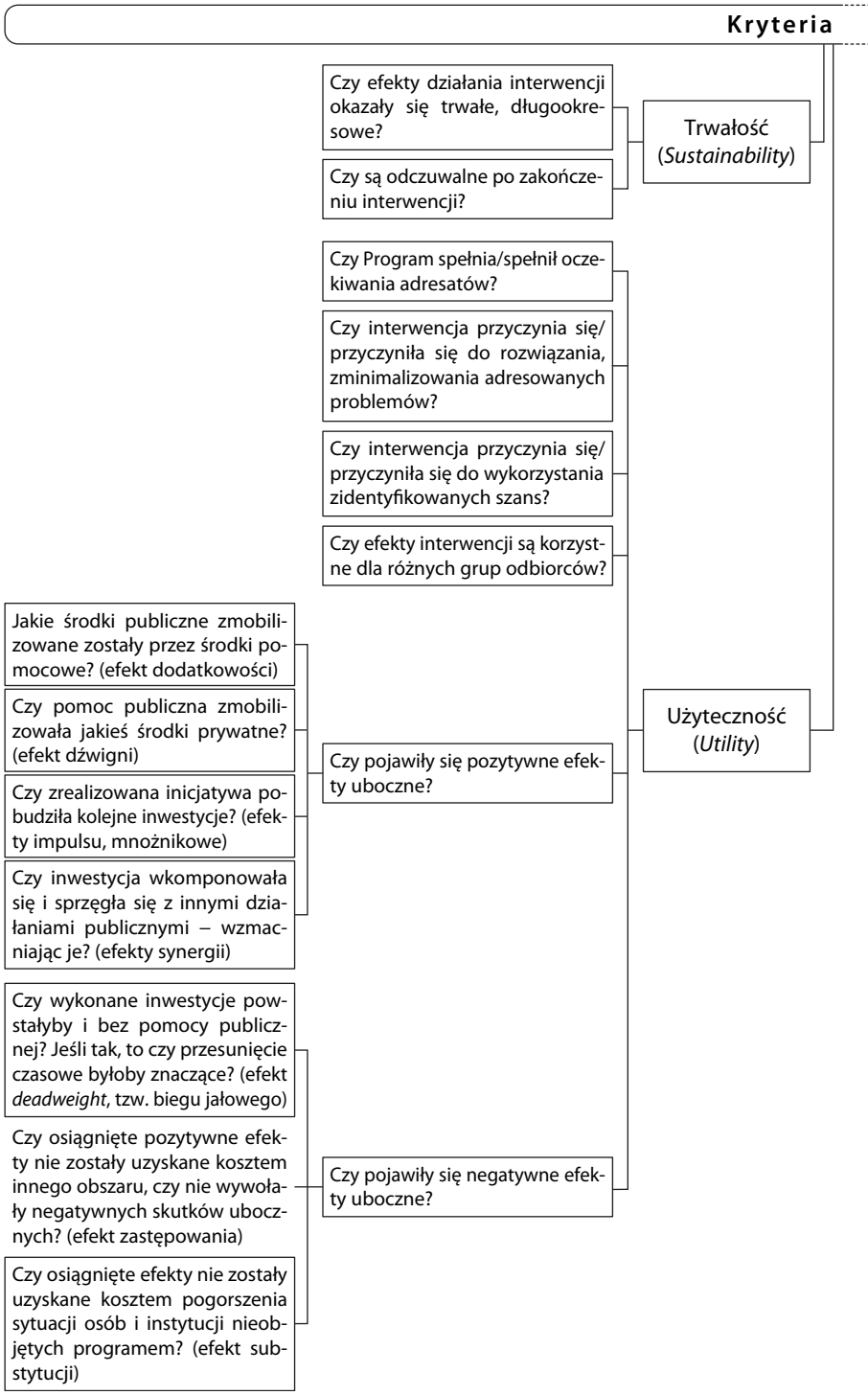
Oczywiście jest to jedynie mapa indykatywna, a nie wyczerpująca lista pytań. Po pierwsze, pomija ona wszelkie typy pytań badawczych prowadzące do pytań normatywnych (a więc pytania opisowe, przyczynowo-skutkowe)[6]. Po drugie, są to pytania szablonowe – pozbawione specyfiki tematycznej i czasowej konkretnych programów. Niemniej jednak może się ona okazać pomocna na etapie formułowania pytań ewaluacyjnych (szerzej o tym procesie piszemy w rozdziałach 2.6 i 2.7.

Tabela 1. Kryteria ewaluacji

Nazwa kryterium	Opis
Trafność (ang. *Relevance*)	Stawia pytania o adekwatność planowanych celów i metod wdrażania interwencji do problemów i wyzwań społeczno-ekonomicznych, które dana interwencja ma rozwiązać. Analiza tego kryterium jest prowadzona przed rozpoczęciem i w pierwszych fazach wdrażania interwencji.
Skuteczność (ang. *Effectiveness*)	Podnosi pytania o stopień realizacji zakładanych celów, skuteczność użytych metod, instytucji oraz wpływ czynników zewnętrznych na ostateczne efekty.
Wydajność (ang. *Efficiency*)	Koncentruje analizę na relacjach między nakładami, kosztami, zasobami (finansowymi, ludzkimi, administracyjnymi) a osiągniętymi efektami interwencji.
Użyteczność (ang. *Utility*)	Stawia pytania o rzeczywiste, osiągnięte efekty i ich adekwatność do sytuacji oraz wyzwań społeczno-ekonomicznych, będących przedmiotem interwencji. W odróżnieniu od kryterium trafności, analiza użyteczności jest prowadzona po zamknięciu interwencji lub w jej końcowej fazie wdrażania.
Trwałość (ang. *Sustainability*)	Obejmuje pytania o ciągłość efektów (przede wszystkim pozytywnych) danej interwencji w perspektywie średnio- i długookresowej.

Źródło: European Commission, 1999; 2007 za Olejniczak 2008: 93.

[6] Typy pytań ewaluacyjnych przedstawiamy w rozdziale 2.1.



Rysunek 2. Drzewo pytań ewaluacyjnych

ewaluacyjne

- Trafność (*Relevance*)
  - Czy cele interwencji odpowiadają zidentyfikowanym potrzebom?
  - Czy interwencja odnosi się do głównych sił i słabości danego układu (regionu, sektora, etc.)?
  - Czy obrana strategia jest spójna?
    - Czy strategia jest spójna wewnętrznie?
    - Czy strategia jest zbieżna z celami innych programów, działań, polityk?
  - Czy logika interwencji jest poprawna? (logika zależności przyczynowo-skutkowych)
  - Czy obrana na wstępie strategia pozostaje aktualna w trakcie wdrażania interwencji?
- Skuteczność (*Effectiveness*)
  - Czy cele interwencji zdefiniowane na etapie programowania są/zostały osiągnięte?
  - Jak silny jest/był wpływ czynników zewnętrznych?
  - Jakie są/były sukcesy interwencji?
  - Jakie problemy napotyka/napotkała interwencja?
  - Czy wybrane instrumenty i rozwiązania okazały się odpowiednie do zidentyfikowanych problemów?
  - Czy system zarządzania i wdrażania okazał się skuteczny?
  - Czy poszczególne instytucje poprawnie wypełniają/wypełniały przypisane im funkcje?
- Wydajność (*Efficiency*)
  - Czy podobne efekty można/można było osiągnąć przy wykorzystaniu innych instrumentów?
  - Czy podobne efekty można/można było osiągnąć przy wykorzystaniu niższych nakładów finansowych?

Tabela 2. Typologie ewaluacji

Kryterium kategoryzacji	Wyróżniane typy
Przedmiot ewaluacji	Ewaluacja projektu – koncentruje się na analizie pojedynczej, małej interwencji (z reguły w polu innowacji lub zasobów ludzkich).
	Ewaluacja programu (najczęściej spotykany typ) – przedmiotem badania może być zarówno pojedynczy program sektora publicznego, jak i cały pakiet interwencji (np. działanie, priorytet).
	Ewaluacja polityk – częściej nazywana analizą polityk (ang. *policy analysis*) – składa się z szeregu badań i analiz (zwykle na przestrzeni wielu lat).
	Ewaluacje tematyczne/horyzontalne (ang. *thematic/horizontal evaluation*) – ich przedmiotem jest konkretna kwestia zarządzania lub aspekt programu (np. zastosowanie zasady partnerstwa); takie badania są prowadzone przekrojowo – dla kilku, kilkunastu interwencji.
	Meta-ewaluacje (ang. *Meta-evaluation*) – badania, których przedmiotem są inne, wcześniejsze ewaluacje. Są to analizy porównawcze, podsumowujące obserwacje, wnioski i rekomendacje kilkunastu badań, oraz oceniające zastosowane techniki, poprawność metodologii.
Cel ewaluacji	Ewaluacja formatywna (operacyjna) – oceniająca formę, plan oraz proces wdrażania danej interwencji przed jej rozpoczęciem lub w czasie jej trwania. Koncentruje się na doradzeniu zarządzającym programem, jakie cele ustalić, lub jak poprawić płynność wdrażania interwencji i podnieść jej ostateczną jakość.
	Ewaluacja podsumowująca (strategiczna) – podsumowuje dobiegającą końca lub zakończoną interwencję, jej rzeczywiste efekty, skalę zmian wywołanych interwencją. Jest podejmowana po zakończeniu lub w końcowej fazie interwencji. Ma charakter opisowy i wyjaśniający, jest adresowana do szerszej publiki.

Moment podjęcia badania	Ewaluacja wstępna (*ex ante*) – prowadzona przed rozpoczęciem interwencji. Jej celem jest zweryfikowanie struktury, logiki i uzasadnienia planowanych działań.
	Ewaluacja bieżąca (ang. *on-going*) wykonywana w trakcie wdrażania badanej interwencji; może być badaniem punktowym (mówimy wtedy o ewaluacji bieżącej) albo składać się z całego szeregu badań wykonywanych po sobie i realizowanych przez jeden zespół badaczy (ewaluacja ciągła). Może dotyczyć badania zarówno procesów, jak i efektów (jeśli program jest zaawansowany).
	Ewaluacja końcowa (*ex post*) – realizowana po zakończeniu interwencji (najczęściej po 2 latach od zakończenia wszystkich działań). Skupia się na ocenie rezultatów i długookresowych efektów (oddziaływania) programu, w tym trwałości uzyskanych pozytywnych zmian i identyfikowaniu ewentualnych efektów ubocznych.
Organizacja badania	Ewaluacja zewnętrzna – badanie zlecane poza instytucję prowadzącą daną interwencję i wykonywane przez zespoły niezależnych ekspertów wyłonione w ramach konkursu.
	Ewaluacja wewnętrzna – prowadzona przez niezależną komórkę ewaluacyjną, umieszczoną jednak w strukturach instytucji publicznej wdrażającej dane interwencje publiczne.
	Autoewaluacja – prowadzona przez osoby z danej instytucji odpowiedzialnej za wdrażanie analizowanej interwencji.

Źródło: opracowanie na podstawie Olejniczak 2008: 25–38.

3. TYPY EWALUACJI

W literaturze przedmiotu funkcjonuje wiele typologii ewaluacji (Greene 2003; Herman et al. 1987; Owen, Rogers 1999; Tavistock Institute et al. 2003). Zróżnicowanie wynika przede wszystkim z przyjmowania różnych kryteriów użytych do kategoryzacji. W niniejszej publikacji przyjęliśmy trzy standardowe wymiary kategoryzowania ewaluacji (podsumowujemy je w tabeli 2):

- ze względu na przedmiot ewaluacji (tzw. ewaluandę – ang. *evaluand*),
- ze względu na ogólny cel ewaluacji,
- ze względu na moment podjęcia badania,
- ze względu na administracyjną organizację badania.

Dodatkowego komentarza wymaga typologia oparta na momencie podjęcia badania. Jak zauważa Górniak (2007: 11), ewaluacja jest pomyślana jako integralny element doskonalenia polityk i programów publicznych. W założeniach Unii Europejskiej wymienione trzy typy ewaluacji wpisane zostały w tzw. cykl życia interwencji publicznych (European Commission, 1999, 2004b). Mają tym samym:

- racjonalizować plany interwencji na etapie ich projektowania (badania *ex ante*);
- wspomagać wdrażanie interwencji, weryfikując aktualność obranych strategii, rzeczywiście poprawiając procesy, a także sprawdzając osiągane cele i wywoływane zmiany (badania bieżące);
- oceniać rzeczywiste efekty zakończonych interwencji (a więc rzeczywistą skuteczność obranych strategii) i przez to wspierać decydentów w podjęciu decyzji o kontynuacji czy zaniechaniu konkretnej strategii (badania *ex post*).

Z tej perspektywy najważniejsza wydaje się więc ewaluacja *ex post*, gdyż to ona dostarcza prawdziwie empirycznych, całościowych dowodów o efektach i użyteczności interwencji publicznych, sile (czy w ogóle występowaniu) relacji przyczynowo-skutkowej między programem a rzeczywistą, pozytywną zmianą.

Jednak, jak zauważa Olejniczak (2008: 91), w praktyce programów UE rola ewaluacji *ex post* w kształtowaniu racjonalnych decyzji o kolejnych interwencjach jest znikoma; ogranicza się głównie do podsumowania i rozliczenia. Wynika to z faktu, że programy UE są planowane i wdrażane w sposób ciągły, podczas gdy szacowanie pełnych efektów wymaga odstępu czasowego (zwykle 2 lat) między zakończeniem interwencji a jej badaniem. Na przykład interwencje na lata 2007–13 były planowane już w 2005 i 2006 roku. Tymczasem ewaluacja *ex post* programów z lat 2000–2006 nastąpiła w 2008 roku, a więc w połowie kolejnego okresu programowania. Nie miała więc bezpośredniego przełożenia na planowanie programów 2007–2013.

Na zakończenie należy zaznaczyć, że powyższe typologie nie są rozłączne[7]. W niniejszej książce będziemy je pokazywali na przykładach – w części poświęconej studiom przypadków.

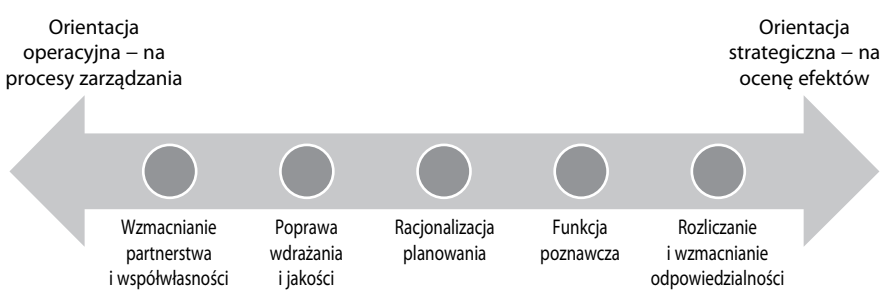
4. FUNKCJE EWALUACJI I DYLEMATY ICH WYBORU

Tak jak sygnalizowaliśmy już w podrozdziale dotyczącym definicji, głównym celem ewaluacji jest wspomaganie instytucji demokratycznych w doborze, nadzorowaniu, wartościowaniu i doskonaleniu programów i polityk publicznych (Mark et al. 2000: 3). Racjonalizując podejmowane działania publiczne (zarówno te poddawane badaniu, jak i przyszłe – analogiczne), ewaluacja przyczynia się więc do dobrobytu społecznego (ang. *social betterment*).

Tak szeroka rola ewaluacji może być opisana za pomocą wielu konkretnych funkcji (czasem określanych mianem celów ewaluacji). W literaturze przedmiotu przedstawiane są różne kategoryzacje i listy funkcji ewaluacji (Bachtler 2001; Batterbury 2006; Chelimsky 1997; European Commission 2007; Górniak 2007). Niezależnie od używanych terminów, we wszystkich tych typologiach można wyróżnić 3 główne motywy prowadzenia ewaluacji: pomoc w planowaniu, pomoc w bieżącym zarządzaniu oraz funkcję szerszego uczenia się (poznawania mechanizmów interwencji oraz jej otoczenia).

W niniejszej książce zdecydowaliśmy się przyjąć, za Ferry i Olejniczakiem (2009, w druku), pięć funkcji ewaluacji w sektorze publicznym. Przedstawiamy je w tabeli 3.

Dotychczas pisaliśmy ogólnie o funkcjach ewaluacji. Odnosząc je jednak do rzeczywistości konkretnych programów publicznych, trzeba wyraźnie stwierdzić, że jedno badanie ewaluacyjne nie może wypełniać wszystkich funkcji. Zarówno wnioski z badań empirycznych (Ferry, Olejniczak



Rysunek 3. Dylematy wyboru funkcji ewaluacji

Źródło: Olejniczak 2008.

[7] Współwystępowanie danych typów ewaluacji zostało omówione w książce Olejniczaka (2008: 37–40).

Tabela 3. Funkcje ewaluacji

Funkcja	Opis funkcji	Jak rozpoznam tę funkcję, gdy ją zobaczę?
Budowanie partnerstwa i współwłasności	■ wzmacnianie powiązań, relacji, współpracy, interakcji między uczestnikami programu; ■ aktywizowanie interesariuszy – włączanie ich w działania programu, dyskusję nad jego oceną i kierunkami rozwoju; ■ uprawnianie interesariuszy – włączanie we współdecydowanie o programie tych grup, które były marginalizowane.	■ dzięki badaniu włączono w program głosy nowych interesariuszy; ■ badanie zintensyfikowało współpracę i dialog między interesariuszami; ■ dzięki badaniu nawiązano kontakty poza programem (np. z instytucjami obsługującymi inne, podobne interwencje).
Poprawa wdrażania i bieżącej jakości	■ poprawienie istniejącej struktury, procedur i procesu zarządzania programem; ■ minimalizowanie negatywnych efektów ubocznych zarządzania opartego na wynikach; ■ podnoszenie efektywności dostarczania programu przez weryfikowanie, na ile program realizuje zakładane cele, jaka jest jakość produktów.	■ po badaniu zmieniono procedury, usprawniono procesy; ■ po badaniu zmieniono kompetencje instytucjonalne ■ wyniki badania pomogły zaprojektować procedury dla nowych programów.
Racjonalizowanie planowania	■ zagwarantowanie racjonalnego uzasadnienia dla polityki/programu przez zweryfikowanie jego spójności zewnętrznej, przystawalności celów do potrzeb i problemów; ■ zapewnienie wydajnej, racjonalnej i skutecznej alokacji zasobów przez zweryfikowanie logiki programu, racjonalności wyznaczonych celów, spójności zewnętrznej planowanej strategii.	■ wprowadzono zmiany w strategii badanego programu; ■ przesunięto środki pomiędzy celami danego programu; ■ wnioski z badania wykorzystano przy tworzeniu nowych strategii.

Funkcja poznawcza	■ wyjaśnianie czynników i procesów warunkujących działania programu, jego sukces lub porażkę (w tym i czynników ubocznych); ■ weryfikowanie teorii leżących u podstaw danych polityk, interwencji; ■ wyciąganie wniosków dla innych programów i promowanie dobrych wzorców; ■ budowanie zdolności ewaluacyjnych danych instytucji zaangażowanych w badanie, lub szerzej – systemu w danym regionie, kraju.	■ badanie wyjaśniło kontekst, mechanizmy programu – jego sukcesu lub porażki; ■ badanie odnosiło się i zweryfikowało teorie, założenia leżące u podstaw programu; ■ badanie prezentowało dobre praktyki; ■ badanie wzmocniło doświadczenie i zrozumienie ewaluacji przez zlecających i wykonawców.
Rozliczanie i wzmocnienie odpowiedzialności	■ informowanie szerszej opinii publicznej o skuteczności programu – na ile cele programu zostały osiągnięte; ■ informowanie szerszej opinii publicznej o wydajności podjętych działań – jak efektywnie użyto przypisanych zasobów publicznych.	■ zidentyfikowano, oceniono i przedstawiono efekty programu interesariuszom; ■ zidentyfikowano, oceniono i przedstawiono efekty programu Komisji Europejskiej i kluczowym decydentom; ■ zidentyfikowano, oceniono i przedstawiono efekty programu opinii publicznej.

■ Źródło: Bachtler 2001; Batterbury 2006; Chelimsky 1997; Tavistock Institute et al. 2003 za Ferry, Olejniczak 2009, w druku.

2009, w druku; Olejniczak 2008; Stern 2005b), jak i logika, pokazują konflikt między niektórymi z funkcji. Chodzi o wybór między pomocą w bieżącym zarządzaniu i analizie procesów, a funkcją rozliczeniową i poznaniem szerszych relacji przyczynowo-skutkowych działań i efektów.

Konieczność wyboru podyktowana jest kilkoma względami. Pierwszym są odmienne perspektywy czasowe. Orientacja na procesy (wdrażanie i wzmacnianie partnerstwa) skupia się na krótkookresowej perspektywie – „tu i teraz" – trwającej interwencji, jej sprawach operacyjnych (problemach we wdrażaniu, aktualności celów etc.). Silna jest też presja czasowa na wyniki badania (szybka diagnoza i sugestia korekt). Natomiast ocena efektów następuje w dłuższej perspektywie – po zakończeniu programu. Ma więc charakter strategiczny.

Kwestią drugą są odmienni adresaci badania oraz relacje badających i badanych. Orientacja operacyjna służy przede wszystkim operatorom danej, trwającej interwencji, zaś ewaluator ma pomóc menedżerom i interesariuszom w rozwiązaniu ich aktualnych problemów i wątpliwości. To wymaga skracania dystansu – ścisłej współpracy między zlecającym a wykonawcą, włączania zlecających w proces badawczy, wymiany informacji. Zbudowane w ten sposób zaufanie i dialog dają możliwość szczerej dyskusji o bieżących problemach, zidentyfikowania ich źródeł, skutecznej mediacji (gdy istnieje konflikt wśród interesariuszy) i wspólnego znalezienia rozwiązań.

Przy oszacowywaniu efektów odbiorcą wniosków jest szersza publika, zaś ewaluator odgrywa – wobec menedżerów programu – rolę oceniającego. To z kolei wymaga większego dystansu między badaczami a badanymi. Ponieważ ewaluatorzy dokonują oceny efektów zakończonych już interwencji (lub zakończonych projektów w ramach trwającej jeszcze, ale zaawansowanej interwencji), to badani postrzegają ich (i słusznie) jako osoby rozliczające efekty ich działań.

Z powyższego wynika kwestia trzecia – wybór odmiennych metod badawczych – i szerzej, podejść do badania.

W przypadku orientacji na procesy i kwestie operacyjne sprawdzają się metody jakościowe, wywiady, rozmowy, nawet spotkania animacyjne z interesariuszami. Najważniejszy jest dialog i wzajemne zrozumienie odmiennych perspektyw wielu interesariuszy. W przypadku oceny efektów szuka się „zewnętrznego" dla wdrażających punktu odniesienia i oceny, a także zweryfikowania relacji przyczynowo-skutkowej nie na podstawie opinii, lecz faktów i udokumentowanych zmian. Poszczególne metody i techniki układają się w szersze całości – tak zwane podejścia/modele ewaluacyjne. Można je uznać za odpowiednik strategii badawczej w badaniach naukowych[8]. W tym wypadku literatura

[8] Mark et al. (2000) uznają, że podejścia do ewaluacji (ang. *evaluation approaches, inquiry modes*) są elementem łączącym i racjonalizującym ogólne cele badania z konkretnymi metodami i technikami.

Tabela 4. Podejścia ewaluacyjne

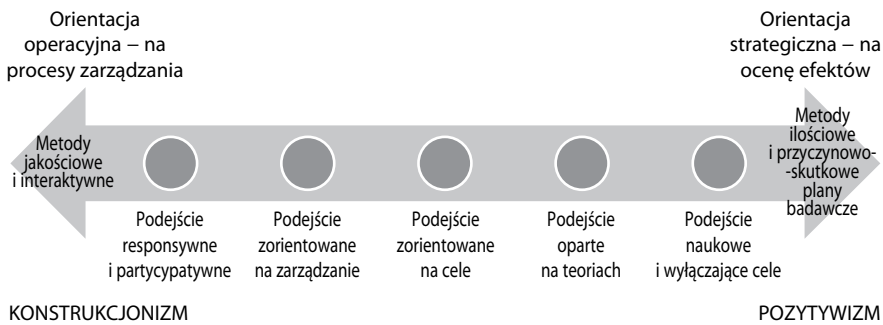
Podejście	Charakterystyka
Podejście naukowe i wyłączające cele programu	Analizuje rzeczywiste efekty programu (w tym efekty uboczne), bada relację przyczynowo-skutkową między działaniami a zmianami na danym obszarze, w grupie docelowej, w środowisku; przewaga metod ilościowych
Podejście oparte na teoriach	Analizuje teorie, założenia i mechanizmy programu w odniesieniu do jego efektów; mieszanka jakościowych i ilościowych metod, użycie modeli badawczych budowanych w oparciu na wcześniejszych, analogicznych ewaluacjach
Podejście zorientowane na cele	Koncentruje się na ocenie, na ile cele zostały zrealizowane – zestawia cele z produktami, rezultatami i oddziaływaniem na daną grupę docelową; przewaga metod ilościowych
Podejście zorientowane na zarządzanie	Koncentruje się na procesach wdrażania, pomiarze bieżącej jakości, usuwaniu blokad, minimalizowaniu konfliktów; przewaga metod jakościowych, użycie modeli logicznych
Podejście responsywne i partycypatywne	Koncentruje się na zidentyfikowaniu i zaspokojeniu bieżących potrzeb informacyjnych głównych interesariuszy, analizuje procesy programu i bieżące wdrażanie ze szczególnym uwzględnieniem roli aktorów procesu, wchodzi w animację współpracy między interesariuszami, budowanie dialogu, wzmacnianie komunikacji; dominujące metody jakościowe, techniki animacji dyskusji

Źródło: Alkin et al. 2006; Dahler-Larsen 2005 za Ferry, Olejniczak 2009, w druku.

przedmiotu również oferuje bogate i zróżnicowane typologie (Alkin et al. 2006; Mark et al. 2000; Stufflebeam Shinkfield 2007). Upraszczając, na potrzeby niniejszej książki zidentyfikowaliśmy 5 podejść (zob. tabela 4).

Podejścia te można ulokować pomiędzy dwoma, zidentyfikowanymi już biegunami (procesy vs efekty). Przedstawiamy to graficznie na rysunku 4.

Warto zwrócić uwagę, że zarysowanej na rysunku 4 biegunowości odpowiadają dwie odmienne filozofie badawcze. Z jednej strony, jest to konstrukcjonizm, w którym rzeczywistość społeczna postrzega się w sposób subiektywny, przez zinstytucjonalizowane interakcje między ludźmi; interpretacja zjawiska zależy od przyjętej perspektywy. Z drugiej, jest to pozytywizm zakładający rzeczywistość społeczną jako obiektywną i zewnętrzną.



Rysunek 4. Dylematy wyboru podejść do ewaluacji

Źródło: Olejniczak 2008.

Ten dualizm wpisuje się w toczącą się w środowisku ewaluacyjnym polemikę dwóch paradygmatów – ilościowego i jakościowego[9]. Dyskusja ta toczyła się głównie w latach 80. w Stanach Zjednoczonych. Obecnie – w praktyce ewaluacji interwencji publicznych – widać wyraźnie przewagę pragmatycznego podejścia (*realistic approach*), które – w zależności od celu badania – odwołuje się do różnych elementów i łączy metody.

Na zakończenie tego podrozdziału raz jeszcze chcielibyśmy podkreślić omawianą kwestię wyborów. Unikanie łączenia sprzecznych funkcji w ramach jednego badania ewaluacyjnego jest jednym z kluczowych czynników warunkujących przydatność badań. Do kwestii tej powracać będziemy wielokrotnie w niniejszej książce, pokazując, na przykładach, jej praktyczne implikacje.

5. EWALUACJA JAKO PROCES UCZENIA SIĘ

Podstawowym założeniem, na którym opiera się ten podręcznik, jest stwierdzenie że **ewaluacja to proces uczenia się**. Założenie to jest powszechnie prezentowane zarówno we współczesnej literaturze przedmiotu, jak i dokumentach z zakresu polityk publicznych.

[9] Nie chodzi tu o kwestie użycia metod jakościowych (np. wywiady) ani ilościowych (np. analiza ankiet), ale o różnice filozoficzne – strategię badawczą. Strategia ilościowa zakorzeniona jest w pozytywizmie, preferuje podejście dedukcyjne, zaś obiektywizm jest jej ontologiczną podstawą. Paradygmat jakościowy preferuje podejście indukcyjne, interpretacyjne i akonstrukcjonizm jako orientację ontologiczną. Krótkie podsumowanie tych paradygmatów można znaleźć w pracy Górniaka (2007: 13–15), zaś wyczerpujące porównanie i streszczenie głównych nurtów dyskusji u Mathisona (2005: 345–350). Co ważne podkreślenia, obydwa paradygmaty – na poziomie metod badawczych – mogą sięgać po metody zarówno jakościowe (np. wywiady), jak i ilościowe, np. ankiety, analizy statystyczne (Bryman 2004: 27).

Osadzając to stwierdzenie w praktyce funduszy strukturalnych (czy – szerzej – logiki zarządzania opartego na wynikach), postanowiliśmy przedstawić je jako **proces produkcji i wykorzystywania wiedzy**.

Wpisaliśmy więc ewaluację w prezentowany już schemat logiczny wdrażania interwencji publicznych (por. rysunek 1). Przyjmujemy więc, że każde z badań ewaluacyjnych ma swój kontekst, nakłady, proces badawczy; przynosi także określone produkty (w postaci raportu i wytworzonej wiedzy), rezultaty (sposoby użycia tej wiedzy) i oddziaływanie (wypełnienie szerszych funkcji ewaluacji). Bezpośrednią inspiracją dla takiego ujęcia tematu były prace Cousins (2003), Marka, Henry'ego (2004), Owena (2007: 106), zaś szczegółowy model jest nieznaczną modyfikacją prac Olejniczaka (2008: 202–205) oraz Ferry'ego, Olejniczaka (2008, w druku). Model procesu i wykorzystania ewaluacji prezentujemy na rysunku 5.

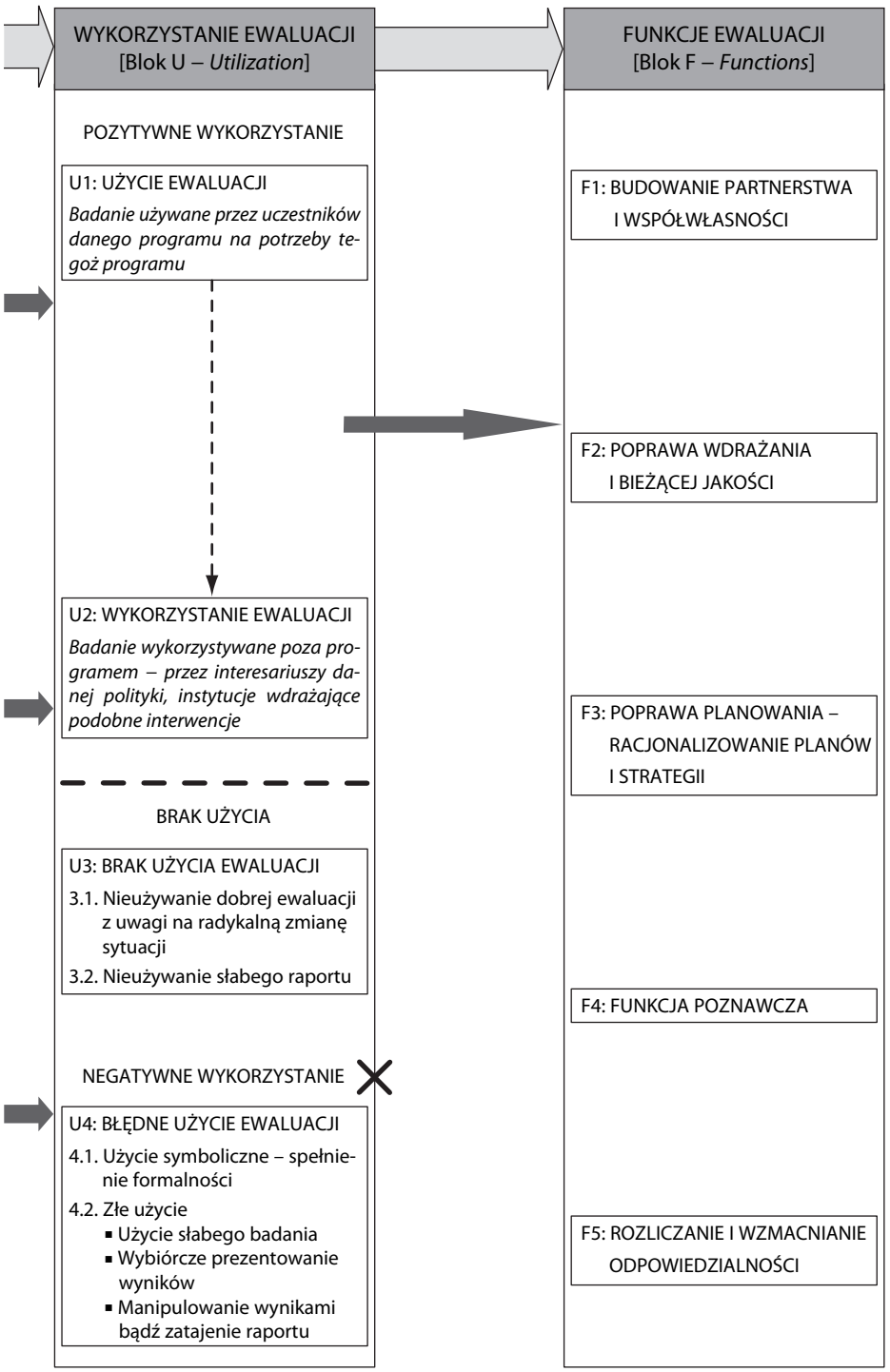
Trzy pierwsze bloki modelu (wkład, proces, produkty) zawierają czynniki warunkujące wykorzystanie ewaluacji (czynniki kontekstowe, charakterystykę uczących się i nauczających, przebieg przekazywania wiedzy oraz typy nabytej wiedzy). Pozostałe dwa bloki natomiast pokazują, jak ta wiedza jest spożytkowana (lub nie) i jakie szersze oddziaływanie przynosi (w naszym modelu przedstawione jako funkcje ewaluacji w sektorze publicznym).

Blok pierwszy (Wkład) przedstawia czynniki będące wkładem w proces ewaluacji. Do roli kultury administracyjnej kraju (potencjału ewaluacyjnego) – zarówno w kontekście państw UE, jak i kontekście polskim odnoszą się rozdziały 1.3 i 1.4. Rolę charakterystyki ewaluandy (przedmiotu badania), momentu podjęcia badania oraz wpływ profilu uczestników są analizowane w studiach przypadków. Do kwestii technicznych i znaczenia ich równowagi (segment I5 modelu) wprowadza rozdział 2.1, zaś szczegółowe omówienie tego problemu z perspektywy zlecającego i wykonawcy znalazło się w rozdziałach 2.6 i 2.7.

W bloku poświęconym procesowi ewaluacji wyróżniliśmy trzy kluczowe segmenty, które w zasadzie nie wymagają dodatkowego komentarza (poza segmentem podejść badawczych, który omówiliśmy już w tym rozdziale). Studia przypadków w dalszych częściach książki ilustrują natomiast funkcjonowanie tych elementów w praktyce badań funduszy strukturalnych.

Krótkiego wyjaśnienia wymagają natomiast segmenty przedstawione w bloku produktów ewaluacji. Pisząc o jakości raportu, odnosimy się do popularnych i powszechnie przyjętych standardów UE (European Commission, 1999: vol. 1, s. 180; 2008) Natomiast wyprodukowana w procesie ewaluacji wiedza ma według nas kilka wymiarów. Pierwszym jest jej nowość i odkrywczość wniosków (oceniana z perspektywy odbiorców raportu). Aspektem drugim jest jej poziom szczegółowości – skoncentrowanie się w raporcie na kwestiach technicznych (procedury,

Rysunek 5. Ewaluacja jako proces produkcji i wykorzystania wiedzy



Źródło: Ferry, Olejniczak 2009.

Tabela 5. Typy wiedzy

Typy wiedzy	Opis
Wiedza diagnostyczna – wyjaśnia, pokazuje szerszy kontekst	
Wiedzieć o (know about)	Wyjaśnia kontekst, w jakim działa program, wskazuje szanse i zagrożenia, także mechanizmy społeczno-gospodarcze, związane z wdrażanym programem
Wiedzieć co (know what)	Pokazuje, co działa, a co nie w programie
Wiedzieć dlaczego (know why)	Wyjaśnia, dlaczego rzeczy działają lub nie, może obejmować również szerszą refleksję nad teoriami leżącymi u podstaw danego programu
Wiedza preskryptywna – ocenia, daje rozwiązania, wskazuje konieczne działania	
Wiedzieć jak (know how)	Pokazuje, jak poprawić, zmienić, udoskonalić rzeczy
Wiedzieć kto (know who)	Informuje, kto powinien podjąć działania, z kim należy współpracować, kogo włączyć
Wiedzieć kiedy (know when)	Informuje, kiedy działania naprawcze powinny zostać podjęte

Źródło: Ferry, Olejniczak 2008, w druku.

cele, wskaźniki) albo też odnoszenie wniosków do szerszych kwestii (racjonalności polityki, skuteczności, a nawet zasadności przyjętego mechanizmu interwencji w świetle najnowszych teorii rozwoju).

Wreszcie trzeci aspekt to charakter diagnostyczny lub preskryptywny wyprodukowanej wiedzy (tabela 5). W tym miejscu wspieramy się typologią rozwiniętą przez Ferry'ego, Olejniczaka (2008, w druku).

W bloku czwartym (wykorzystanie ewaluacji) założyliśmy, że dobre wykorzystanie ewaluacji ma dwie perspektywy czasowe – krótko- i średnioterminową. Ewaluacja może być wykorzystywana także źle lub wcale. Przyjęliśmy więc typologię, którą przedstawia tabela 6.

Nasz model zamyka odniesienie do omawianych już funkcji ewaluacji. Przyjmujemy, że jeśli badanie ewaluacyjne jest dobrze wykorzystywane, to przyczynia się do realizacji jednej lub więcej funkcji w sektorze publicznym.

Prezentowany model jest rdzeniem niniejszej książki. Posłużył nam jako punkt odniesienia w kolejnych rozdziałach poświęconych zagadnieniom ogólnym i metodologii. Dla rozdziałów zawierających studia przypadków stał się natomiast ich dokładną „mapą drogową" wytyczającą kolejne elementy analizy danego badania ewaluacyjnego i jego efektów.

Tabela 6. Typy i podtypy wykorzystania ewaluacji

Typ	Podtypy i ich opis
Pozytywne wykorzystanie	**Pozytywne użycie** ewaluacji, czyli użycie danego badania przez instytucje zaangażowane w dany program i w odniesieniu do tego programu
	Pozytywne wykorzystanie wyników badania przez aktorów niezaangażowanych bezpośrednio w dany program, do działań poza programem
Brak użycia	**Brak użycia** dobrej ewaluacji, która straciła swoją aktualność ze względu na szybkie zmiany kontekstu programu i jego zawartości
	Brak użycia słabej ewaluacji
Błędne użycie	**Symboliczne użycie** – wykorzystywanie wniosków do pozornego uzasadnienia już podjętych decyzji
	Złe użycie – używanie słabego jakościowo badania, manipulacja wynikami – prezentowanie wyników wybiórczo, by uzasadnić jakąś tezę, ukrywanie negatywnych wniosków

Źródło: Alkin, Coyle 1988; Alkin, Taut 2003; Cousins 2004; Leviton 2003 za Ferry, Olejniczak 2009, w druku.

W ten sposób staraliśmy się zapewnić Czytelnikom w miarę jednolity i spójny obraz prezentowanych badań (przy zachowaniu specyfiki omawianych przypadków), a także ułatwić możliwość samodzielnych porównań i poszukiwania odniesień.

Mamy więc nadzieję, że proponowany model okaże się przydatną mapą drogową w podróży przez prezentowane studia przypadków, a także cennym punktem odniesienia do własnych, podjętych przez Czytelników, przemyśleń na temat ewaluacji.

BIBLIOGRAFIA

Alkin M.C., Christie C.A., Rose M. (2006), *Communicating evaluation*, w: Shaw I., Greene J.C., Mark M.M. (eds.), *Handbook of evaluation: policies, programs and practices*, London, Sage.

Alkin M.C., Coyle K. (1988), Thoughts on evaluation utilization, misutilization and non-utilization, *Studies in Educational Evaluation* **14(3)**, 331–340.

Alkin M.C., Taut S.M. (2003), Unbundling evaluation use, *Studies in Educational Evaluation* **29**, 1–12.

Bachtler J. (2001), Quod erat demonstrandum? Ewaluacja polityki regionalnej, *Studia Regionalne i Lokalne* **4(7)**.

Batterbury S. (2006), Principles and Purposes of European Union Cohesion Policy Evaluation, *Regional Studies* **40(2)**, 179–188.

Bingham R.D., Felbinger C.L. (2002), *Evaluation in practice: a methodological approach*, New York, London, Seven Bridges Press.

Bryman A. (2004), *Social research methods*, Oxford, Oxford University Press.

Chelimsky E. (1997), *The Coming Transformation in Evaluation*, w: Chelimsky E., Shadish W.R. (eds.), *Evaluation for the 21st century: a handbook*, Thousand Oaks, Calif., London, Sage.

Council Of The European Economic Community (1993), *Council Regulation (EEC) No 2082/93 of 20 July 1993 amending Regulation (EEC) No. 4253/88 laying down provisions for implementing Regulation (EEC) No 2052/88 as regards coordination of the activities of the different Structural Funds between themselves and with the operations of the European Investment Bank and the other existing financial instruments*, 2082/1993, OJ L 193, 31/07/1993.

Council Of The European Union (1999), *Council Regulation (EC) No. 1260/1999 of 21 June 1999 laying down General Provisions on the Structural Funds*, 1260/1999, OJ L161, 26.6.1999.

Cousins B.J. (2003), *Utilization effects of participatory evaluation*, w: Kellaghan T., Stufflebeam D.L., Wingate L.A. (eds.), *International handbook of educational evaluation*, Dordrecht; London, Kluwer Academic.

Cousins B.J. (2004), Commentary: Minimizing evaluation misuse as principled practice, *American Journal of Evaluation* **25(3)**, 391–397.

Dahler-Larsen P. (2005), *Evaluation In Public Management*, in: Ferlie E., Lynn L.E., Pollitt C. (eds.) *The Oxford handbook of public management*, Oxford, Oxford University Press.

Davidson E.J. (2005), *Evaluation methodology basics: the nuts and bolts of sound evaluation*, Thousand Oaks, Calif., Sage Publications.

European Commission (1999), *MEANS collection: evaluating socio-economic programmes*. DG-Regio, Office for Official Publications of the European Communities, Luxembourg.

European Commission (2000), *Focus on Results: Strengthening Evaluation of Commission Activities*, SEC2000/1051.

European Commission (2001), *Project Cycle Management Manual*, EUROPEAID.

European Commission (2004a), *Evaluating EU activities – A practical guide for the Commission services*, DG-BUDGET, Office for Official Publications of the European Communities, Luxembourg.

European Commission (2004b), *Project Cycle Management guidelines*, vol. 1, EUROPEAID, Office for Official Publications of the European Communities, Luxembourg, http://ec.europa.eu/europeaid/evaluation/methodology/index_en.htm

European Commission (2006), *Working Document no. 2 – Indicative Guidelines on Evaluation Methods: Monitoring and Evaluation Indicators*, DG-REGIO, The New Programming Period 2007–2013 Working Documents, Office for Official Publications of the European Communities, Luxembourg, http://ec.europa.eu/regional_policy/sources/docoffic/working/sf2000_en.htm

European Commission (2007), *Evalsed – the resource for the evaluation of socio-economic development*, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/index_en.htm [12.04.2008].

European Commission (2008), *Evaluation quality assessment form and guide on scoring the criteria*, DG-BUDGET, http://ec.europa.eu/budget/documents/evaluation_en.htm#standards

Ferry M., Olejniczak K. (2008, w druku), *The use of evaluation in management of EU programmes – evidences from Poland*, Ernst & Young, Program „Sprawne Państwo", 2nd revision, www.sprawnepanstwo.pl

Forunier D.M. (2005), *Evaluation*, w: Mathison S. (ed.), *Encyclopedia of evaluation*, Thousand Oaks, Calif., London, Sage.

Frechtling J.A. (2007), *Logic modeling methods in program evaluation*, San Francisco, Calif., Jossey-Bass [Chichester: John Wiley, distributor].

Górniak J. (2007), Ewaluacja w cyklu polityk publicznych, w: Mazur S. (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Kraków, Uniwersytet Ekonomiczny w Krakowie.

Greene J.C. (2003), Understanding Social Programs through Evaluation, w: Denzin N.K., Lincoln Y.S. (eds.), *Collecting and interpreting qualitative materials*, Thousand Oaks, Calif., London, Sage.

Greene J.C., Caracelli V.J., Graham W.F. (1989), Toward a Conceptual Framework for Mixed-Method Evaluation Designs, *Educational Evaluation and Policy Analysis* **11(3)**, 255–274.

Guba E.G., Lincoln Y.S. (1989), *Fourth generation evaluation*, Newbury Park, Sage.

Herman J.L., Morris L.L., Fitz-Gibbon C.T. (1987), *Evaluator's handbook*, Newbury Park, Sage & University of California Los Angeles – Center for the Study of Evaluation.

Keehley P., Abercrombie N. (2008), *Benchmarking in the public and nonprofit sectors: best practices for achieving performance breakthroughs*, San Francisco, Calif., Jossey-Bass.

Kierzkowski T. (2002), *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, Warszawa, Polska Agencja Rozwoju Przedsiębiorczości.

Korporowicz L. (red.) (1997), *Ewaluacja w edukacji*, Warszawa, Oficyna Naukowa.

Langbein L., Felbinger C.L. (2006), *Public Program Evaluation. A statistical guide*, London, New York, M.E.Shape.

Leviton L.C. (2003), Evaluation use: advances, challenges and applications, *American Journal of Evaluation* **24(4)**, 525–535.

Mark M.M., Henry G.T. (2004), The mechanisms and outcomes of evaluation influence, *Evaluation* **10(1)**, 35–57.

Mark M.M., Henry G.T., Julnes G. (2000), *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs*, San Francisco, Calif., Jossey-Bass.

Mathison S. (ed.) (2005), *Encyclopedia of evaluation*, Thousand Oaks, Calif.; London, Sage.

OECD (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, OECD-Development Assistance Committee, Paris.

Olejniczak K. (2007), *Ewaluacja jako narzędzie zarządzania w sektorze publicznym*, w: Gorzelak G., Tucholska A. (red.), *Rozwój, region, przestrzeń*, ss. 361–396, Warszawa, Ministerstwo Rozwoju Regionalnego.

Olejniczak K. (2008), *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INTERREG III*, Warszawa, Wydawnictwo Naukowe Scholar.

Owen J.M. (2007), *Program evaluation: forms and approaches*, 3rd edition, London, The Guilford Press.

Owen J.M., Rogers P.J. (1999), *Program evaluation: forms and approaches*, London, Sage.

Oxford University Press (2002), Oxford English Dictionary, Oxford University Press.

Paddock S.C. (1998), *Evaluation*, w: Shafritz J.M. (ed.), *International Encyclopedia of Public Policy and Administration*, ss. 818–823, Boulder, Colo., Westview Press.

Patton M.Q. (2004), *Evaluation Research*, w: Lewis-Beck M.S., Bryman A., Liao T.F. (ed.) *The Sage encyclopedia of social science research methods*, ss. 337–340, Thousand Oaks, Calif., Sage.

Pedersen L., Rieper O. (2008), Is Realist Evaluation a realistic approach for complex reforms?, *Evaluation* **14(3)**, 271–298.

Prime Minister's Strategy Unit (2003), *The Magenta Book Guidance Notes for Policy Evaluation and Analysis*, Government Chief Social Researcher's Office, UK Prime Minister's Cabinet Office, London, http://www.nationalschool.gov.uk/policyhub/magenta_book/

Rossi P.H., Lipsey M.W., Freeman H.E. (2004), *Evaluation: a systematic approach*, Thousand Oaks, Calif., London, Sage.

Rozporządzenie Rady (WE) (2006), *Rozporządzenie Rady (WE) nr 1083/2006 z dnia 11 lipca 2006 r. ustanawiające przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego oraz Funduszu Spójności i uchylające rozporządzenie (WE) nr 1260/1999*, 1083/2006, OJ L210, 31.7.2006.

Scriven M. (1991), *Evaluation thesaurus*, Newbury Park; London, Sage.

Stern E. (2005a), *Evaluation research methods*, London, Sage.

Stern E. (2005b), *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, 16.09.2005.

Stufflebeam D.L. (2001), Evaluation Models, *New Directions for Evaluation* 89(Spring).

Stufflebeam D.L., Shinkfield A.J. (2007), *Evaluation theory, models, and applications*, San Francisco, Calif., Jossey-Bass; Chichester, John Wiley [distributor].

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

Weiss C.H. (1998), *Evaluation: methods for studying programs and policies*, Upper Saddle River, N.J. Prentice Hall.

World Bank (2008), Independent Evaluation Group – Overview, http://www.worldbank.org/oed/ieg_overview.html

Worthen B.R. (1990), *Program evaluation*, w: Walberg H.J., Haertel G.D. (eds.), *The International encyclopedia of educational evaluation*, Oxford, England; New York, Pergamon Press.

ROZDZIAŁ 1.2.

Marek Kozak

POLITYKA SPÓJNOŚCI JAKO PRZEDMIOT BADAŃ EWALUACYJNYCH

Polityka spójności zajmuje szczególne miejsce w systemie europejskim. Jak zauważa Willem Molle, dla jej zwolenników, polityka spójności jest cementem utrzymującym całą konstrukcję zróżnicowanej wewnętrznie Unii. Spotyka się także głosy odmienne, iż polityka spójności nie przynosi obiecywanych korzyści, jest nieskuteczna i nieefektywna (Molle 2007, Kozak 2007a). Nie rozstrzygając tej kwestii, zauważmy, że mowa jest o kluczowym problemie oceny polityki spójności, jej użyteczności dla procesów rozwojowych, integracji europejskiej czy pozycji Unii na scenie światowej.

Jeśli ewaluacja polityki spójności ma być długofalowo użyteczna, musi uwzględniać szersze jej uwarunkowania. Rozdział ten ma ukazać związki, jakie łączą politykę spójności ze zjawiskami zachodzącymi w jej otoczeniu. Podstawowa teza głosi, że polityka spójności jest kształtowana przez wiele czynników wobec niej zewnętrznych, a bez ich znajomości każda szersza ocena może prowadzić do nietrafnych wniosków, a w efekcie do suboptymalnej alokacji środków.

Związki te będą omawiane w odniesieniu do dwóch grup czynników:

- poziomu rozwoju i zróżnicowań w UE,
- czynników instytucjonalnych (w tym stosowany paradygmat rozwoju, proces integracji europejskiej, system zarządzania działania innych polityk unijnych).

Mówiąc językiem ewaluacji, przyjrzymy się, czy istnieje odpowiedniość między stanem rzeczy a odpowiedzią instytucjonalną.

1. ZRÓŻNICOWANIA TERYTORIALNE UE

Fundamentem interwencji w zakresie polityki spójności winien być poziom rozwoju i stan zróżnicowań terytorialnych UE, a miarą jej skuteczności zanikanie różnic (a miarą nieskuteczności ich wzrost, czyli konwergencja lub dywergencja). Zróżnicowania te możemy mierzyć na różnych poziomach: krajowym bądź regionalnym, przy czym, jak dowodzi praktyka, na każdym z nich kierunek zmian może być odmienny. Pomiaru dokonuje się rutynowo, z uwzględnieniem trzech aspektów: ekonomicznego, społecznego i – ostatnio – terytorialnego. Główny miernik różnic gospodarczych to PKB, czasem także uwzględniający czynniki jakościowe HDI, zmienność tego ostatniego jednak i tak w zdecydowanej większości jest wyjaśniana przez poziom dochodu. W kwestiach spójności społecznej pomiaru dokonuje się, badając stopę bezrobocia, a ostatnio coraz częściej stopę zatrudnienia. Największy dotąd problem wiąże się z pomiarem spójności terytorialnej: stosowane są różne wskaźniki (dostępność komunikacyjna, dostępność usług publicznych, badanie relacji między obszarami funkcjonalnymi, np. centrum – peryferia, miasta – wsie, ostatnio coraz częściej obszary metropolitalne – reszta). Ponieważ istnieje obszerna i aktualna literatura dotycząca stanu zróżnicowań w Unii (w tym czwarty *Raport o spójności gospodarczej i społecznej*), nie ma potrzeby szczegółowego omówienia różnic. Przypomnijmy tylko główne ich wymiary wyznaczane przez następujące osie:

- rdzeń gospodarczy Europy – obszary peryferyjne,
- obszary metropolitalne – obszary inne.

Istotną cechą współczesnych różnic jest to, że ulegają one szybkiemu zmniejszaniu w zakresie wyposażenia w tradycyjną infrastrukturę techniczną (zwłaszcza lokalną), w niektórych porównaniach (na poziomie krajowym) także pod względem dochodu. Utrzymuje się natomiast wysokie zróżnicowanie w zakresie innowacyjności i rozwoju społeczeństwa informacyjnego, co uważa się dziś – jak widać w praktyce nie wszędzie – za motor rozwoju. Warto zwrócić uwagę, że pozycja poszczególnych państw w rankingu może zaskakująco szybko się zmieniać. Badanie okresu 1990–2004 pokazuje, że PKB Niemiec w odniesieniu do średniej UE zmniejszył się ze 116 do 109%, Holandii wzrósł ze 100 do 125, Wielkiej Brytanii z 99 do 119, Irlandii z 74 do 141, a Francji ze 109 do 111% (Molle 2007: 28).

Polska należy do słabiej rozwiniętych krajów Unii (z PKB równym 52% średniej). Odznacza się wysokim tempem wzrostu PKB (w okresie 2001–2007 w cenach stałych o 131,5%, podczas gdy w UE27 o 115,9, a w OECD bez krajów grupy wyszehradzkiej o 113,5% – GUS 2008: 13). W ostatnich latach, paradoksalnie, mimo wysokiej dynamiki wzrostu, Polska cofa się w rankingu PKB Unii, mając za sobą już tylko Bułgarię

i Rumunię. Jest za wcześnie, żeby orzekać, czy jest to trend trwały. Jeśli tak, to polska polityka rozwoju oraz system instytucjonalny wymagają głębokich reform[1].

2. PARADYGMAT ROZWOJU A POLITYKA SPÓJNOŚCI

W ciągu ostatnich 40 lat świat przechodzi od gospodarki industrialnej do postprzemysłowej, gdzie głównym źródłem dochodu są usługi, gospodarka jest oparta na wiedzy, a kulturową dominantą jest postmodernizm. Realizuje się „trzecia fala" w dziejach ludzkości (Toffler 1997). To rewolucja, której podlegają także reprezentujące różny poziom rozwoju kraje UE, mogąca prowadzić do daleko idących zmian na scenie globalnej. Jakie czynniki przesądzają dziś o rozwoju? W co inwestować, by poprawić pozycję konkurencyjną kraju czy regionu?

Teorie rozwoju różnią się zwłaszcza w dwóch kwestiach:

- roli państwa i charakteru stosowanych instrumentów w rozwoju,
- czynników rozwoju.

W pierwszej sprawie, przyjął się ogólny podział na koncepcje neokeynesowskie, bazujące na podejściu popytowym, oraz neoklasyczne, podażowe. W ogólnym zarysie doktryna Keynesa (i pochodne) zakładała konieczność aktywizacji polityki gospodarczej państwa dla redukcji skutków niedoskonałości rynku. Wzrost roli państwa jako aktora na rynku oraz regulatora rynku miał zapobiegać cyklom gospodarczym i zwłaszcza przeciwdziałać kryzysom i ich skutkom. Zamówienia państwowe (naturalnie możliwe tylko dzięki podniesieniu podatków), przyjmujące także postać inwestycji infrastrukturalnych i biznesowych, miały sprzyjać utrzymaniu produkcji i miejsc pracy, a nawet rozwojowi. Po okresie dominacji podejścia keynesowskiego (od II wojny światowej do końca lat 60.), w świetle coraz liczniejszych dowodów nieskuteczności odgórnej polityki industrializacji, do łask wróciły doktryny neoklasyczne, ograniczające znaczenie państwa w gospodarce do niezbędnego minimum, pozostawiające mu funkcję regulatora rynku i arbitra, ale nie aktywnego przedsiębiorcy.

Odrębną kwestią jest przyjęcie założeń dotyczących czynników rozwoju we współczesnej gospodarce, charakteryzującej się cechami wyraźnie odmiennymi od ery przemysłowej. Jest to gospodarka o globalnym charakterze, co przejawia się zarówno w decyzjach o lokalizacji firm, poszerzaniu konkurencji na obszary technologii informacyjnych i usług,

[1] Na tle innych państw, oceny polskiego systemu instytucjonalnego (Bank Światowy, za: Strategia rozwoju kraju 2006), warunków prowadzenia przedsiębiorstw (Bank Światowy, za Baj 2008), innowacyjności (European Innovation Scoreboard, ProInno 2007) czy przygotowania do społeczeństwa informacyjnego (ESPON 1.2.3, 2007), wypadają bardzo źle i Polska zajmuje dalekie pozycje w rankingach.

mobilności technologii, jak i globalizowaniu rynku pracy. Wyróżnia ją znaczne przyspieszenie procesów: cykl życia produktów uległ skróceniu, a proces zmian technologicznych przyspieszeniu (stąd i potrzeba ciągłego podnoszenia kwalifikacji). Współczesna gospodarka to gospodarka oparta na wiedzy, co oznacza nie tylko wzrost zatrudnienia w sektorach związanych bezpośrednio z przetwarzaniem informacji, ale też wzrost wymagań wobec kwalifikacji. Obserwowana specjalizacja firm wiedzie do wzrostu ich powiązań z innymi firmami, instytucjami, dostawcami, a zarazem podnosi elastyczność i zdolność do rozwoju. (Blakely, Bradshaw 2002: 5). Im bardziej dana gospodarka jest zdolna do zaoferowania podmiotom gospodarczym cech wynikających z powyższych warunków, tym większe są jej szanse na odniesienie sukcesu.

Spotyka się też podejście eksponujące znaczenie globalizacji, konkurencji i innowacji, jako trzech ściśle powiązanych czynników wchodzących w skład współczesnego modelu rozwoju. Zgodnie z tym podejściem rozwojem cieszyć się mogą współcześnie tylko te obszary, które są zdolne do uzyskania przewagi konkurencyjnej dzięki swej zdolności do innowacji technologicznej, organizacyjnej czy produktowej (Gorzelak 2007). Inne podejście nadaje szczególną wagę czynnikom instytucjonalnym, podkreślając, że współcześnie system instytucjonalny wyznacza szanse rozwoju przez determinowanie możliwości efektywnego wykorzystania dostępnych zasobów[2] (Amin, Thrift 2000; Keating, Koughlin, Deschouwer 2003; Hryniewicz 2004; Kukliński 2007). Wielu zwolenników ma koncepcja podnosząca znaczenie kultury (Landes 2005; Harrison, Huntington 2003). W gruncie rzeczy oba te podejścia – kulturowe i instytucjonalne – nie pozostają w sprzeczności, bowiem, jak nietrudno zauważyć, kultura wyznacza ramy instytucji, te zaś oddziałują na kulturę.

Praktyczną wykładnią zmian globalnych są decyzje lokalizacyjne przedsiębiorców, mające ogromne znaczenie dla każdej polityki rozwoju, która może być o tyle tylko skuteczna, o ile zdoła uwzględnić preferencje biznesu. Jeszcze przed nieco ponad ćwierćwieczem ważnymi czynnikami były: dostępność energii, koszt gruntów i budynków, tania siła robocza, dostęp do podstawowej infrastruktury technicznej (linii kolejowych, dróg). Dziś na czoło wysuwają się: dostęp do strategicznej infrastruktury transportowej (autostrady, lotniska, porty), wykwalifikowana, elastyczna siła robocza, oraz, w rosnącym stopniu, tzw. miękkie czynniki lokalizacyjne – klimat dla biznesu, dostępność zaplecza naukowo-badawczego, szkolnictwo, instytucje kultury, atrakcyjność miejsca (Dziemianowicz, Jałowiecki 2004). Wszystko to sprzyja współcześnie inwestycjom na obszarach metropolitalnych, te bowiem odznaczają się największym nagromadzeniem pożądanych cech. Zjawisko to wyznaczyło

[2] Przez zasoby rozumiem występujące obiektywnie atrybuty środowiska naturalnego i społecznego.

nową oś zróżnicowań terytorialnych (Jałowiecki 2005). Dobrze widocznym zjawiskiem jest również relatywne (czasem też bezwzględne) ubożenie peryferyjnych obszarów wiejskich, podlegających dezindustrializacji (Blakely, Bradshaw 2002) i depopulacji. Przykładem tego ostatniego zjawiska są zarówno regiony Polski wschodniej, jak i byłego NRD, dotkniętego wręcz katastrofą demograficzną na obszarach wiejskich. Do nowych zjawisk w sferze społecznej zalicza się segmentację rynku pracy, przejawiającą się w podziale na zajęcia pierwszorzędne i marginalne, niewymagające kwalifikacji i niskopłatne. Prowadzi to np. w USA do obserwowanego dziś chronicznego ubóstwa, dotykającego nie tylko osób trwale bezrobotnych, ale też podejmujących wspomniane już zajęcia marginalne, uzależniające od pomocy socjalnej (Hirszowicz, Neyman 2004; Jarosz 2005).

W ciągu swego stosunkowo niedługiego istnienia w formie rozwiniętej, europejska polityka spójności zaczęła się jako polityka o mieszanych podstawach koncepcyjnych, z naciskiem na elementy doktryny neokeynesowskiej, by w miarę upływu czasu w coraz większym stopniu uwzględniać podejście neoliberalne.

Polityka spójności od początku waha się między wyrównywaniem poziomu rozwoju a maksymalizacją efektywności interwencji. Jest to kwestia praktycznego wyboru między przyspieszeniem rozwoju kraju jako całości (dzięki koncentracji nakładów w miejscach najbardziej konkurencyjnych) kosztem pogłębienia się zróżnicowań międzyregionalnych, a zdecydowanym nastawieniem na wyrównywanie poziomu rozwoju regionów kosztem wolniejszego rozwoju kraju.

Przyjęcie określonego paradygmatu ma oczywisty wpływ na cele i strukturę interwencji polityki spójności. W ciągu ostatnich lat obserwujemy stały wzrost nacisku na nowy paradygmat, reprezentowany w polityce spójności przez Strategię Lizbońską. Podobnie jak Strategia Lizbońska, polityka spójności w tej mierze jest modyfikowana przez presje społeczne, płynące z odmiennych wyobrażeń na temat czynników i celów rozwoju. Dlatego, tak jak pierwotna wersja Strategii Lizbońskiej została poważnie pozbawiona swej tożsamości przez dodanie nowego celu, jakim stało się idealistyczne „pełne zatrudnienie", tak strategia spójności, nakazała od 2007 roku w celu konwergencji na cele tzw. lizbońskie przeznaczyć nie mniej niż 75% nakładów budżetowych w państwach UE15 i 60% w nowych państwach członkowskich, ale te ostatnie natychmiast uwolniła od tego zobowiązania. Było to ewidentnym zgniłym kompromisem między potrzebą i chęcią wdrożenia nowego paradygmatu w skali Unii a oporem w krajach i regionach słabiej rozwiniętych, które w większości nadal hołdują teorii zależności (*dependency theory*) i tkwią w przeświadczeniu, że rozwój wymaga powtórzenia ścieżki, jaką przeszły najbogatsze kraje. Główny czynnik, przesądzający o przewadze teorii zależności w mniej rozwiniętej Europie (także Grecji czy Portugalii), to głęboko zakorzenione przekonanie, że

podstawowy warunek rozwoju i czynnik lokalizacji to infrastruktura techniczna. Nic bardziej mylnego – byłe NRD ma zapewne najnowocześniejszą infrastrukturę i mimo gigantycznych transferów z budżetu centralnego przechodzi stałe problemy rozwojowe. Ośrodki miejskie Polski wschodniej dysponują nową infrastrukturą, ale wbrew oczekiwaniom inwazja inwestorów nie nastąpiła. Włochy południowe, mimo wielkich inwestycji infrastrukturalnych, nadal należą do najbiedniejszych obszarów w Europie.

Warto zauważyć, że polityka rozwojowa zarówno UE, jak i krajów członkowskich odwołuje się deklaratywnie do nowego paradygmatu rozwoju. Im bliżej jednak konkretnych działań i projektów, tym silniej zaznacza się kompromis z minionym paradygmatem. Niestety, najlepiej jest on widoczny w słabiej rozwiniętych krajach Unii. W efekcie, mimo że Polska za cel strategiczny stawia budowę gospodarki opartej na wiedzy i społeczeństwa informacyjnego, to w latach 2007–2013 ze środków polityki spójności nadal planuje się wydać ponad 60% na infrastrukturę, a tylko do 15% na rozwój zasobów ludzkich i do 25% na przedsiębiorczość (Kozak 2007b; 2008).

3. POLITYKA SPÓJNOŚCI NA TLE INTEGRACJI EUROPEJSKIEJ

Państwa założycielskie Wspólnot borykały się z istotnymi różnicami regionalnymi (w przypadku Włoch można mówić o różnicach dramatycznych), pomimo dawno osiągniętego sukcesu w przezwyciężeniu wojennego i powojennego niedoboru żywności. Polityka rolna stała się więc głównym konsumentem środków wspólnego budżetu. Ani polityka regionalna, ani spójności, nie znalazły precyzyjnego zapisu w Traktatach Rzymskich, nie mówiąc już o środkach pieniężnych. W preambule traktatu o EWG wymienia się „wzmocnienie jedności swoich systemów gospodarczych i zapewnienie ich harmonijnego rozwoju przez zmniejszanie istniejących różnic między poszczególnymi regionami oraz usuwanie opóźnienia mniej uprzywilejowanych regionów", potem jednak w tekście nie znajdziemy doprecyzowania metod i instrumentów osiągania tego celu. Kluczowy dla Traktatu art. 2 stanowi ogólnie: „Zadaniem Wspólnoty jest popieranie – przez utworzenie wspólnego rynku oraz stopniowe ujednolicanie polityki gospodarczej Państw Członkowskich – harmonijnego rozwoju działalności gospodarczej na terenie całej Wspólnoty, stałego i równomiernego rozwoju, umacniania stabilności, szybszego podnoszenia poziomu życia i ściślejszych stosunków między Państwami należącymi do Wspólnoty". Kładziono wyraźny nacisk na instrumenty horyzontalne, mające, przez zniesienie barier i wprowadzenie wspólnego rynku, doprowadzić do harmonijnego rozwoju Wspólnot. Nic dziwnego, że przy takim podejściu, wobec braku solidnej podstawy traktatowej, polityka spójności, obejmująca instrumenty terytorialne

w początkowym okresie istnienia Wspólnot, właściwie nie istniała, a koordynacja polityk rozwojowych w poszczególnych państwach nie przynosiła efektu. Okres ten nazywano często sytuacją orkiestry bez dyrygenta. Podejście takie brało się ze wspomnianego przekonania, że proces tworzenia wspólnego rynku doprowadzi do dyfuzji korzyści rozwojowych z obszarów najwyżej we Wspólnotach rozwiniętych do obszarów najsłabszych, co jednak nie nastąpiło Jednocześnie rozwój Wspólnot o nowe państwa (1973: Wielka Brytania, Dania, Irlandia) wnosił nowe elementy do już istniejących zróżnicowań i – szczególnie w przypadku Wielkiej Brytanii – także nowe oczekiwania, w tym zwłaszcza w zakresie utworzenia możliwości rekompensowania Brytyjczykom nieproporcjonalnie wysokich wpłat do wspólnego budżetu[3]. Wskutek tego w 1975 roku doszło do powołania Europejskiego Funduszu Rozwoju Regionalnego, wyposażonego jednak w stosunkowo skromne środki. W pierwszej dekadzie swego istnienia dysponował w sumie budżetem 11,7 mld ECU, przeznaczonym głównie (a jakże!) na infrastrukturę (Balchin et al. 1999).

W historii Wspólnot (potem Unii) kamieniem milowym był przełom dokonany w połowie lat 80. Po okresie trwającego od lat 50. europejskiego cudu gospodarczego, od roku 1973, w następstwie wstrząsu wywołanego kryzysem naftowym i koniecznością dostosowań mikro- i makroekonomicznych, gospodarki europejskie zaczęły spowalniać tempo wzrostu, wręcz stagnować. Ponownie szybko zaczął rosnąć dystans dzielący Wspólnoty od USA. Na tle ekonomicznym dochodziło do spirali roszczeń socjalnych. Według głośnego Raportu Cecchiniego (*Cost of non-Europe*, 1988), główna przyczyna kryzysu gospodarczego Wspólnot brała się z zahamowania, wręcz zahibernowania procesów integracji europejskiej w połowie lat 60. Straty ponoszone wskutek powyższego szacowano na 4,25 do nawet 6,25% PKB (European Parliament, 2008). Raport dowodził, że dalsze pogłębienie integracji przyniesie wymierne korzyści ekonomiczne, pozwoli skorzystać z efektu skali, obniży koszt tworzenia PKB, przywróci konkurencyjność europejskiej gospodarce. Receptą na sukces miało być stworzenie Unii Gospodarczej i Walutowej. Wejście (właściwie wchłonięcie) Grecji (1981), a następnie akcesja Hiszpanii i Portugalii (1986) dały dodatkowy argument na rzecz potrzeby reform. Zapisana w Jednolitym Akcie Europejskim (JAE, 1986) daleko idąca reforma instytucjonalna została zaplanowana na przeszło dekadę.

---

[3] Wielka Brytania, jako najbardziej umiędzynarodowiona gospodarka w Europie, wnosiła do budżetu Wspólnot wysokie wpłaty związane z rozwiniętym handlem międzynarodowym. Nie była jednak zainteresowana korzystaniem ze środków wspólnej polityki rolnej (brytyjskie rolnictwo od lat było konkurencyjne i traktowane na równych prawach z innymi sektorami gospodarki). Polityka regionalna dawała możliwość zbilansowania wkładów i korzyści Wielkiej Brytanii, przy okazji będąc dobrą sposobnością do zredukowania nakładów na nieosiągającą założonych celów Wspólną Politykę Rolną.

Opierała się na założeniu, że wdrożenie Unii Gospodarczej i Walutowej (zniesienie barier w obrocie wewnętrznym, wprowadzenie Europejskiego Banku Centralnego i wspólnej waluty), mogą prowadzić do osłabienia ochrony rynków narodowych przed konkurencją. W przypadku regionów słabiej rozwiniętych groziłoby to szokiem gospodarczym i społecznym (Molle 2007).

W JAE po raz pierwszy wprowadzono pojęcie spójności (Tytuł V *Spójność Gospodarcza i Społeczna*). Artykuł 130a stanowił: „W celu wspierania harmonijnego rozwoju całej Wspólnoty, rozwija ona i prowadzi działania służące wzmocnieniu jej spójności gospodarczej i społecznej. W szczególności Wspólnota zmierza do zmniejszenia dysproporcji w poziomach rozwoju różnych regionów oraz zacofania regionów najmniej uprzywilejowanych".

W wymiarze spójnościowym reforma przyjęła postać tzw. Pakietów Delorsa (od nazwiska ówczesnego przewodniczącego KE i promotora integracji europejskiej). W wyniku pierwszego pakietu, na lata 1989–1993 zdefiniowano cele polityki strukturalnej, wprowadzono kategoryzację jednostek terytorialnych NUTS i rozpoczęto systematyczny pomiar spójności na potrzeby prowadzonej polityki, podwojono nakłady (z 7,2 mld ECU w 1987 do 14,5 mld w 1993 roku) i sformułowano zasady polityki. W *Traktacie o Unii Europejskiej*, przyjętym w 1992 roku, w art. B jako pierwszy cel Unii wymienia się „popieranie zrównoważonego i trwałego postępu gospodarczo-społecznego, w szczególności przez stworzenie obszaru pozbawionego wewnętrznych granic, wzmacnianie spójności ekonomicznej i społecznej i ustanowienie unii gospodarczo-walutowej, docelowo ze wspólną walutą, zgodnie z postanowieniami niniejszego Traktatu". W Tytule XIV *Spójność Gospodarcza i Społeczna* w art. 130a–e powtarza się cel i określa z pewnymi zmianami w porównaniu z JAE postanowienia nt. polityki spójności. Istotną zmianą jest decyzja o powołaniu w 1993 roku Funduszu Spójności. Ewolucję budżetu polityki spójności pokazuje tabela 1.

Tabela 1. Budżet UE i polityki strukturalnej (w mld EUR, dane w cenach bieżących)

	1989–93	1994–99	2000–06	2007–13
Budżet UE	244,8	529,9	646,2	862,3
Fundusze strukturalne	53,1	162,2	195,0[4]	249,7
Fundusz Spójności	–	15,15	18,0[4]	58,3

Źródło: opracowanie na podstawie danych KE.

[4] Bez środków dla 10 nowych państw członkowskich (łącznie 39 mld euro, w tym około 13 mld na Fundusz Spójności).

W odniesieniu do systemu zarządzania, fazę tę charakteryzowała przede wszystkim niska jakość przepisów. Problemem okazał się złożony, wieloetapowy system programowania, taki sam dla każdego państwa i programu[5]. Słabością o szczególnym znaczeniu okazał się też brak rozwiniętego systemu wskaźników realizacji celów, monitoringu i ewaluacji. W efekcie dochodziło do częstego przesunięcia celów (cele realizowane odbiegały od założonych), a ostateczne precyzyjne ustalenie, czy i w jakim stopniu zostały osiągnięte, było właściwie niemożliwe.

Drugi pakiet Delorsa (na lata 1994–1999) bazował na analizie postępów procesu budowy UGW i jej skutków, efektów dotąd prowadzonej polityki spójności oraz zmian zachodzących w gospodarce europejskiej i globalnej (Pietrzyk 2001; Balchin et al. 1999). Przejawiło się to m.in. w przeformułowaniu dwóch celów horyzontalnych i redukcji inicjatyw wspólnotowych z 14 do 13. Istotniejsze zmiany główne dotyczyły jednak powołania nowych instytucji, przede wszystkim wspomnianego Funduszu Spójności oraz Finansowego Instrumentu Rozwoju Rybołówstwa, który jako przejściowe narzędzie miał pomóc łagodzić skutki restrukturyzacji obszarów zależnych od przestarzałego, tradycyjnego połowu ryb. Z okazji akcesji Austrii, Finlandii i Szwecji w 1995 roku wprowadzono nowy cel polityki strukturalnej (zwany potocznie szóstym, choć w rzeczywistości był to cel 7. w kolejności), udzielający wsparcia regionom NUTS I o gęstości zaludnienia poniżej 8 osób/km^2. Dysponując kwotą ok. 700 mln EUR, stanowił raczej symboliczną rekompensatę wkładów wnoszonych do wspólnego budżetu przez nowe państwa członkowskie. W związku z ujawnionymi poważnymi rozbieżnościami w ocenie efektów interwencji europejskiej, czego przyczyną w znacznym stopniu był brak mierników postępu oraz kryteriów oceny, podjęto decyzję o wypracowaniu i wdrożeniu metodologii ewaluacji. Istotną zmianę w systemie zarządzania wprowadzono w 1994 roku, gdy obok tzw. klasycznej ścieżki programowania dopuszczono w pewnych wypadkach możliwość stosowania ścieżki uproszczonej[6].

Rok 1999 to z jednej strony początek ostatniego etapu UGW, z drugiej zaś rok intensywnych przygotowań do zbliżającej się największej akcesji w dziejach Wspólnot, obejmującej 10–12 państw Europy Środkowo-Wschodniej. Ponownie zreformowano politykę strukturalną, zmieniając liczbę celów (3), redukując liczbę i budżet inicjatyw wspólnotowych,

---

[5] Dokumenty wymagane to krajowy plan rozwoju, podstawy wsparcia wspólnoty, program operacyjny, uzupełnienie programu operacyjnego. Uciążliwość wiązała się z potrzebą przedstawiania Komisji i negocjacji kolejno opracowywanych dokumentów (formalnie z wyłączeniem podstaw wsparcia, ale w rzeczywistości znaczne partie tego dokumentu były także przygotowywane w krajach członkowskich).

[6] Wymagane do opracowania dokumenty to jednolity dokument programowy i uzupełnienie programu. Jeśli chodzi o zawartość treściową, JDP zawierał ten sam zakres informacji, co plan rozwoju, podstawy wsparcia wspólnoty i program operacyjny.

nadając status funduszu strukturalnego FIRR. Okres ten miał dwie cechy charakterystyczne. Pierwsza to nieproporcjonalnie mały – w obliczu bezprecedensowej akcesji przewidywanej pierwotnie na 2002 rok – wzrost planowanych nakładów na politykę strukturalną (tabela 1). Drugi – nie bez związku z nieprawidłowościami wykrytymi w latach 90. – to wprowadzenie zaostrzonych rygorów dostępu do środków, wzmocnienie roli ewaluacji w zarządzaniu (European Commission 1999) oraz znaczące uszczegółowienie procesu programowania, co było reakcją na niedoregulowanie okresów wcześniejszych. Przejawiło się to zwłaszcza w konieczności sporządzania tzw. uzupełnień do każdego programu operacyjnego. Dokumenty te (formalnie tylko do wiadomości Komisji, w praktyce zatwierdzane), zamiast przyczynić się do lepszego poinformowania beneficjentów i wykonawców o trybie realizacji programów, a tym samym do lepszych wyników, przez swą szczegółowość powodowały niemałe trudności, bowiem jakakolwiek zmiana zapisów (czytaj: usprawnienie) wymagała skomplikowanych i czasochłonnych procedur, przy okazji eliminując i tak skromne elementy zarządzania menedżerskiego[7].

Mająca miejsce w tym okresie akcesja 2004 roku okazała się bezbolesna, a korzyści płynące z niej dla całej Unii szybko się ujawniły. Potwierdziła też znaczenie Funduszu Spójności (czyli infrastruktury strategicznej) w rozwoju obszarów słabo rozwiniętych i peryferyjnych, a takie w większości weszły w skład Unii. Poszerzenie Unii o 12 nowych państw prowadziło do gwałtownego wzrostu zróżnicowań regionalnych (z poziomu 1: 2,3 do 1: 5,6) i spadku średniego dochodu w Unii o blisko 15%. Mimo to, w 2006 roku, pod naciskiem 6 państw – płatników netto do budżetu, nastąpiło zmniejszenie wielkości budżetu UE na okres 2007–2013 z poziomu 1,24% do 1,045% PNB Unii. Jak widać, budżet UE i nakłady na politykę spójności nie zależą od poziomu zróżnicowań. Jednocześnie zmieniono strukturę celów, wprowadzając cel *Konkurencyjność* i podporządkowując politykę spójności realizacji Strategii Lizbońskiej. Warto wiedzieć, że o 63% (do 50 mld EUR) zwiększono nakłady na VII Program Ramowy (badania naukowe).

Od roku 2007, zachowując środki służące przejrzystości i dyscyplinie operacji finansowych[8], podjęto próbę uproszczenia zarządzania

---

[7] W państwach o stosunkowo małym doświadczeniu w programowaniu i realizacji działań rozwojowych na dużą skalę, co gorsza o niskiej jakości systemu instytucjonalnego, jak Polska, zaprojektowano system wyjątkowo scentralizowany i niewydolny, który poddawano stałym usprawnieniom, przyczyniającym się do dodatkowych trudności realizacyjnych (zob. Grosse 2004, Kozak 2006).

[8] Jednym z narzędzi dyscyplinujących, zastosowanych eksperymentalnie w programach przedakcesyjnych m.in. w Polsce, była zasada N+2 (przejściowo dla EU12 N+3), wyznaczająca limit czasu na zakończenie realizacji projektów z danego roku budżetowego (roku N). W 2007 roku zastosowano tę zasadę także do projektów Funduszu Spójności, co jest uważane za rozwiązanie kontrowersyjne.

i wzmocnienia wymiaru strategicznego. Podstawą filozofii zarządzania jest wspieranie osiągania celów strategicznych, przy pozostawieniu większej niż dotąd swobody w dobieraniu niezbędnych środków i narzędzi i proporcjonalnym zwiększeniu odpowiedzialności państwa i regionu za ostateczny efekt. Otwarto wreszcie większe pole dla innowacyjności i zarządzania menedżerskiego. Symbolem tych zmian było po raz pierwszy opracowanie wytycznych strategicznych (czyli zaleceń, a nie, jak dotąd, nakazów) oraz rezygnacja z uzupełnień[9]. Na ogół zmiany te przyjęto w państwach członkowskich z dużym zadowoleniem, w niektórych przypadkach (np. w Polsce) zdecydowano się jednak zachować uzupełnienie jako dokument krajowy, nie unijny, nadając mu nową nazwę uszczegółowienia. To kolejny przykład, jak wielka potrafi być inercja administracji.

4. POLITYKA SPÓJNOŚCI A INNE POLITYKI

Politykę spójności zazwyczaj postrzega się samoistnie, mimo że jest tylko jedną z wielu polityk europejskich. Efekty polityki spójności w dużym stopniu zależą także od innych polityk. W gruncie rzeczy niezwykle trudno jest udzielić pełnej odpowiedzi na pytanie, jaki konkretnie wpływ na rozwój Unii ma polityka spójności, jaki zaś inne czynniki, jak inne polityki unijne czy koniunktura światowa. Stosowane w UE główne modele analityczne (QUEST, HERMIN) mają charakter prognozy (nic nie mówią o stanie rzeczy), a wartość ich predykcji wpływu polityki spójności na rozwój zależy od przyjętych założeń. Jeśli przyjrzeć się ich ustaleniom, to nawet z tych generalnie optymistycznie usposobionych modeli wyciągnąć można niepokojący wniosek, że w zasadzie wymierny wpływ polityki spójności trwa mniej więcej tyle, co jej wydatki (KE 2007). Tymczasem istnieje co najmniej kilka polityk, które mają oczywisty, choć nie zawsze doceniany, wpływ na stan rzeczy.

Pierwsza to Wspólna Polityka Rolna, konsumująca początkowo około 80, a dziś około 50% budżetu unijnego, przed którą postawiono cele: podniesienie wydajności rolnictwa, zapewnienie odpowiedniego standardu życia zatrudnionym w rolnictwie, zagwarantowanie stałości zaopatrzenia i utrzymania korzystnych cen dla konsumentów. Żaden z tych celów nie został osiągnięty[10], wobec czego poddano ją stopniowym reformom,

[9] Wymagane od państw dokumenty to narodowe strategiczne ramy odniesienia (NSRO) oraz programy operacyjne.

[10] Rolnictwo zaczęło produkować namiar żywności (generując koszty przechowania i utylizacji), rolnicy nadal zarabiają znacznie mniej niż zatrudnieni w innych sektorach, a mimo gigantycznych dopłat żywność na rynku europejskim należy do najdroższych na świecie (Weidenfeld, Wessels 1995).

a część budżetu przeznaczono na rozwój polityki regionalnej (szerzej: spójności). Biorąc pod uwagę, że rolnictwo w UE odgrywa coraz mniejszą rolę, wytwarza zaledwie ok. 4% dochodu i zatrudnia do 6% siły roboczej, nie ma uzasadnienia dla WPR w tak ogromnej skali. Tym bardziej, że efekty są dwuznaczne[11].

Druga bardzo ważna z punktu rozwojowego polityka dotyczy ochrony środowiska. Wysokie normy ochrony środowiska narzucane w UE wpływają na zachowania producentów, konsumentów i inwestorów, także publicznych. Stawiają też wysokie wymagania inwestycjom publicznym, co nastręcza niemałe problemy zwłaszcza w krajach prowadzących wielkie inwestycje infrastrukturalne (które lepiej rozwinięte państwa realizowały w przeszłości nieskrępowane tak rygorystycznymi przepisami) i nie dysponujących sprawną, doświadczoną administracją[12]. Ostatnie lata pokazały silne związki ochrony środowiska i polityki energetycznej UE. Ostatnie decyzje w tej sferze (przydział kwot gazów cieplarnianych na poszczególne państwa) mogą mieć fatalne skutki dla krajów słabiej rozwiniętych wskutek wpływu na koszt inwestycji.

Oczywiste znaczenie ma – znów zwłaszcza w krajach uboższych, budujących nowoczesną infrastrukturę – polityka transportowa, mogąca się przyczynić do przezwyciężenia negatywnych skutków peryferyjności i marginalizacji. Nie bez znaczenia są też polityki makroekonomiczne i monetarna (UGW), w rosnącym stopniu także B+R.

Wpływy i relacje między tymi politykami są trudno czytelne. Tabela 2 przedstawia próbę szacunku wpływu wybranych polityk unijnych na poziom spójności UE (zatem i na politykę spójności). Pokazuje też, jak niejednoznaczny bywa ten wpływ i jak trudno jest mierzalny.

Nie można jednak zapominać, że UE nie jest wyspą, a jej gospodarki podlegają wpływom zewnętrznym, na które każda reaguje nieco inaczej. Niezwykła koniunktura światowa ostatnich kilku lat miała zapewne większy wpływ na tempo rozwoju UE niż prowadzone przez nią polityki. Co ciekawe, niektóre państwa odniosły wyraźnie większe korzyści (Wielka Brytania, Holandia, Szwecja), niż pozostałe, co dowodzi, że

---

[11] Z jednej strony jej działania wspierają modernizację gospodarstw rolnych, z drugiej jednak pomoc dostaje się głównie najbogatszym właścicielom ziemskim (np. brytyjskiej królowej czy w Polsce byłemu senatorowi S.), co gorsza przez dotacje bezpośrednie sprzyja petryfikacji struktury własności rolnej, zwłaszcza w krajach o najbardziej zacofanej strukturze (np. Grecja, Litwa, Polska). W efekcie podtrzymuje przy życiu ekonomicznie nieopłacalne gospodarstwa.

[12] Zwłaszcza w państwach, w których przez dziesięciolecia administracja była narzędziem w ręku opresyjnej władzy, a nie instytucją służącą społeczeństwu. Wymóg konsultacji społecznych dla wielu instytucji administracyjnych wydawał się tylko utrudnieniem w pracy. Przykład Rospudy pokazał, że to najzupełniej błędne, a bywa, że i kosztowne, przekonanie.

Tabela 2. Szacunkowy wpływ polityk unijnych na trzy wymiary spójności (symbole: „0" wpływ neutralny, „+" pozytywny, „–" negatywny, „tp" – niedający się określić)

	Ekonomiczny	Społeczny	Terytorialny
WPR (wsparcie cenowe)		tp	
WPR (wsparcie dochodów)	0	tp/+	+
Rynek wewnętrzny, liberalizacja	+	–	tp
Rynek wewnętrzny, konkurencja	+	tp	tp
Makro i monetarna	+	tp	tp
Transportowa	0/+	0/+	–/+
Środowiskowa	–	+	tp
B+R	0	0	–

Źródło: na podstawie Molle 2007: 273.

jakość krajowego systemu instytucjonalnego ma wielkie znaczenie. Nie chodzi o pomniejszanie wagi polityk europejskich, ale raczej o ich nieprzecenianie. Przykładowo, w 2006 roku napływ środków unijnych netto do Polski wyniósł 2,71 mld EUR, prywatne transfery netto 4,74 mld, a bezpośrednie inwestycje zagraniczne 15,3 mld EUR[13]. Jednocześnie, jak co roku, około 4 mld EUR wydano z polskiego budżetu tylko na subwencje do ubezpieczenia społecznego rolników. Nawet w kwestii źródeł finansowania jest większe pole manewru niż nam się na ogół wydaje, a środki unijne to tylko fragment układanki. Mało znane w Europie są doświadczenia amerykańskie, z których wynika, że USA w ciągu mniej więcej 7 dziesięcioleci odnotowały wyraźną konwergencję[14]. Ważną cechą ostatnio prowadzonej amerykańskiej polityki regionalnej wobec słabiej rozwiniętych regionów (stanów) jest koncentracja na realizacji strategii rozwojowych, opartych na innowacji i rozwoju technologii, rozwój istniejących technologii, bardziej systemowe podejście (wsparcie planowania i rozwoju zdolności innowacyjnych), rozwój klastrów regionalnych, nacisk na uczenie się, benchmarking, wdrażanie przez PPP lub sieci regionalne (Shapira 2005). Podobne trendy odnotowuje się także w niektórych (lepiej rozwiniętych) krajach europejskich.

[13] UKIE, 2008, Gospodarka, efekty członkostwa, prezentacja, http://www.ukie.gov.pl/HLP/files.nsf/0/CC0321C3DC5F14D9C125748300282DD7/$file/2008.07.10Ukie-FINAL.pdf [31.07.2008].

[14] Niepełna porównywalność danych zmusza do ostrożności w porównaniach z Europą.

5. PODSUMOWANIE I WNIOSKI

Polityka spójności w swej niedługiej historii przeszła znaczące zmiany założeń, celów, metod, instrumentów. Nic nie wskazuje, by w przyszłości było inaczej.

Zmiany, jakim podlegała i podlega polityka spójności przez cały okres jej trwania, dokonywały się pod znaczącym wpływem procesu integracji europejskiej. W gruncie rzeczy jej nagły awans miał niewiele wspólnego z poziomem zróżnicowań terytorialnych, za to wiele z przypisaną jej funkcją narzędzia przeciwdziałania potencjalnym negatywnym skutkom Unii Gospodarczej i Walutowej. Przypomnijmy, że wzrost zróżnicowań po ostatnich akcesjach, spowodował, że zamiast proporcjonalnie zwiększyć budżet UE i polityki spójności, Rada zmniejszyła budżet Unii i prawie nie zwiększyła nakładów na politykę spójności. Cel wyrównywania różnic w rozwoju coraz bardziej przegrywa z potrzebą maksymalizacji efektywności działań.

Rosnące znaczenie ma też globalizacja i widoczne względne pogarszanie się pozycji konkurencyjnej Unii na scenie światowej. Odpowiedzią jest niezbyt skuteczna na razie Strategia Lizbońska, której realizacji od 2007 roku w znacznym stopniu podporządkowano środki polityki spójności, wymuszając dostosowania do celów SL i większe uwzględnienie nowego paradygmatu rozwoju.

Poziom spójności i tempo rozwoju terytorium Unii są wypadkową wielu czynników o charakterze krajowym, europejskim i światowym oraz zdolności systemów instytucjonalnych do ich absorpcji i wykorzystania na potrzeby rozwojowe. Postępująca globalizacja rodzi potrzebę bardziej systemowego podejścia do polityk rozwojowych Unii i mocniejszego powiązania ich z efektywnością.

Jakie wnioski płyną z dotychczasowych przemian dla ewaluacji polityki spójności?

Zmienność założeń, celów i narzędzi polityki spójności wymaga odpowiednich dostosowań po stronie ewaluacji. Rzecz główna polega na dostosowaniu i poprawie jakości kryteriów oceny, tak by lepiej mierzyły rzeczywisty poziom osiągnięć, uwzględniając także szerszy kontekst (w tym nowe czynniki rozwoju) i długofalowe konsekwencje badanych projektów i programów. Wymagać to będzie nie tylko technicznej poprawności ewaluacji, ale też dobrej znajomości trendów zachodzących w gospodarce światowej oraz koncepcji rozwoju leżących u podstaw interwencji. Rosnący nacisk na rozwój oparty na innowacjach, rozwoju przedsiębiorczości i zasobów ludzkich stawia też większe wymagania w kontekście innowacyjności samej ewaluacji. Wobec rosnącej złożoności przedmiotu badania na znaczeniu winna zyskać metoda triangulacji, pozwalająca łączyć różne metody i punkty widzenia. Zdecydowanie większy nacisk powinien być też położony na bliższe zintegrowanie

ewaluacji i zarządzania, tak by skutecznie eliminować przypadki nietrafionych przedsięwzięć (nie wspominając o nieprawidłowościach). Wreszcie ewaluacja, posługując się odpowiednio dobranymi wskaźnikami, mocno osadzonymi w nowym paradygmacie, winna w większym stopniu badać skutki długofalowe.

BIBLIOGRAFIA

Amin A., Thrift N. (eds.) (2000), *Globalization, Institutions and Regional Development in Europe*, Oxford, Oxford University Press.

Baj L. (2008), Polska między Kiribati a Macedonią, *GW*, Gospodarka, 4 lipca 2008, s. 29.

Balchin P., Sykora L., Bull G. (1999), *Regional policy and planning in Europe*, London, Routledge.

Blakely E.J., Bradshaw T.K. (2002), *Planning Local Economic Development. Theory and Practice*, Sage Publications Inc.

Dziemianowicz W., Jałowiecki B. (współpraca – M. Krajewska) (2004), *Polityka miejska a inwestycje zagraniczne w polskich metropoliach*, Warszawa, EUROREG i Wydawnictwo Naukowe Scholar.

European Commission (1999), *Evaluating Socio-Economic Programs*, Luxembourg.

European Parliament Fact Sheets, 3.1.0. Principles and general completion of the internal market, http://www.europarl.europa.eu/factsheets/3_1_0_en.htm (marzec 2008)

Gorzelak G. (2007), *Rozwój – region – polityka*, w: Gorzelak G., Tucholska A. (red.), *Rozwój, region, przestrzeń*, Warszawa, Ministerstwo Rozwoju Regionalnego.

Grosse T.G. (red.) (2004), *Polska wobec nowej polityki spójności Unii Europejskiej*, Warszawa, Instytut Spraw Publicznych.

GUS (2008), *Kwartalnik statystyki międzynarodowej*, luty 2008.

Harrison L.E., Huntington S.P. (red.) (2003), *Kultura ma znaczenie*, Poznań, Zysk i S-ka.

Hirszowicz M., Neyman E. (2004), Nowa bieda w świecie obfitości, *Kultura i Społeczeństwo* **2**.

Hryniewicz J.T. (2004), *Polityczny i kulturowy kontekst rozwoju gospodarczego*, Warszawa, Wydawnictwo Naukowe Scholar.

Jarosz M. (red.) (2005), *Wygrani i przegrani polskiej transformacji*, Warszawa, Oficyna Naukowa.

Keating M., Loughlin J., Deschouwer K. (2003), *Culture, Institutions and Economic Development, A Study of Eight European Regions*, Cheltenham UK, Northampton USA, Edward Elgar.

Komisja Europejska (2007), *Rozwijające się regiony, rozwijająca się Europa, 4 raport na temat spójności gospodarczej i społecznej*, Luksemburg.

Kozak M. (2006), System zarządzania europejską polityką regionalną w Polsce w pierwszym okresie po akcesji, *Studia Regionalne i Lokalne* **2**.

Kozak M. (2007a), *Rola funduszy strukturalnych w rozwoju regionów*, w: Gorzelak G., *Polska regionalna i lokalna w świetle badań EUROREG*, Warszawa, EUROREG i Wydawnictwo Naukowe.

Kozak M. (2007b), *Problem strategii w dokumentach rozwoju regionalnego*, w: Gorzelak G., Tucholska A. (red.), *Rozwój, region, przestrzeń*, Warszawa, MRR i EUROREG.

Kozak M. (2008), *Wykluczenie-rozwój-przestrzeń,* w: Jarosz M. (red.), *Wykluczeni. Wymiar społeczny, materialny i etniczny*, ISP PAN, Warszawa.

Kukliński A. (2007), *Towards a New Model of Regional Policy*, w: Landabaso M., Kukliński A., Roman C. (eds.), *Europe – Reflections on Social Capital, Innovation and regional Development*, Nowy Sącz, Wyższa Szkoła Biznesu – NLU.

Landes D.S. (2005), *Bogactwo i nędza narodów*, Warszawa, Muza SA.

Molle W. (2007), *European cohesion policy*, London and New York, Routledge.

Pietrzyk I. (2001), *Polityka regionalna Unii Europejskiej i regiony w państwach członkowskich*, Warszawa, WN PWN.

Shapira Ph. (2005), *Innovation Challenges and Strategies In Catch-Up Regions*, w: Fuchs G., Shapira Ph. (eds.), *Rethinking Regional Innovation and Change. Path dependency or Regional Breakthrough?*, Springer.

Strategia rozwoju kraju (2006), Warszawa, Ministerstwo Rozwoju Regionalnego.

Toffler A. (1997), *Trzecia fala*, Warszawa, PIW.

UKIE (2008), *Gospodarka, efekty członkostwa*, prezentacja, http://www.ukie.gov.pl/HLP/files.nsf/0/CC0321C3DC5F14D9C125748300282DD7/$file/2008.07.10UkieFINAL.pdf, pobrano: 31 lipca 2008.

Weidenfeld W., Wessels W. (1995), *Europa od A do Z*, Gliwice, Wyd. „Wokół Nas".

ROZDZIAŁ 1.3.

John Bachtler

EWALUACJA REGIONALNEJ POLITYKI W EUROPIE: KULTURA, ZAANGAŻOWANIE I POTENCJAŁ

1. WSTĘP

Ewaluacja ma w Europie długą historię. Podobnie jak w Ameryce Północnej, wzrost wydatków publicznych w latach 50. i 60., związany z rosnącym interwencjonizmem rządu wywołał potrzebę prowadzenia badań nad najlepszym sposobem wykorzystania funduszy publicznych w ramach wsparcia celów rządowych. Trend ten był szczególnie widoczny w Niemczech i Szwecji. Natomiast w latach 70. i 80. ubiegłego wieku, z powodu coraz silniejszych ograniczeń wydatków rządowych, ewaluacji zaczęto używać do rozliczeń – jako uzasadnienia ograniczenia, kontynuacji, lub racjonalizacji programów (Derlien i Rist 2002).

Większość modeli ewaluacji polityki regionalnej w Europie pochodzi z Wielkiej Brytanii oraz Niemiec i Holandii (Nicol 1982; Wadley 1986). Istotne bodźce do działań ewaluacyjnych w latach 90. ubiegłego wieku były następstwem polityki Unii Europejskiej, głównie w ramach rozporządzeń dotyczących Funduszy Strukturalnych, zgodnie z którymi państwa członkowskie UE oraz Komisja Europejska musiały oceniać produkty, rezultaty i oddziaływanie programów rozwoju regionalnego na różnych etapach realizacji cyklu wydatków. Te wymogi regulacyjne UE wywarły znaczący wpływ na ewaluacje krajowej polityki regionalnej, które stały się bardziej popularne w Europie w ciągu ostatnich 10–15 lat (Bachtler i Michie 1995; Bachtler i Wren 2006).

Niezależnie od nasilenia działań ewaluacyjnych w ostatnich latach, aktualne pozostaje pytanie o charakter kultury ewaluacyjnej w Europie. Mimo postępów w profesjonalizacji tej dyscypliny – odzwierciedlonej w tworzeniu stowarzyszeń ewaluacyjnych i nowych publikacji w zakresie ewaluacji – stopień instytucjonalizacji ewaluacji jest różny w różnych administracjach europejskich. Pojawiła się istotna luka pomiędzy bieżącymi debatami na temat teorii i metodologii ewaluacji a wymogami władz administracyjnych zamawiających takie ewaluacje.

W tym kontekście poniżej przedstawiona zostanie analiza ewaluacji polityki regionalnej w Europie na podstawie szerokiego przeglądu literatury oraz badań praktyk ewaluacyjnych, przeprowadzanych w krajowej polityce regionalnej państwa członkowskiego i w polityce spójności UE (Polverari et al. 2006; Polverari i Bachtler 2004; Raines i Taylor 2002). Rozdział rozpoczyna się od krótkiej dyskusji na temat kultury ewaluacji w Europie, ze szczególnym uwzględnieniem prac porównawczych Furubo et al. (2002). Następnie omówiona zostanie ewaluacja polityki regionalnej w różnych grupach państw – tych, w których polityka regionalna jest realizowana od dłuższego czasu, a także w państwach, w których ewaluacja polityki regionalnej rozpoczęła się w latach 90. oraz wschodzących praktyk ewaluacyjnych w państwach członkowskich Europy Środkowej i Wschodniej. Rozdział kończy się obserwacjami na temat przyszłości ewaluacji polityki regionalnej.

2. KULTURA EWALUACJI W EUROPIE

W ciągu ostatnich 20 lat w całej Europie rosło znaczenie ewaluacji polityki publicznej. Zjawisko to stanowi odzwierciedlenie kilku czynników. Największe znaczenie mają kontrola zasadności wydatków publicznych oraz odczuwana przez polityków potrzeba zademonstrowania, że rządy rozsądnie i dobrze wydatkują powierzone im pieniądze. Nacisk kładziony na ewaluację stanowił również element trendów w zarządzaniu sektorem publicznym w latach 80. i 90. ubiegłego wieku, opartym na kulturze postrzegania obywatela jako klienta lub konsumenta, promowania wyborów organizacyjnych w świadczeniu usług publicznych (zarówno w sektorze prywatnym, jak publicznym) i wymagającego większej odpowiedzialności.

Kolejnym czynnikiem jest rosnąca złożoność wyzwań stojących przed politykami, wymagająca, by mieli oni dostęp do właściwych, dokładnych i aktualnych informacji służących do opracowywania celów polityki, ustalania priorytetów i planowania instrumentów polityki. Regionalizacja i rosnąca instytucjonalizacja, zwiększająca się liczba graczy biorących udział w projektowaniu i wdrażaniu polityki oraz interakcje przez różnego rodzaju stosunki, partnerstwa i sieci na poziomie krajowym, regionalnym, lokalnym i gminnym także przyczyniły się do wzrostu

popytu ze strony interesariuszy na ewaluację w ramach rządu i poza jego sferą działania.

Niektóre z tych nacisków na ewaluację trwają już od dawna. Na przykład w RFN, wszystkie środki interwencji publicznej były już od lat 70. przedmiotem regularnych „kontroli sukcesów" (*Erfolgskontrollen).* W Holandii i Szwecji formalne rządowe i parlamentarne ewaluacje polityki są już od dawna stałym elementem kultury tworzenia polityki. Także w Wielkiej Brytanii ewaluacja polityki stosowana jest od ponad 30 lat. W ostatnich latach odnotowano wzrost zainteresowania tą problematyką (Polverari i Bachtler 2004). Na przykład w Holandii w 2002 roku utworzono nowe Ministerstwo Regulacji Finansowych ds. Wdrożenia i Ewaluacji Polityki (*Regeling Prestatiegegevens en Evaluatieonderzoek Rijksoverheid*) w celu zapewnienia, że funkcja ewaluacji została wdrożona w ramach rządu w sposób zadowalający, oraz w celu upewnienia się, że szacunki i rozliczenia budżetu departamentalnego są oparte na rzetelnych informacjach. Brytyjskie Ministerstwo Finansów opracowało w 2003 roku Zieloną Księgę *Oceny i Ewaluacji w Rządzie Centralnym*, zapewniającą wytyczne dla ministerstw rządowych i agencji wykonawczych, w której stwierdzono, że „wszystkie nowe polityki, programy i projekty o charakterze przychodowym, kapitałowym czy regulacyjnymi powinny zostać poddane wszechstronnej, ale proporcjonalnej ocenie w każdej sytuacji, gdy jest to praktyczne w celu promowania interesu publicznego" (HMT, 2003).

Od lat 90. ubiegłego wieku wiele innych państw europejskich – Austria, Dania, Finlandia, Irlandia, Włochy, Polska, Hiszpania – także przyznało ewaluacji większe znaczenie. Rosnący rynek z zakresie ewaluacji (Leuuw 2004) potwierdza fakt pewnej instytucjonalizacji ewaluacji w ramach ministerstw rządowych – przynajmniej w niektórych obszarach polityki – przez ustanowienie komórek ewaluacji, regularne zlecanie badań ewaluacyjnych w ramach cykli polityki oraz publikację sprawozdań ewaluacyjnych. W niektórych krajach, takich jak Dania, Finlandia i Irlandia, tego rodzaju proces instytucjonalizacyjny przebiegał w niezwykle szybkim tempie przez ostatnich 10 lat.

Znaczenie przypisywane ewaluacji stanowi część szerzej idących zmian w projektowaniu i realizacji polityki, które mają dla twórców polityki funkcje zarówno „informacji" jak „alokacji" (Derlien, Rist, 2002), takich jak:

- popyt na tworzenie polityki na podstawie dowodów w celu zrozumienia złożonych wyzwań dla polityki, postrzeganych jako dobre zarządzanie w odpowiedzi na wyzwania ekonomiczne (OECD, 2007a);
- wykorzystanie technik prognozowania strategicznego dla celów myślenia strategicznego i planowania polityki w warunkach rosnącej niepewności (np. w dziedzinie nauki i technologii, rozwoju regionalnego) w celu analizowania przyszłych wyzwań, z perspektywy

multidyscyplinarnej oraz w celu określenia opcji i strategii polityki (Gavigan, Scapolo 2002a, 2002b);

- zastosowanie budżetowania opartego na wynikach (ang. *performance based budgeting*) – opracowanie wskaźników, poziomów odniesienia (benchmarków) i celów w ramach procesu opracowywania budżetu, po to, aby przeprowadzić ewaluację oceny i zademonstrować wartości za pieniądze dla społeczeństwa (OECD, 2007b);
- *benchmarking* – porównanie wyników osiągniętych przez regiony, sektory czy obszary polityki w celu przeprowadzenia oceny względnych wyników oraz oceny potencjału lepszego wykorzystania zasobów;
- zarządzanie na podstawie wyników – zarządzanie zasobami organizacyjnymi oraz zastosowanie procesów w celu osiągnięcia właściwych i mierzalnych wyników (OECD, 2001).

Dyscyplina ewaluacji rozwijała się także wraz z powstawaniem stowarzyszeń ewaluacyjnych – w Wielkiej Brytanii (1994), Szwajcarii (1996), Niemczech (1997), Włoszech (1997), Finlandii (1999), Francji (1999), Danii (2000), Polsce (2000), Hiszpanii (2001), Szwecji (2003), Węgrzech (2006), Słowacji (2006), Czechach (2007), Holandii (2007) – oraz pojawianiem się nowych publikacji na temat ewaluacji, internetowych grup dyskusyjnych i sieci, opracowywaniem standardów i wytycznych ewaluacji oraz organizacją regularnych konferencji ewaluacyjnych. Jak stwierdziła Międzynarodowa Organizacja Współpracy w Ewaluacji (International Organisation for Cooperation on Evaluation), wszystko to stanowi część tendencji o zasięgu globalnym: „w latach osiemdziesiątych ubiegłego wieku istniały tylko trzy krajowe i regionalne stowarzyszenia ewaluacji, pod koniec lat dziewięćdziesiątych było ich dziewięć, a na początku XXI wieku ich liczba wzrosła do pięćdziesięciu”[1].

Istotnym bodźcem dla rozwoju ewaluacji w Europie była polityka UE. Wykorzystanie ewaluacji w ramach usług Komisji i na mocy określonych polityk wspólnotowych rosło od lat 80. ubiegłego wieku. Williams et al. (2002) określili „trzy szczegółowe grupy praktyk ewaluacyjnych” w politykach UE:

- wykorzystanie zarządzania cyklem projektu i metody analizy logicznej w zagranicznych projektach rozwojowych i innych politykach zewnętrznych (jak np. PHARE, TACIS);
- systemy ewaluacji i monitoringu z niezależnymi panelami ekspertów oraz metody przeglądów *peer review2* w politykach badań;
- ramy metodologiczne służące ocenie rozwoju regionalnego i interwencji w sprawie bezrobocia w ramach polityki spójności UE.

---

[1] http://ioce.net/overview/general.shtml

[2] (przyp. tłum: termin ten dosłownie oznacza „recenzja równych”; w praktyce oznacza system oceniania prac naukowych przed ich publikacją przez ekspertów z danej dziedziny pod kątem prawidłowości, ważności i oryginalności)

Inicjatywy takie zostały wzmocnione po przyjęciu ostatnich reform UE, promujących „rzetelne i wydajne zarządzanie", zachęcających do stosowania ewaluacji jako części skutecznego projektowania i wdrażania programów. Od 2000 roku służby Komisji mają obowiązek prowadzenia ewaluacji w odniesieniu do działań ukierunkowanych na odbiorców lub beneficjentów spoza instytucji europejskich, które to zobowiązania zostały ujęte w planie działania ewaluacji dla Dyrekcji Generalnej – Budżet Komisji Europesjkiej w 2007 roku (European Commission, 2000; 2007). Inne naciski międzynarodowe na zastosowanie ewaluacji w celu usprawnienia procesu projektowania i realizacji polityki były wywierane przez OECD, UNDP i Bank Światowy.

Wpływ UE na rozwój ewaluacji był szczególnie silny w Irlandii i Hiszpanii, które otrzymały znaczące finansowanie z Funduszy Strukturalnych i Funduszu Spójności UE (Derlien i Wrist 2002). W Irlandii O'Hara et al. (2007) uznali, że bez zewnętrznych nacisków ze strony UE (a także OECD), kultura ewaluacji w ogóle by się nie pojawiła. W Hiszpanii Fernández-Ramirez i Rebolloso (2006: 137) stwierdzili, że „dyrektywy Unii Europejskiej były podstawową siłą napędową" szybkiego rozwoju praktyki ewaluacyjnej od lat 90. ubiegłego wieku. To samo można powiedzieć na temat rozwoju ewaluacji w państwach członkowskich Europy Środkowej i Wschodniej.

Niezależnie od powyższych zdarzeń w europejskiej kulturze ewaluacji, niektórzy badacze argumentują, że w porównaniu do Ameryki Północnej dyscyplina ta nadal „raczkuje" z powodu niedostatecznego kształcenia ewaluatorów na poziomie uniwersyteckim, wykorzystania quasi-profesjonalnych zespołów ewaluacyjnych w badaniach wielonarodowych oraz nadmiernej uwagi przywiązywanej w badaniach ewaluacyjnych do zasobów i praktyk administracyjnych zamiast do produktów i wpływów (Leeuw 2004, Schröter 2004).

Tabela 1 przedstawia ranking państw na podstawie wskaźników kultury ewaluacyjnej, z uwzględnieniem stopnia, w jakim ewaluacja ma miejsce w różnych obszarach polityki, podaży ewaluatorów, krajowych dyskusji ewaluacyjnych oraz profesjonalizacji i instytucjonalizacji ewaluacji. W rankingu tym za *benchmark* przyjęto kulturę ewaluacji w Ameryce Północnej. Mimo postępów poczynionych w latach 90. ubiegłego wieku, kraje europejskie nadal w zasadzie pozostają w tyle. Jedynie Szwecji, Holandii i Wielkiej Brytanii udało się zbliżyć do ocen osiąganych w Stanach Zjednoczonych i Kanadzie, a kilka państw – jak Włochy, Irlandia i Hiszpania – otrzymało bardzo niskie oceny. Wiele państw zostało bardzo nisko ocenionych pod względem stopnia instytucjonalizacji ewaluacji (w ramach rządu lub parlamentu), ograniczonego udziału podejmowanych ewaluacji produktów czy (niedostatecznego) wykorzystania ewaluacji przez krajowe izby kontroli.

W tabeli 1 przedstawiono także znaczne rozbieżności występujące w Europie Zachodniej (państwa Europy Środkowej i Wschodniej nie

zostały uwzględnione w badaniu). Argumentowano też, że różnice w kulturze ewaluacji są związane z systemami prawa i administracji (Bœuf, 2001; Casavola i Tagle 2003). Na przykład tzw. kontynentalna struktura prawa i administracji, istniejąca we Francji, Włoszech, Portugalii i Hiszpanii, została uznana z mniej korzystną dla rozwoju kultury ewaluacji. Dla kontrastu anglosaski system prawa zwyczajowego i administracji opartej na zarządzaniu publicznym zostały uznane za bardziej sprzyjające rozwojowi takiej kultury, z położeniem nacisku na wartość pieniądza i odpowiedzialność.

Tabela 1. Ranking państw na podstawie wskaźników kultury ewaluacyjnej

	Wskaźniki kultury ewaluacyjnej									
	I	II	III	IV	V	VI	VII	VIII	IX	Razem
Stany Zjednoczone	2	2	2	2	2	2	2	2	2	18
Kanada	2	2	2	2	2	1	2	2	2	17
Szwecja	2	2	2	1	2	1	2	2	2	16
Holandia	2	2	2	1	2	1	2	2	1	15
Wielka Brytania	2	2	2	2	1	1	2	1	2	15
Niemcy	2	2	1	2	1	1	2	1	1	13
Dania	2	2	2	1	1	0	2	1	1	12
Francja	2	1	1	2	2	1	1	1	0	11
Finlandia	2	1	1	1	1	1	1	1	1	10
Szwajcaria	1	1	2	2	0	0	2	0	0	8
Irlandia	1	1	1	0	1	0	1	1	1	7
Włochy	1	1	1	2	0	0	1	1	0	7
Hiszpania	1	0	1	2	1	0	0	0	0	5

I – Ewaluacja ma miejsce w głównych obszarach polityki.
II – Podaż ewaluatorów w różnych dyscyplinach.
III – Krajowa dyskusja nad ewaluacją
IV – Organizacje zawodowe
V – Stopień instytucjonalizacji – rząd.
VI – Stopień instytucjonalizacji – parlament.
VII – Pluralizm instytucji lub ewaluatorów, wykonujących ewaluacje w każdym z obszarów polityki.
VIII – Ewaluacja w ramach Najwyższej Instytucji Kontroli.
IX – Udział ewaluacji produktów w ewaluacjach produktów i procesów.

Źródło: wyciąg z tabeli 1.1. w Furubo et al. (2002).

Inne czynniki, tłumaczące zróżnicowaną pozycję ewaluacji, obejmują podatność władz rządowych na wyniki badań empirycznych, stopień, w jakim orientacja i programowanie wyników stanowią naturalną część administracji, naturę dyskusji publicznej, dostępność danych i innych informacji oraz podaż szkoleń i studiów w zakresie ewaluacji (Beywl 2006).

Tradycje konstytucjonalne, prawne i administracyjne oraz wartości kulturowe mogą stanowić główne przeszkody, tak jak miało to miejsce w wypadku Belgii, Włoch i Hiszpanii. Historycznie niski status ewaluacji w Belgii, która nie została ujęta w rankingu w tabeli 1, przypisuje się zjawisku partiokracji (niechęci do wyzwań dla konsensusu politycznego), słabości parlamentu i federalizacji kraju, przy jednoczesnym braku koordynacji pomiędzy poziomami (Varone et al. 2005). W wypadku Włoch Stames (1998: 92) wyjaśnił brak tradycji ewaluacji bardzo prostymi słowami: „nikomu nie zależy na zademonstrowaniu skuteczności programu, a nawet jego wyraźnie widoczna porażka będzie tolerowana, jeżeli może zostać przypisana utajonemu konsensusowi". O ile w latach 90. ubiegłego wieku we Włoszech rosła świadomość potencjału ewaluacji, o tyle głównym jej ograniczeniem był wystarczający poziom wiedzy. Zainteresowanie ewaluacją rosło także w Hiszpanii od lat 80. ubiegłego wieku, wraz z pojawianiem się instytucji, publikacji i programów ewaluacyjnych; nadal jednak uważa się, że ewaluacja nie wyszła poza „wczesny etap rozwoju" (Pazos i Zapico-Gõni 2002). Fernández-Ramirez i Rebolloso (2006: 144) stwierdzili, że:

> „jest to rozszerzający się obszar w procesie konsolidowania i formalizowania, z mnogością profesjonalistów i naukowców, któremu nadal jednak brakuje tożsamości i który charakteryzuje się ograniczonym networkingiem. Wysiłki podejmowane przez nielicznych w celu uzgodnienia ewaluacji nałożyły się na kontrasty wsparcia politycznego poprzez brak definicji tego, co oznacza ewaluacja, jakie są jej techniczne kompetencje lub cele i procedury".

Irlandia zaszła znacznie dalej w instytucjonalizacji ewaluacji, szczególnie w drodze niedawnych reform służby publicznej, w tym jednak wypadku jedną z potencjalnych trudności jest brak systematycznych inicjatyw, podejmowanych w kierunku rozwoju umiejętności ewaluacyjnych (Boyle 2002). W kontekście Irlandii argumentowano, jak już wspomniano, że ewaluacja rozwinęła się głównie jako reakcja na naciski zewnętrzne, przy niewielkim „apetycie na politykę ukierunkowaną na dowody pomiędzy wyższymi rangą liderami politycznymi lub sektora publicznego" (O'Hara et al. 2007).

Nawet w krajach, które regularnie uzyskują wysokie oceny według tabeli 1, pojawiają się wątpliwości dotyczące kultury ewaluacji, szczególnie w odniesieniu do standardów ewaluacji. W Niemczech na przykład stwierdzono, że mimo prowadzenia działalności ewaluacyjnej od lat 60.,

zawód ewaluatora rozwinął się dopiero w połowie lat 90. ubiegłego wieku (Struhkamp 2005). W Danii, gdzie ewaluacja rozwijała się gwałtownie od lat 90., stając się „niemal rutyną w wielu obszarach sektora publicznego", jakość ewaluacji jest nadal uważana za wysoce niespójną (Albæk i Rieper 2002). W Szwecji, mimo długiej tradycji ewaluacji, wciąż występują duże różnice w działalności ewaluacyjnej pomiędzy sektorami (oraz instytucjonalizacji ewaluacji w ministerstwach rządowych), przy ograniczonej koordynacji i wymienia najlepszych praktyk (Forss 2008).

W porównaniu do Europy Zachodniej, państwa Europy Środkowej i Wschodniej nie mają tradycji ewaluacji polityki, a w przeszłości termin ten był kojarzony z monitoringiem, audytem i kontrolą. Podczas przygotowań do akcesji do UE podjęto znaczne wysiłki w celu wypracowania kultury ewaluacji, wspierane przez programy przedakcesyjne PHARE. Jednak szczególnie w latach 90. ubiegłego wieku zakres wykorzystania ewaluacji w ramach administracji rządowej był ograniczony brakiem stabilności instytucjonalnej, niekompletnymi reformami administracyjnymi, brakiem wykwalifikowanego personelu i doświadczenia w stosowaniu narzędzi i systemów rozwoju programów (Salamin i Peti 2002; Toulemonde i Bjornkilde 2003; Blazek i Vozab 2006).

Problemy te utrzymywały się jeszcze do niedawna. Wyniki przeprowadzonego w 2006 roku badania potencjału ewaluacji w nowych państwach członkowskich wykazały, że mimo usprawnień (takich jak dostępność ewaluatorów), kultura ewaluacji nie przyjęła się w stopniu porównywalnym z UE15, głównie ze względu na brak procedur zapewniających przyczynienie się ewaluacji do procesu tworzenia polityki (CSES, 2006). Podobnie jak w niektórych państwach członkowskich EU15, zastosowanie ewaluacji w kilku państwach Europy Środkowej i Wschodniej (jak np. Rumunia) zależy od bodźców zewnętrznych i raczej nie wydaje się, by sektor publiczny okazywał zainteresowanie rozwojem krajowej wiedzy w dziedzinie ewaluacji (Mihalache 2006)[3]. Natomiast przeprowadzona niedawno na Litwie ocena ustaleń ewaluacyjnych wykazała rosnący popyt na ewaluację ze strony władz rządowych, ograniczany niewielką podażą (Jarmalavičiūtė 2008); dyskusje w DG REGIO Evaluation Network sugerują, że te same zjawiska mogą występować w innych, niewielkich państwach[4]. O rosnącym zainteresowaniu profesjonalizacją tej dyscypliny świadczy fakt, że do istniejącego już od 1999 roku Polskiego Towarzystwa Ewaluacyjnego niedawno doszła Węgierska Sieć Ewaluacyjna, która rozpoczęła działalność we

[3] Podobne działania są podejmowane w państwach sąsiadujących z Unią Europejską, jak Mołdawia, Rosja, Turcja i Ukraina, gdzie działania ewaluacyjne są uzależnione od międzynarodowego networkingu ewaluacyjnego oraz wymogów ewaluacyjnych zagranicznych organizacji finansujących (Karimov et al., 2007; UNDP, 2007).

[4] Zob. protokół z zebrania DG REGIO Evaluation Network (oraz dyskusja z 2007–2013 ustalenia ewaluacyjne) pod adresem: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/tech_en.htm

wrześniu 2006 roku; Słowackie Towarzystwo Ewaluacyjne, utworzone w 2006 roku; Czeskie Towarzystwo Ewaluacyjne, utworzone w 2007 roku oraz Rumuńska Nieformalna Sieć Ewaluacyjna (obecnie w trakcie rejestracji).

3. EWALUACJA REGIONALNEJ POLITYKI W EUROPIE

3.1. Czynniki sprawcze ewaluacji polityki regionalnej

Ewaluacja polityki regionalnej jest podstawowym przykładem omówionych trendów. Twórcy polityki regionalnej w państwach europejskich stosują ewaluację w większym stopniu niż kiedykolwiek przedtem do oceny skuteczności i wydajności polityki, projektowania nowych polityk i wspierania wdrożenia programów i instrumentów polityki. Jest to silnie związane z głębokimi zmianami, jakie zachodziły w polityce regionalnej w ciągu ostatnich 15 lat, czasami określanych jako zmiana paradygmatu polityki regionalnej (Bachtler 2001; Pezzini 2003; Bachtler i Yuill 2007). W porównaniu do tradycyjnego podejścia do polityki, współczesne polityki regionalne charakteryzują się kilkoma szczególnymi cechami. Są one ukierunkowane jednocześnie na równość i wydajność, przenosząc centralny punkt polityki z redystrybucji na konkurencyjność. Faworyzują instrumenty strony podażowej i oddolne inicjatywy lokalnego rozwoju gospodarczego, ucieleśniające silniejsze ukierunkowanie przestrzenne, ale także tematyczne i sektorowe zasobów. Są one wdrażane i realizowane przez różnych (szeroko pojętych) graczy i mechanizmy, zgodnie z szerszą rolą przeznaczoną dla regionalnych, lokalnych publicznych i prywatnych aktorów, oraz planowane w wieloletnich ramach strategicznych.

Polityka regionalna przypisuje obecnie także większe znaczenie ewaluacji. Jak stwierdzono, odzwierciedla to tendencje organów publicznych do podejmowania odpowiedzialności za wydatki rządowe oraz rządów, wymagających dowodów w celu usprawnienia zarządzania i realizacji takich polityk. Z paradygmatem nowej polityki regionalnej łączy się integralnie potrzeba lepszego zrozumienia sposobu funkcjonowania gospodarki regionalnej oraz skuteczność różnych kombinacji wkładów do polityki, szczególnie w kontekście większego zregionalizowania zarządzania regionalnymi politykami rozwoju w wyniku dewolucji, dekoncentracji lub innych tendencji decentralizacyjnych.

Poza zmianami w projektowaniu i treści polityki regionalnej, istotną siłą sprawczą zmian polityki rozwoju regionalnego, a szczególnie zastosowania ewaluacji, była polityka spójności UE (Bachtler i Wren 2006; Bachtler i Michie 1995). Od czasu reformy funduszy strukturalnych w 1988 roku, ewaluacja polityki spójności otrzymała znacznie wyższy status w rozporządzeniach UE i wymaga od służb Komisji i władz

zarządzających w państwach członkowskich monitorowania i ewaluacji programów funduszy strukturalnych na różnych etapach cyklu zarządzania projektami. Od 1988 roku przygotowanie regionalnych programów rozwoju musiało obejmować ocenę *ex ante* oczekiwanych wpływów przewidzianych programem interwencji wraz ze skwantyfikowanymi celami, ewaluacji, która była w coraz większym stopniu podejmowana przez ewaluatorów zewnętrznych w okresach programowych 1994–1999 i 2000–2006. Działania te uzupełniała ewaluacja śródokresowa postępów w realizacji programów, podejmowana w połowie okresu programowego, realizowana przez niektóre programy w latach 1989–1993 i 1994–1999 oraz systematycznie przez wszystkie programy w latach 2000–2006.

Pierwszy okres zastosowania tej metody przyniósł różne doświadczenia. Początkowe oceny wykazały brak przestrzegania przez państwo członkowskie wymogów ewaluacyjnych, brak spójności metodologii ewaluacji oraz niedostateczny potencjał i umiejętności wewnątrz służb Komisji (Levy 1996). Jednak od początku lat 90. Komisja zaczęła wywierać silne naciski na państwa członkowskie w sprawie bardziej poważnego traktowania wymogów regulacyjnych, dotyczących ewaluacji (Bachtler i Mendez 2007; Raines i Taylor 2002) i podjęła ewaluację *ex post* wszystkich interwencji pod koniec każdego okresu ewaluacji. Komórka ewaluacji DG REGIO również zleciła szeroki zakres badań nad ewaluacją strategiczną i tematyczną oraz zapewniła wsparcie dla rozwoju potencjału ewaluacyjnego w postaci wytycznych, inwestycje w metodologie i zasoby ewaluacji[5], a także w organizację sieci administratorów ewaluacji i międzynarodowe konferencje na temat ewaluacji.

Osiągnięto trzy rodzaje efektów (Bachtler i Wren 2006; Polverari et al. 2006; Polverari i Bachtler 2004; Raines i Taylor 2002). Po pierwsze, władze krajowe i regionalne odpowiedzialne za zarządzanie funduszami strukturalnymi w całej Europie musiały spełnić wymogi ewaluacyjne UE, inwestując czas i zasoby w planowanie i zlecanie badań ewaluacyjnych – w wielu przypadkach po raz pierwszy. Wymogi pobudziły także pojawienie się podaży usług ewaluacyjnych oferowanych przez konsultantów, uniwersytety i ośrodki badawcze. Po drugie, mimo że wiele władz mogło przyjąć „podejście zgodności" do ewaluacji, spełniając swoje obowiązki dla celów związanych z odpowiedzialnością, są dowody coraz szerszego uznania wartości ewaluacji dla skutecznego projektowania i zarządzania programami. Po trzecie, skutkiem tych dwóch trendów był efekt „rozlania się" ewaluacji krajowej polityki publicznej, wynikający z przyjęcia procedur ewaluacji UE w niefinansowanych przez UE obszarach interwencji rządowych, a bardziej ogólnie wpływ na krajową kulturę ewaluacji.

---

[5] Program MEANS w latach 90. i zob. EVALSED w latach dwutysięcznych (http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evaluation_en.htm).

3.2. Krajowe studia przypadków ewaluacji polityki regionalnej

W celu ilustracji opisanych tendencji, w niniejszym rozdziale przedstawiono krótkie podsumowanie rozwoju ewaluacji w kilku państwach członkowskich UE. Podsumowanie, oparte na wynikach badań Polverari i Bachtlera (2004), rozpoczyna się od omówienia państw, w których ewaluacja polityki regionalnej ma długie tradycje, takich jak Niemcy, Wielka Brytania, Holandia, Szwecja. Następnie omówiono państwa członkowskie, w których ewaluacja polityki regionalnej rozpoczęła się w latach 90. – Austria, Finlandia, Irlandia, Włochy – na skutek wpływu zmiany polityki i wymogów funduszy strukturalnych. Na koniec przedstawiono ostatnie przykłady ewaluacji regionalnej w nowych państwach członkowskich, w tym Polski.

Polityka regionalna była w **Niemczech** (podobnie jak w innych obszarach wsparcia polityki) przez ponad 30 lat poddawana regularnej „kontroli sukcesu" (*Erfolgskontrolle*). Podejście przyjęte do ewaluacji łączyło w sobie roczny monitoring na poziomie projektów w celu oceny inwestycji i zatrudnienia w związku z korzystającymi ze wsparcia projektami, oraz badania na poziomie krajowym lub regionalnym, oceniające wpływy ekonomiczne polityki interwencji na różnice regionalne. Jednak w latach 80. ubiegłego wieku doszło do wstrzymania prób ewaluacji na poziomie mezo, polegających na analizie pomocy regionalnej otrzymywanej przez indywidualne firmy z powodu przyjęcia ustawodawstwa o ochronie danych. Od połowy lat 90. specjalistyczny instytut (IAB) rozpoczął opracowywanie na zlecenie Federalnego Ministerstwa Gospodarki nową koncepcję ewaluacji, obejmującą sondaże firm i badania w obszarze rynku pracy (Deutscher Bundestag, 2007). Umożliwiło to dopasowanie danych w szeregach czasowych, dotyczących firm, którym udzielono pomocy, do grupy odniesienia, składającej się z około siedmiu milionów firm, pod względem zatrudnienia, inwestycji, płac, stabilności firm oraz wkładu do regionalnego dochodu i zatrudnienia, zapewniając przy tym nowe informacje na temat skuteczności instrumentów polityki regionalnej (IAB, 2004).

Niezależnie od tej tradycji ewaluacji polityki regionalnej w odniesieniu do pomocy regionalnej, należy tutaj zwrócić uwagę na fakt, że od 1989 roku fundusze strukturalne poszerzyły wykorzystanie ewaluacji na inne obszary rozwoju regionalnego w Niemczech. Stwierdzono, że wartość przypisywana ewaluacji w Niemczech zmieniała się gwałtownie na podstawie doświadczenia w ocenie funduszy strukturalnych w latach 1994–1999, aczkolwiek z pewnymi uwagami co do sposobu spełnienia obowiązku ewaluacji w UE (Scheffel 2001). Jakoby (2006) stwierdził, że na poziomie landów, w Północnej Nadrenii-Westfalii „systematyczne ewaluacje polityki regionalnej nie były podejmowane przed fundamentalną reformą funduszy strukturalnych w latach 1988/89"; od tamtej pory ewaluacja była w coraz większym stopniu uznawana ze względu

na jej potencjał „uczenia się" oraz zakres katalizujących zmian polityk i programów.

Wielka Brytania ma już od dawna kulturę ewaluacji, zwłaszcza w obszarze polityki regionalnej, gdzie ewaluacja była prowadzona przez ponad 40 lat. Skuteczność i wydajność polityki regionalnej jako całości oraz indywidualnych instrumentów polityki były w tym okresie poddawane okresowej ewaluacji (zob. np. Arup 2000), głównie przez badania wspomaganych firm oraz modelowanie, z silnym naciskiem na analizę tworzenia zatrudnienia oraz koszty podejmowanych działań w przeliczeniu na miejsce pracy. W ciągu ostatniej dekady ewaluacja otrzymała nowy impuls w postaci zaangażowania rządu brytyjskiego w „tworzenie polityki na podstawie dowodów", co wymagało, by nowe polityki były od samego początku realizacji poddawane ewaluacji ich skuteczności (Cabinet Office, 2000). Stanowiło to element szerszej agendy reformy polityki i modernizacji służby publicznej, w których ewaluacja była stosowana przez rząd centralny jako „agent zmiany", na przykład w celu zachęcenia władz lokalnych do poprawy wyników świadczenia usług (Martin i Sanderson 1999).

Siłą napędową polityki regionalnej była także zmieniająca się przez rząd brytyjski ocena „problemu regionalnego", czyli postrzegania różnic regionalnych głównie jako wyniku długoterminowych różnic w osiąganych wynikach w stosunku do kilku „driverów" produktywności, przy jednoczesnym żądaniu bardziej nowoczesnego podejścia do ewaluacji, z uwzględnieniem horyzontu długoterminowego o różnych poziomach złożoności (Ling 2003). Dalsze bodźce były związane z dewolucją (do Szkocji, Walii i Irlandii Północnej) oraz regionalizacją w ramach Anglii, co prowadziło do powstania nowych instytucji i strategii rozwoju regionalnego oraz powiązanego z nimi popytu na analizę, monitoring i ewaluację w różnych skalach przestrzennych (Martin i Thurston 2001). Ewaluacja w skali całej Wielkiej Brytanii jest uregulowana postanowieniami *Zielonej Księgi*, wydanej przez Ministerstwo Finansów, zapewniającymi wiążące wytyczne dla ministerstw i agend wykonawczych odnośnie do oceny i ewaluacji. Ministerstwo Przedsiębiorczości i Reformy Regulacyjnej (dawniej Ministerstwo Handlu i Przemysłu), które odpowiada za politykę regionalną, ma swój własny Zespół ds. Wyników i Ewaluacji (Performance and Evaluation Team), koordynujący toczący się program ewaluacji. Podobne jednostki działają w zdecentralizowanych jednostkach administracji.

W **Holandii** polityka ewaluacji jest już od wielu lat częścią reżimu polityki regionalnej. Polityka regionalna jest realizowana w cyklach cztero- lub pięcioletnich. Zgodnie z powszechnie przyjętą praktyką, instrumenty polityki są poddawane ewaluacji pod koniec każdego cyklu w celu poprawy i ponownego ukierunkowania przyszłej polityki. W jej wyniku powstaje memorandum polityki przestrzennej, określające cele i priorytety polityki (zob. np. Ministerstwo Spraw Ekonomicznych, 2000; 2004). Dodatkowym wzmocnieniem było przyjęcie pod koniec lat 90. ubiegłego

wieku przez Ministerstwo Finansów systemu kontroli zarządzania, obejmującego cały rząd, tworzącego ramy budżetowe, w których ministerstwa rządu centralnego musiały połączyć ze sobą swoje cele, działania oraz alokację zasobów w celu zapewnienia „odpowiedzialności ministra za wydajność i przydatność polityki i *jej* okresowego badania" (Ministerstwo Finansów, 1999; 2002). W konsekwencji Ministerstwo Spraw Ekonomicznych (odpowiedzialne za politykę regionalną) opracowało własny *Podręcznik Ewaluacji i Wdrożenia Polityki* (*Handboek Beleidsevaluatie & Prestatiegegevens*), zawierający wytyczne dla ewaluacji, którymi kierowano się podczas ostatnich ewaluacji wszystkich głównych działań podejmowanych w ramach polityki regionalnej oraz polityki jako całości (IBO, 2004).

W **Szwecji** ewaluacja została umiejscowiona w procesie tworzenia polityki w postaci Krajowych Przeglądów Publicznych (SOU), tzn. ewaluacji realizowanych przez Komisje Parlamentarne w celu przeglądu polityki lub jej realizacji. Ostatni taki przegląd polityki regionalnej został przeprowadzony w 2000 roku i posłużył jako podstawa dla fundamentalnej zmiany polityki regionalnej w następnym roku (Regeringskansliet, 2000). Niezależnie władze krajowe także podejmują okresowe ewaluacje konkretnych instrumentów lub programów polityki. Przystąpienie do UE i wdrożenie programów funduszy strukturalnych podniosły percepcję znaczenia ewaluacji wśród twórców polityki i wpłynęły także na sposób zarządzania ewaluacjami. Spowodowało to na przykład wzrost wykorzystania zaproszeń do przetargów, zawierających dokładne specyfikacje zadań do wykonania; promowało ponadto zastosowanie harmonogramu ewaluacji *ex ante*/*mid-term*/*ex post* także w odniesieniu do polityki krajowej (np. Regionalne Programy Wzrostu krajowej polityki regionalnej zostały poddane ewaluacji w ramach tego cyklu).

Zwiększone zainteresowanie ewaluacją wynika także z ustawodawstwa dotyczącego polityki regionalnej z 2001 roku, którego celem było poprawienie bazy analitycznej dla rozwoju polityki. Zamówiony przez rząd raport Szwedzkiego Instytutu Badań nad Polityką Wzrostu (ITPS) zalecał, by gracze centralni i regionalni zostali wyposażeni w systematyczne instrumenty w postaci holistycznych podstaw analitycznych i nadrzędnych celów. W dokumencie zaproponowano wspólny zbiór wskaźników, stosowanych przy formułowaniu rządowej polityki rozwoju regionalnego, a także służących jako podstawa do określenia szczególnych regionalnych problemów rozwojowych czy niewykorzystanego potencjału rozwoju, czy też jako podstawa do działań następczych i ewaluacji (Kolmodin i Melkersson 2004).

Przeciwieństwem takich ugruntowanych tradycji ewaluacji polityki regionalnej są praktyki stosowane przez niektóre państwa europejskie, w których działania ewaluacyjne pojawiły się raczej niedawno, szczególnie po latach 90. ubiegłego wieku, często w związku z wymogami ewaluacyjnymi polityki spójności UE.

Austria jest jednym z państw należących do tej grupy. Ewaluacja polityki regionalnej – i bardziej ogólnie ewaluacja polityki – zostały opracowane głównie w związku z wymogami UE (Holzinger 2001). Poza działalnością Austriackiego Trybunału Audytorów, który koncentruje się głównie na aspektach finansowych i prawnych, stosowano tam jedynie ograniczoną ewaluację polityki regionalnej. Huber (2006) stwierdził, że przed przystąpieniem do UE ewaluacja nie była wymagana prawem ani nie stanowiła części kultury politycznej. O ile niektóre instrumenty polityki regionalnej były poddawane ewaluacji, o tyle działania takie były dobrowolne i niesystematyczne. Istniał także zdecydowany sceptycyzm odnośnie do ewaluacji polityki regionalnej (Holzinger 2001). Drugim czynnikiem wyjaśniającym jest fakt, że krajowa polityka regionalna Austrii jest względnie ograniczona pod względem zasobów i instrumentów, nie tylko ze względu na niezbyt ostry charakter różnic regionalnych, ale także z powodu skutecznych mechanizmów wyrównawczych zastosowanych przez redystrybucję funduszy między regionami. Istotny jest także fakt, że długotrwałe stosunki współpracy pomiędzy austriackimi instytucjami rządowymi a innymi aktorami, sprzyjały raczej tradycji dobrowolnego uczenia się przez nieformalne sieci niż formalnym praktykom ewaluacyjnym (Huber 2006).

Od przystąpienia do UE ewaluacja stała się bardziej zintegrowana z zarządzaniem i wdrażaniem polityki, zarówno w odniesieniu do funduszy strukturalnych, jak i programów krajowych. Huber (2006: 279) stwierdza, że „Austria postanowiła skorzystać z możliwości rozpoczęcia wspólnego procesu uczenia się zamiast pójścia łatwą drogą i ograniczenia ich wysiłków ewaluacyjnych do minimum, wymaganego dla czysto formalnych celów". Wytyczne dla krajowych inicjatyw polityki regionalnej (*Richtlinien*) obejmują obecnie szczegółowe postanowienia odnośnie do wdrożenia zarówno interwencji, jak i ich ewaluacji, a inicjatywy te są systematycznie poddawane ewaluacji. Ustanowiono także forum – wstępnie znane jako „Checkpoint EVA" do 1999 roku, następnie jako „KAP EVA" w latach 2000–2006 – które ułatwiło współpracę nad ewaluacją pomiędzy poziomem federalnym a poziomem landów (przy czym landy odpowiadają za programy funduszy strukturalnych oraz wiele inicjatyw rozwoju regionalnego), podniosło potencjał ewaluacji wśród administratorów i ewaluatorów oraz w sposób bardziej ogólny przyczyniło się do rozwoju kultury ewaluacji w Austrii (Strohmeier i Holzinger 2004; 2006).

W **Finlandii**, podobnie jak w innych obszarach polityki regionalnej, ewaluacja polityki regionalnej znajdowała się pod silnym wpływem UE. Mimo że badania ewaluacyjne można prześledzić wstecz do lat 70. ubiegłego wieku, Ahonen et al. (2002) argumentują, że pierwsze wszechstronne ewaluacje zostały przeprowadzone przez rząd na początku lat 90. w ramach oceny reform administracyjnych dokonanych przez Ministerstwo Finansów, oraz że „adwent" prawdziwej ewaluacji programów

w rządzie Finlandii zbiegł się w czasie z przystąpieniem Finlandii do Unii Europejskiej" (ibidem: 50). Zastosowanie ewaluacji w tworzeniu polityki regionalnej zaczęło się gwałtownie rozwijać po przystąpieniu Finlandii do UE w 1995 roku, szczególnie ze względu na wzajemne stosunki pomiędzy polityką krajową a programami UE (fundusze strukturalne są jednym ze źródeł finansowania strategicznych programów regionalnych) oraz na utworzenie Rad Regionalnych, zarządzających funduszami strategicznymi oraz krajowymi programami rozwoju regionalnego.

Kolejnym, związanym z powyższym czynnikiem, była ewolucja polityki regionalnej w Finlandii. Na początku XXI wieku rozpoczęto szeroki przegląd polityki rozwoju regionalnego, który doprowadził do zmiany fińskiego podejścia do polityki regionalnej. W ramach ustawy o rozwoju regionalnym, przyjętej w 2002 roku, wprowadzono serię wymogów obowiązujących w stosunku do ewaluacji wszelkich programów podlegających ustawie. Ewaluacji prowadzonej przez ewaluatorów zewnętrznych miały podlegać trzy odrębne komponenty, przy czym ewaluacja taka miała odbyć się przynajmniej raz podczas lub pod koniec okresu ich realizacji. Były to: strategiczne czteroletnie programy regionalne, opracowywane i realizowane przez każdą z Rad Regionalnych; regionalne plany i cele, uzgadniane i wdrażane przez określone sektory; oraz specjalne programy, przyczyniające się do osiągnięcia krajowych celów rozwoju regionalnego (takich jak Program Centrum Wiedzy Fachowej, Program Rozwoju Centrum Regionalnego, Program Polityki dla Obszarów Wiejskich oraz Program Rozwoju Wysp). Ciągła ocena skuteczności polityki, zwłaszcza przez ewaluację, stanowiła także istotny element wprowadzenia w życie ustawy o pomocy dla biznesu (2000), która regulowała przyznawanie pomocy regionalnej (oraz bardziej ogólnie – pomoc dla firm). Na koniec, w 2004 roku, w Decyzji Rządu stwierdzono, że „monitoring, skuteczność, ewaluacja i przewidywanie odnośnie do alokacji rządowych na rzecz rozwoju regionalnego w różnych sektorach administracyjnych zostaną wzmocnione", co częściowo dotyczy corocznej ewaluacji wpływu polityki regionalnej na propozycje, składane corocznie w odniesieniu do budżetu Państwa (Ministerstwo Spraw Wewnętrznych, 2004). Łącznym efektem tych kroków był szeroki program regularnej ewaluacji zarówno programów funduszy strukturalnych, jak i inicjatyw krajowych, które zapewniły istotne informacje na temat skuteczności programów polityki regionalnej (OECD, 2005).

Irlandia stanowi jeden z najbardziej uderzających przykładów wpływu zasad polityki spójności UE na zastosowanie i kulturę ewaluacji. Podobnie jak w wypadku Austrii i Finlandii, w latach 70. i 80. ubiegłego wieku ewaluacja polityki i programów była ograniczona i miała peryferyjne znaczenie dla procesu decyzyjnego, mimo istnienia względnie dobrze rozwiniętego systemu ewaluacji projektów (Tutty 1994). Zapotrzebowanie na ewaluację programów finansowanych przez UE oraz programu reform władz publicznych przyczyniły się do znaczącej zmiany

roli i wagi ewaluacji w Irlandii, uwzględniając nie tylko politykę regionalną, ale także całą, szeroko pojętą politykę publiczną.

W latach 90. ubiegłego wieku wprowadzono ustawodawstwo w celu zapewnienia okresowego przeglądu wydatków publicznych oraz promocji rozwoju procedur ewaluacyjnych w ministerstwach. Wprowadzona w 1993 roku ustawa o kontrolerze i audytorze generalnym (poprawka) oraz wprowadzona w 1997 roku ustawa o zarządzaniu służbą publiczną, zapewniły łącznie mandat prawny oraz wymóg rozwoju ewaluacji we wszystkich ministerstwach. W konsekwencji, w 1997 roku przyjęto *Inicjatywę Przeglądu Wydatków* (*Expenditure Review Initiative* – ERI) w celu zapewnienia systematycznej analizy osiągnięć dzięki wydatkom w ramach każdego programu oraz zapewnienia podstaw do podejmowana opartych na dobrych informacjach decyzji w sprawie priorytetyzacji wydatków w ramach programów i pomiędzy nimi. W ramach ERI wszystkie obszary wydatków są poddawane przeglądowi przynajmniej raz na trzy lata. Ponadto wprowadzono obowiązek przeglądów polityki przemysłowej w odstępach trzyletnich w celu przeprowadzenia analizy wdrażanej polityki, zarysowania zmian kontekstowych oraz określenia koniecznych zmian kierunku przyszłej polityki.

Fundusze strukturalne wywarły zasadniczy wpływ na rozwój irlandzkiego podejścia do ewaluacji (Hegarty 2003). Niezależną ewaluację jednostek ustanowiono na początku lat 90. ubiegłego wieku w ministerstwach zarządzających programami operacyjnymi zasobów ludzkich i rozwojem przemysłowym, a w latach 90. podjęto szeroki program działań ewaluacyjnych w związku z funduszami strukturalnymi. Działania te rozszerzyły się na krajowe praktyki ewaluacyjne. Należy tutaj podkreślić, że w ramach Krajowego Planu Rozwoju na lata 2000–2006 przyjęto metodę ewaluacji na podstawie regulacji dotyczących funduszy strukturalnych z 1999 roku, mimo że kwota zasobów NDP, zapewniona przez fundusze europejskie, była względnie niska. Podobnie jak w wypadku wymogów UE, metoda ta działała na podstawie cyklu badań *ex ante,* międzyokresowych, i *ex post*, uzupełnionych programem bieżących ewaluacji tematycznych.

Na koniec, w ramach tej grupy państw, ewaluację polityki regionalnej we **Włoszech** można powiązać z procedurami i praktykami, wprowadzonymi przez fundusze strukturalne. Historycznie rzecz ujmując, Włochy nie miały silnej kultury ewaluacji, przynajmniej w odniesieniu do rozwoju regionalnego i ekonomicznego (inne sfery polityki, takie jak sektor zdrowotny czy polityka rynku pracy mają silniejsze tradycje ewaluacji). Do końca lat 80. ewaluacja była związana głównie z projektami oraz badaniami *ex ante* w zakresie rozwoju regionalnego i gospodarczego. Nie prowadzono ewaluacji *ex post* ani ewaluacji w skali całych programów. Dopiero po wprowadzeniu w życie w 1988 roku regulacji dotyczących funduszy strukturalnych i związanych z nimi wymogów, ewaluacja objęła całe programy.

Podobnie jak Irlandia, także Włochy przyjęły szczególną metodę zastosowania obowiązków wobec UE jako części proaktywnej (chociaż nie zawsze spójnej) strategii budowy potencjału ewaluacji. Casavala i Tagle (2003: 2) stwierdzili, że: „Włochy zaczęły budować swój potencjał od szczególnie niskiego poziomu. Wymogi europejskie odegrały istotną rolę w kształtowaniu sposobu prowadzenia ewaluacji". Szczególnie twórcy polityki regionalnej, zatrudnieni w Ministerstwie Rozwoju i polityki spójności zastosowali ramy ewaluacji przewidziane dla funduszy strukturalnych jako dźwignię służącą rozwojowi zastosowania ewaluacji we wszystkich politykach przyjmowanych dla regionów problemowych („obszary niedostatecznie wykorzystane"). W ramach szerszej, „strategicznej reformy włoskiej polityki regionalnej w celu poprawy skuteczności i wydajności interwencji, usprawnienia praktyk ewaluacyjnych były postrzegane jako sposób na podwyższenie weryfikacyjności polityki i usprawnienia procesu projektowania polityki" (Barca, 2006). Dokonano tego zwłaszcza przez budowanie zapotrzebowania na ewaluację oraz przez generowanie większego rozumienia zastosowania ewaluacji pośród osób zaangażowanych w programowanie i wdrożenie, głównie przez budowanie potencjału i działania networkingowe. Składanie parlamentowi raportów przez ministerstwa zapewniające pomoc regionalną (Ministerstwo Gospodarki i Finansów, Ministerstwo Działalności Produkcyjnej) w coraz większym stopniu nabiera charakteru ewaluacji. Wstępne oceny nowej strategii zostały ocenione jako „wyjątkowe" (Barca 2006), zachęcając do przyjęcia bardziej przejrzystego procesu decyzyjnego w świetle oczekiwanych i faktycznych wyników i większej odpowiedzialności. Wpływ długoterminowy jest mniej przejrzysty, szczególnie w odniesieniu do kwestii, czy wprowadzenie ewaluacji zostało faktycznie wbudowane w strukturę władz rządowych, odpowiedzialnych za rozwój regionalny na poziomie krajowym i regionalnym.

W przypadku państw Europy Środkowej i Wschodniej ewaluacja dopiero raczkuje. Jednak także w ich przypadku wymogi ewaluacyjne polityki spójności UE odgrywają istotną rolę w zachęcaniu do szerszego wykorzystania ewaluacji w dziedzinie rozwoju regionalnego. Na największą uwagę spośród tych państw członkowskich zasługuje **Polska**, gdzie ewaluacja polityki regionalnej rozwijała się szybciej i w bardziej istotny sposób niż w którymkolwiek z pozostałych nowych państw członkowskich. Ewaluacja polskiej polityki regionalnej została wprowadzona początkowo na mocy przepisów PHARE, a następnie przepisów polityki spójności. Jednak na początku tej drogi, o ile urzędnicy administracji centralnej byli gotowi do wdrożenia procedur monitoringu i ewaluacji, potencjał ewaluacji nie istniał. Olejniczak (2002: 8) stwierdził, że: „nie istnieją gracze zdolni to robić, ani wiedza teoretyczna i metodologiczna na temat ewaluacji polityk publicznych. Koncepcja ewaluacji interwencji czy wsparcia publicznego była nadal obca nie tylko przedstawicielom

administracji, ale także świata akademickiego". Kolejną trudnością było upolitycznienie systemu.

Mimo tego, w ramach regionalizacji administracji publicznej oraz wprowadzenia planowania strategicznego w okresie przedakcesyjnym, wymogi ewaluacyjne zostały także wprowadzone jako element krajowego ustawodawstwa, regulującego rozwój regionalny. W ministerstwach utworzono jednostki ewaluacyjne, wypracowano system monitoringu i zlecono pierwsze ewaluacje (Olejniczak 2002). Po akcesji działalność ewaluacyjna nabrała tempa; według obliczeń Krajowej Jednostki Ewaluacji, w latach 2004–2006 przeprowadzono 135 badań ewaluacyjnych związanych z polityką spójności, aczkolwiek większość z nich została zlecona przez jednostki ewaluacyjne działające przy zaledwie dwóch instytucjach zarządzających rozwojem regionalnym i zasobami ludzkimi (Ministerstwo Rozwoju Regionalnego, 2008). Według Jednostki, podaż usług ewaluacyjnych i jakość ewaluacji wzrosły, również dzięki wsparciu inwestycji w szkolenia i regularne konferencje ewaluacyjne, promujące wymianę doświadczeń w skali krajowej i międzynarodowej.

4. WNIOSKI

Istotnym czynnikiem rozwoju ewaluacji w Europie była Unia Europejska. Toczące się stale debaty na poziomie UE na temat zarządzania finansowego i dyscypliny budżetowej sugerują, że ewaluacja, wraz z pomiarami wyników, analizą kosztów i korzyści oraz oceną wpływów, zachowają w przyszłości swój istotny wpływ na tworzenie polityki UE[6]. W odniesieniu do polityki regionalnej, przepisy funduszy strukturalnych odegrały istotną rolę w wielu państwach członkowskich UE w promowaniu kultury ewaluacji wraz z większą świadomością ewaluacji i zaangażowaniem w nią. Jednak niektórzy komentatorzy kwestionowali fakt wbudowania tego zaangażowania, oraz perspektywy przetrwania redukcji lub wycofania interwencji funduszy strukturalnych. Podejścia, podejmowane przez państwa członkowskie w okresie programowania 2007–2023, mogą dostarczyć pierwsze wskazówki dotyczące zasadności tych uwag: wymogi regulacyjne UE zostały zaadaptowane w sposób mniej preskryptywny i powierzają państwom członkowskim większą odpowiedzialność (w pewnych granicach) za określenie, kiedy i w jaki sposób podejmowana jest ewaluacja programu. Może to zachęcić do bardziej przemyślanego i skutecznego wykorzystania ewaluacji przez władze krajowe i regionalne; może też jednak skutkować obniżeniem znaczenia przypisywanego ewaluacji, szczególnie w krajach/regionach otrzymujących ograniczone finansowanie UE.

[6] Zob. np. wypowiedzi państw członkowskich w bieżącej dyskusji na temat reformy budżetu http://ec.europa.eu/budget/reform/issues/read_en.htm

Należy także zwrócić uwagę na ograniczenia bieżącej praktyki ewaluacyjnej. Leuuw (2004) wskazał, że ewaluacje europejskie wykazywały tendencje do skupiania się raczej na procesach i wynikach niż na wpływach, co wpływało na stopień, w jakim ewaluacja może przyczynić się do projektowania polityki. Doświadczenia Holandii są typowe także dla innych państw: pięć lat po wprowadzeniu nowego holenderskiego podejścia do odpowiedzialności za politykę, w oficjalnej ocenie stwierdzono, że o ile „Budżety polityki oraz odpowiedzialność za politykę stymulowały ludzi do lepszego przemyślenia zamierzonych wyników polityki (...) związek pomiędzy celami, instrumentami i zasobami nadal nie jest wystarczająco jasny" (IOFEZ, 2004: 42). Ponadto, stan ewaluacji rządu centralnego oceniono jako rozczarowujący: „odbywa się wiele ewaluacji *ex post*, ale nie wiąże się to z przeglądem skutków polityki, a badania *ex ante* prawie nie są podejmowane" (ibidem). Niezależnie od potencjału przeprowadzenia liczniejszych ocen wpływów, istotne jest zrozumienie ograniczeń tego, co jest możliwe, szczególnie w świetle złożoności, nakładania się i wzajemnych związków obecnego rozwoju polityk, co sugeruje, że oczekiwania co do wyników takiej ewaluacji mogą być zbyt wysokie. Istnieje także możliwość wystąpienia niezamierzonych lub niepożądanych skutków, związanych z położeniem (nadmiernego) nacisku na ewaluację, gdy skuteczne zarządzanie organizacyjne lub polityka mogą utrudnić koncentrowanie się na miarach wyników i metodach ewaluacji, które nie ułatwiają dialogu i uczenia się (Leuuw 2000; Van Der Knaap 2006).

W tym kontekście istotną kwestią stają się metody ewaluacji. W niniejszym rozdziale skoncentrowano się głównie na kulturze ewaluacji w Europie, przywiązując niewielką wagę do metod ewaluacji polityki regionalnej. Przedstawiona zmiana paradygmatu w polityce regionalnej stanowi zasadnicze wyzwanie dla ewaluacji, wraz z wymogiem przyjęcia nowych lub innych rozwiązań metodologicznych, które dopiero muszą zostać docenione w praktyce ewaluacyjnej. Diez (2002), Stern (2003) oraz McVittie i Swales (2006) podają kilka cech polityki regionalnej, które komplikują ewaluację polityki regionalnej i wymagają przyjęcia nowego sposobu myślenia metodologicznego. Obejmują one: fakt, że innowacje polityki opierają się na nowych teoriach, które są trudniejsze do przetestowania, szczególnie w sytuacji braku odpowiednich danych; obecność niemierzalnych celów, takich jak tworzenie wiedzy, uczenie się i budowanie potencjału; integracja tematów przekrojowych z polityką regionalną, takich jak równość szans czy trwałość środowiska; złożoność stosunku przyczynowo-skutkowego w procesach angażujących dużą liczbę agentów i organizacji, a także mnogość źródeł finansowania; systemowa natura interwencji w ramach polityki regionalnej, ukierunkowanych na różne etapy procesów innowacyjnych lub przedsiębiorczych; trudności z uwzględnieniem kontekstu kulturowego i politycznego oraz warunków ekonomiczno-społecznych, w których

ugruntowane zostały nowe polityki regionalne; dynamika i elastyczność procesów rozwojowych; rola regionów jako „animatorów", angażująca większy udział aktorów regionalnych; oraz rozszerzone skale czasowe, w jakich będą podejmowane interwencje w ramach polityki regionalnej.

Na koniec pojawia się kwestia rozwoju potencjału. Dotyczy to nie tylko instytucjonalizacji ewaluacji w ramach administracji rządowych, na różnych poziomach, ale także profesjonalizacji tej dyscypliny. Wraz z rozwojem towarzystw ewaluacji w wielu państwach europejskich w ciągu ostatnich 5–10 lat pojawił się potencjał prowadzenia bardziej systematycznych szkoleń – szczególnie w ramach studiów uniwersyteckich – oraz rozwoju standardów ewaluacji w skali UE, które mogą przyczynić się do podniesienia jakości ewaluacji polityki publicznej.

BIBLIOGRAFIA

Ahonen P., Virtanen P. and Uuskiylä P. (2002), *Evaluation in Finland*, w: Furubo J.-E. et al. (eds.) *International Atlas of Evaluation*, New Brunswick and London, Transaction Publishers.

Albæk E. and Rieper O. (2002), *Evaluation in Danish Governance*, w: Furubo J.-E. et al. (eds.) *International Atlas of Evaluation*, New Brunswick and London, Transaction Publishers.

Arup (2000), *Evaluation of Regional Selective Assistance, 1991–1995*, Report to the DTI, National Assembly for Wales, Scottish Executive, Arup Economics & Planning, London.

Bachtler J. and Michie R. (1995), A new era in EU regional policy evaluation? The appraisal of the Structural Funds, *Regional Studies* **29**, 745–752.

Bachtler J. and Wren C. (2006), Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges, *Regional Studies* **40(2)**, 143–145.

Barca F. (2006), European Union Evaluation between Myth and Reality: Reflections on the Italian Experience, *Regional Studies* **40(2)**, 273–276.

Beywl W. (2006), The role of evaluation in democracy: Can it be strengthened by evaluation standards: a European perspective, *Journal of MultiDisciplinary Evaluation* **6(3)**, 10–29.

Bœuf J-L. (2001), L'evaluation des politiques publiques, *Problèmes politiques et sociaux* **853**.

Boyle R. (2002), *A Two-Tiered Approach: Evaluation Practice in the Republic of Ireland*, w: J.-E. Furubo et al. (eds.) *International Atlas of Evaluation*, New Brunswick and London, Transaction Publishers.

Casavola P. and Tagle L. (2003), *Building Capacity for Evaluation: Lessons from Italy*, Paper presented at the Fifth European Conference on Evaluation of the Structural Funds, Budapest, 26–27.06.

CSES (2006), *Evaluation Capacity in the New EU Member States*, Paper to the European Evaluation Society and UK Evaluation Society, Joint Conference, London, 4–6.10.

Derlien H.-U. and Rist R.C. (2002), Policy Evaluation in International Comparison, w: J.-E. Furubo et al. (eds.) *International Atlas of Evaluation*, New Brunswick and London, Transaction Publishers.

Deutscher Bundestag (2003), *Sechsunddreißigster Rahmenplan der "Gemeinschaftsaufgabe Verbesserung der regionalen Wirtschaftsstruktur" für den Zeitraum 2007 bis 2010*, Deutscher Bundestag, Drucksache 16/5215, 27.04.

Diez M.-A. (2002) Evaluating New Regional Policies: Reviewing the Theory and Practice, *Evaluation* **8(3)**, 285–305.

European Commission (2000), *Focus on Results: Strengthening Evaluation of Commission Activities*, Communication to the Commission, SEC(2000)1051, Brussels, 26.07.

European Commission (2007), *Responding to Strategic Needs: Reinforcing the use of evaluation*, Communication to the Commission, SEC(2007)213, Brussels, 21.02.

Forss K (2008), *Evaluation in Sweden – the persistence of traditions*, Paper to the French Evaluation Society meeting, Strasbourg.

Furubo J.-E., Rist R.C. and Sandahl R (eds.) (2002), *International Atlas of Evaluation*, New Brunswick and London, Transaction Publishers.

Gavigan J.P. and Scapolo F. (2002a), *Regional Foresight – Future-Proofing and Validating Development Strategies*, ITPS Futures Group.

Gavigan J.P. and Scapolo F. (2002b), *Foresight and the Long-Term View for Regional Development*, IPTS Futures Group.

Hegarty D. (2003), *Framework for the Evaluation of the Structural Funds in Ireland*, Paper to the Fifth European Conference on the Evaluation of the Structural Funds, Budapest, 26–27.06.

Holzinger E. (2001), *Sechs Jahre Programmevaluation – eine Bilanz*, Wien, Österreichisches Institut für Raumplanung (ÖIR).

Huber W. (2006), Evaluation of European Union Cohesion Policy: Window-dressing, Formal Exercise or Coordinated Learning Process? *Regional Studies* **40(2)**, 277–281.

HMT (2003), *The Green Book. Appraisal and Evaluation in Central Government*, Treasury Guidance, Her Majesty's Treasury, London, The Stationery Office.

IAB (2004), *Wie erfolgreich sind Subventionen? Investitionsförderung auf dem Prüfstand (How effective are subsidies? investment assistance under examination)* Gutachten im Rahmen der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur", IAB-Gutachten Nr 1/2004, Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit, Nürnberg.

IBO (2004), *De effecteviteit van regionaal beleid (The effectiveness of regional policy)* Interdepartementaal Beleidsonderzoek (IBO) Regionaal Economisch Beleid, Haga, June.

IOFEZ (2004), *Policy budgets and policy accountability: evaluation – Lessons from practice*, Interministerial Consultations for Financial and Economic Affairs (IOFEZ), Haga, December.

Jakoby H. (2006), Evaluation as Part of the Regional Policy Life Cycle: The Example of North Rhine-Westphalia, Germany, *Regional Studies* **40(2)**, 281–285.

Jarmalavičiūtė N. (2008), *Evaluation arrangements in Lithuania 2007–2013*, Presentation to the DG REGIO, spotkanie Evaluation Network, Bruksela, 26.06.

Karimov A., Borovykh A., Kuzmin A., Abdykadyrova A., Efendiev D., Greshnova E., Konovalova E., Frants I., Palivoda L., Usifili S. and Balakirev V. (2007), Program evaluation development in the Newly Independent States, *Journal of MultiDisciplinary Evaluation* **4(7)**, 84–91.

Kolmodin A. and Melkersson M. (2004), *Den nye regionala utvecklingspolitiken – Hur följa upp och effektutvärdera?* Swedish Institute for Growth Policy Studies, ITPS, Stockholm.

Leuuw F.L. (2002), Evaluation in Europe 2000: Challenges to a Growth Industry, *Evaluation* **8(1)**, 5–12.

Leuuw F.L. (2004), Evaluation in Europe, w: R. Stockmann (ed.) *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*, 2nd edition, Opladen, Leske & Budrich.

Levy R. (1996), Managing Value-for-Money Audit in the European Union: The Challenge of Diversity, *Journal of Common Market Studies* **34(4)**, 510–529.

Ling T. (2003), Ex Ante Evaluation and the Changing Public Audit Function: the Scenario Planning Approach', *Evaluation* **9(4)**, 437–452.

Martin S. and Sanderson I. (1999), Evaluating Public Policy Experiments: Measuring Outcomes, Monitoring Processes or Managing Pilots? *Evaluation* **5(3)**, 245–258.

Martin S. and Thurston R. (2001), *Evaluation Post Devolution: Challenges and Opportunities*, A Discursive Symposium, UK Evaluation Society Annual Conference.

Ministerstwo Finansów (1999), *From Policy Budgets to Policy Accountability* (*Van beleidsbegroting tot beleidsverantwoording*), Ministerie van Financien, Haga, May.

Ministerstwo Finansów (2002), *Management Control Framework: The frame of reference statement on operational management*, Ministerie van Financien, Haga, April.

Ministerstwo Rozwoju Regionalnego (2008), *Evaluation of the cohesion policy in Poland: Summary of previous experience. Plans and challenges for the future*, Krajowa Jednostka Ewaluacji, Departament Koordynacji Polityki Strukturalnej, Warszawa.

Ministerstwo Spraw Gospodarczych (2000), *White Paper on Spatial Economic Policy in the Netherlands: Dynamics in networks*, Ministerie van Economische Zaken, Haga.

Ministerstwo Spraw Gospodarczych (2004), *Peaks in the Delta: Territorial Economic Prospects – Making room for enterprise*, Ministerie van Economische Zaken, Haga.

Ministerstwo Spraw Wewnętrznych (2004), *Government Decision on Development Targets under the Regional Development Act*, Sisäasiainministeriö, Helsinki, 15.01.2004.

Mihalache R. (2006), *Evaluation and knowledge management in the regional and national context*, Paper to the UNDP Evaluation Policy Regional Workshop, Stambul, 6–8.12.

McVittie E. and Swales J.K. (2006), Regional Policy Evaluation: Ignorance, Evidence and Influence, *Strathclyde Discussion Papers in Economics,* No. 03–06, Glasgow, Department of Economics, University of Strathclyde.

Nicol W.R. (1982), Estimating the effects of regional policy: A critique of the European experience, *Regional Studies* **16(3)**, 199–210.

O'Hara J., McNamara G., Boyle R. and Sullivan C. (2007), Contexts and constraints: An analysis of the evolution of evaluation in Ireland with particular reference to the education system, *Journal of MultiDisciplinary Evaluation* **4(7)**, 75–83.

OECD (2001), *Results-based management in development cooperation agencies: a review of experience*, DAC Working Party on Aid Evaluation, Paris.

OECD (2005), *OECD Territorial Development Reviews: Finland*, Paris.

OECD (2007a), *Globalisation and Regional Economies: Can OECD Regions Compete in Global Industries?*, Paris.

OECD (2007b), *Performance Budgeting in OECD Countries*, Paris.

Olejniczak K. (2002), *Towards the evaluation culture in Poland – experiences and prospects*, Annual Conference of Regional Studies Association – *Evaluation and EU Regional Policy: New Questions and New Challenges*, Regional Studies Association, Aix-en-Provence, June.

Pazos M. and Zapico-Gõni R. (2002), *Program evaluation in Spain: Taking off at the edge of the Twenty-First Century?*, w: Furubo J.-E. et al. (eds.), *International Atlas of Evaluation*, New Brunswick and London, Transaction Publishers.

Polverari L. and Bachtler J. (2004), *Assessing the Evidence: The Evaluation of Regional Policy in Europe*, EoRPA Paper 04/5, Glasgow, European Regional Policy Research Consortium, European Policies Research Centre, University of Strathclyde.

Polverari L., Mendez C., Gross F. and Bachtler J. (2004), Making Sense Of European Cohesion Policy 2007–13: On-Going Evaluation And Monitoring Arrangements, *IQ-Net Thematic Paper* **21(2)**, Glasgow, European Policies Research Centre, University of Strathclyde.

Raines P. and Taylor S. (2002), Mid-term Evaluation of the 2000–06 Structural Funds Programmes, *IQ-Net Thematic Paper* **11(2)**, Glasgow, European Policies Research Centre, University of Strathclyde.

Regeringskansliet (2000), *Regionalpoltisika utredningens slutbetänkande,* Statens offentliga utredningar, SOU 2000:87, Stockholm, Näringsdepartementet.

Salamin G. and Peti M. (2002), *First steps towards systematic regional programme evaluation: experiences of an accession country, Hungary*, Paper to the RSA Conference *Evaluation and EU Regional Policy: New Questions and New Challenges*, Aix-en-Provence, 31.05–10.6.

Scheffel K. (2001), Evaluierung – Chance oder Last? Eine Betrachtung aus Bundessicht, *Informationen zur Raumentwicklung* **6**/**7**, 315–319.

Schröter D. (2004) Evaluation in Europe: An Overview, *Journal of MultiDisciplinary Evaluation* **1**, 58–62.

Stame N. (1998), Evaluation in Italy: Experience and Prospects, *Evaluation* **4(1)**, 91–103.

Stern E. (2003), *The Rationale & Challenges for Evaluation of Public Policies*, Keynote Speech to the Fifth European Conference on Evaluation of the Structural Funds, Budapest, 26–27.06.

Strohmeier G. and Holzinger E. (2004), *KAP-EVA Koordinations- und Arbeitsplatform Evaluierung: Bilanzbericht*, Österreiches Raumordnungskonferenz, Wien.

Strohmeier G. and Holzinger E. (2006), *Die Koordinations- und Arbeitsplatform KAP-EVA: Ein Lernprozess in der Evaluierung der Strukturfonds in Österreich*, Österreiches Raumordnungskonferenz, Wien.

Struhkamp G. (2005), Evaluation in Germany: An Overview, *Journal of Multi-Disciplinary Evaluation* **2(3)**, 180–194.

Toulemonde K. and Bjornkilde T. (2003), *Building evaluation capacity: experience and lessons in Member States and Acceding Countries*, Paper to the Fifth European Conference on Evaluation of the Structural Funds, Budapest, 26–27.06.

Tutty M. (1994), *Improving Evaluation of Policies and Programmes in the Irish Public Service,* University of Dublin, Trinity College, unpublished Msc thesis (quoted in Boyle R. (1997), *Evaluating Public Expenditure Programmes: Determining a Role For Programme Review*, Dublin, Institute of Public Administration.

UNDP (2007), *Workshop Report*, UNDO Regional Workshop on the UNDP Evaluation Policy: European and Commonwealth of Independent States, Stambul, 6–8.12.

Van Der Knaap P. (2006), Responsive Evaluation and Performance Management, *Evaluation* **12(3)**, 278–293.

Varone F., Jacob S. and De Winter L. (2005), Polity, Politics and Policy Evaluation in Belgium, *Evaluation* **11(3)**, 253–273.

Wadley D. (1986), *Restructuring the Regions: Analysis, Policy Model and Prognosis*, Paris, Organisation for Economic Cooperation and Development,

Williams K., de Laat B. and Stern E. (2002), *The Use of Evaluation in the Commission Services*, Final Report to the European Commission, Paris, Technopolis France.

ROZDZIAŁ 1.4.

Piotr Żuber
Stanisław Bienias

SYSTEM EWALUACJI W POLSCE – DOTYCHCZASOWE DOŚWIADCZENIA I WYZWANIA NA PRZYSZŁOŚĆ[1]

1. WPROWADZENIE

Celem niniejszego rozdziału jest zapoznanie Czytelnika z dotychczasową praktyką systemowego wykorzystania narzędzia, jakim jest ewaluacja w funkcjonowaniu administracji publicznej. Na kolejnych stronach omówiono kształt systemu ewaluacji w Polsce – wraz z rysem historycznym, który wpłynął na jego obecną postać. Następnie zaprezentowany został zakres badań zrealizowanych w latach 2004–2007 oraz dalsza perspektywa wykorzystania ewaluacji w okresie 2008–2015. Rozdział kończy się zdefiniowaniem kluczowych wyzwań stojących przed procesem ewaluacji w naszym kraju.

Poniższy tekst koncentruje się w głównej mierze na systemowym wykorzystaniu ewaluacji w obszarze realizacji polityki spójności. Należy

[1] W niniejszym artykule wykorzystano tezy raportu: *Proces ewaluacji polityki spójności w Polsce. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość* opracowanego przez Krajową Jednostkę Oceny w Departamencie Koordynacji Polityki Strukturalnej MRR, Warszawa, kwiecień 2008.

jednak pamiętać, że ewaluacja była już wykorzystywana w ramach pomocy przedakcesyjnej. Ewaluacje odnoszące się do konkretnych projektów i programów były wykonywane na zamówienie Komisji Europejskiej (np. ewaluacja programu PHARE CBC – Współpracy przygranicznej); można także wymienić kilka pozytywnych przykładów instytucji, w których systematyczne oceny skuteczności i – w mniejszym stopniu – efektywności podejmowanych działań były dokonywane np. przez Polską Agencję Rozwoju Regionalnego, Polską Agencję Rozwoju Przedsiębiorczości oraz Urząd Komitetu Integracji Europejskiej. W tym czasie powstało również wiele ciekawych opracowań, które zostały następnie wykorzystane przy budowie potencjału dla ewaluacji polityki spójności[2]. Niemniej jednak, te pojedyncze działania nie były ze sobą skorelowane i nie wpływały zasadniczo na poprawę funkcjonowania polityk krajowych, w tym postrzegania ewaluacji jako narzędzia poprawy sposobu zarządzania w sferze publicznej. Doświadczenia z okresu przedakcesyjnego miały ponadto bardzo ograniczony wpływ na tworzenie się obecnie funkcjonującego systemu ewaluacji. Jedyną instytucją, która zachowała ciągłość funkcjonowania i w sposób płynny przeszła od ewaluacji interwencji przedakcesyjnych do ewaluacji polityki spójności, była PARP. Instytucja ta, z racji swojego usytuowania w systemie wdrażania polityki spójności (instytucja wdrażająca), miała jednak ograniczony wpływ na kształtowanie się systemu ewaluacji, jaki w chwili obecnej funkcjonuje w naszym kraju. Z powyższych względów podjęto decyzję o skoncentrowaniu problematyki niniejszego artykułu w obszarze ewaluacji polityki spójności od momentu przystąpienia Polski do struktur europejskich, czyli od roku 2004. Nie można jednak nie doceniać działań, jakie zostały podjęte w okresie wcześniejszym[3].

2. EWALUACJA JAKO NARZĘDZIE POPRAWY SKUTECZNOŚCI I EFEKTYWNOŚCI FUNKCJONOWANIA POLITYKI SPÓJNOŚCI W POLSCE

Na tle innych państw europejskich, w których ewaluacja od wielu lat funkcjonuje jako nieodzowny element procesu zarządzania w sferze publicznej, polskie doświadczenia w tym zakresie są niewielkie. Przed przystąpieniem Polski do Unii Europejskiej nie wprowadzano bowiem na szerszą skalę do polskiej administracji zasad związanych z zarządzaniem ukierunkowanym na rezultaty oraz nie stosowano ewaluacji jako instrumentu oceny efektywności i skuteczności wydatków publicznych.

[2] M.in. znakomite opracowanie Kierzkowski 2002.

[3] Więcej na temat doświadczeń Polski w ewaluacji projektów i programów przedakcesyjnych zob. Olejniczak 2002.

Wynikało to zarówno z dziedzictwa systemu społeczno-gospodarczego, w którym znaczenie słów „efektywność" i „skuteczność" było nieznane, jak i z generalnego braku na poziomie politycznym wizji modelu funkcjonowania nowoczesnego państwa, a co się z tym wiąże – niedocenianiem zagadnień związanych z organizacją systemu zarządzania w sferze publicznej.

Sytuacja ta uległa jednak od 2004 roku pozytywnym zmianom. Członkostwo w Unii Europejskiej i udział Polski w polityce spójności wymusiły stosowanie zasad UE w zakresie ewaluacji wynikających z rozporządzeń dotyczących funduszy strukturalnych oraz przygotowanie odpowiednich przepisów krajowych (m.in. ustawa z dnia 20 kwietnia 2004 roku o Narodowym Planie Rozwoju[4]). Realizacja Narodowego Planu Rozwoju/Podstaw Wsparcia Wspólnoty 2004–2006 stały się punktem zwrotnym w historii ewaluacji Polsce – rozpoczęła się systematyczna, skoordynowana i oparta na jednolitym systemie wskaźników ocena efektywności stosowanych instrumentów w ramach polityki spójności, ocena skuteczności działań administracji w tym zakresie, a także całościowa ocena oddziaływania funduszy strukturalnych na sytuację społeczno-gospodarczą.

W celu zapewnienia funkcjonowania systemu ewaluacji w Polsce w obszarze polityki spójności od kilku lat trwa proces budowy odpowiednich struktur administracyjnych, rozwijanie potencjału administracyjnego i, równolegle, proces zwiększania podaży do wykonywania prac ewaluacyjnych ze strony środowiska naukowego (szkoły wyższe, instytuty naukowo-badawcze) i profesjonalnego (firmy badawcze, konsultingowe, doradcze). Początkowo procesy te koncentrowały się, czy też były inicjowane, w administracji rządowej szczebla centralnego (zwłaszcza w Ministerstwie Rozwoju Regionalnego, zapoczątkowane powołaniem Krajowej Jednostki Oceny[5]); obecnie wzrasta znaczenie budowy potencjału na poziomie regionalnym.

W ciągu ostatnich kilku lat (do końca 2007 roku) można w tych pracach wyróżnić trzy wyraźne etapy, a mianowicie:

- **etap przygotowawczy** (obejmujący lata 2004–2005), który związany był z budową od podstaw potencjału ewaluacyjnego w administracji centralnej oraz z uruchomieniem pierwszych ewaluacji;
- drugi etap (2006 rok), związany z **ewaluacją** *ex ante* **dokumentów programowych na okres 2007–2013** – ugruntowujący i rozwijający wypracowane wcześniej rozwiązania organizacyjne (w zakresie

[4] DzU, Nr 116, poz. 1206.

[5] Krajowa Jednostka Oceny (KJO) została powołana zgodnie z art. 58 ust.1 ustawy z dnia 20 kwietnia 2004 roku o Narodowym Planie Rozwoju w maju 2004 roku jako Zespół ds. Oceny w strukturze Departamentu Koordynacji Polityki Strukturalnej byłego Ministerstwa Gospodarki i Pracy. Obecnie obsługą merytoryczną i organizacyjną prac KJO zajmuje się Wydział Ewaluacji w Departamencie Koordynacji Polityki Strukturalnej Ministerstwa Rozwoju Regionalnego (szerzej o zadaniach KJO w *Planie ewaluacji NSRO na lata 2007–2013*, dostępnym na www.ewaluacja.gov.pl).

zlecania ewaluacji, współpracy z wykonawcami i upowszechniania wyników ewaluacji);

- trzeci etap, obecnie wdrażany, związany z uruchomieniem **zintegrowanego systemu ewaluacji bieżącej**, obejmującej interwencje zarówno okresu 2004–2006 (w ramach NPR), jak i lata 2007–2013 (NSRO).

Podstawowe zasady, na których opiera się obecnie funkcjonujący system ewaluacji, to przede wszystkim:

- zasada integracji dwóch równoległych procesów ewaluacji toczących się w ramach okresu programowania 2004–2006 oraz w ramach okresu programowania 2007–2013 – ewaluacja całej polityki spójności realizowanej w naszym kraju, a nie wyłącznie poszczególnych dokumentów programowych;
- zasada decentralizacji systemu ewaluacji (w ślad za decentralizacją systemu wdrażania funduszy strukturalnych) – realizowana przez powierzanie kompetencji w zakresie ewaluacji na niższe szczeble wdrażania, tak by ewaluacje mogły być realizowane przez instytucje znajdujące się najbliżej badanej problematyki;
- zasada partnerstwa w procesie ewaluacji – realizowana przez zaangażowanie w proces ewaluacji (m.in. na forum grup sterujących) przyszłych interesariuszy wyników badania[6];
- zasada planowania działań ewaluacyjnych – realizowana przez obowiązek opracowania planów okresowych i strategicznych.

Na podstawie wymienionych zasad, w 2007 roku uruchomiony został zintegrowany i zdecentralizowany system ewaluacji interwencji współfinansowanych z funduszy strukturalnych UE, bazujący na strukturze organizacyjno-instytucjonalnej, gwarantującej realizację zasady partnerstwa w możliwie najszerszym zakresie. Głównym celem takiej organizacji systemu jest zapewnienie najwyższego stopnia użyteczności uzyskiwanych wyników badań ewaluacyjnych przez przyznanie kompetencji w zakresie zlecania badań ewaluacyjnych instytucjom najbardziej zainteresowanym wnioskami i rekomendacjami z ewaluacji na właściwym poziomie wdrażania oraz przez zaangażowanie w proces konsultacji szerokiego grona uczestników systemu wdrażania, w celu trafnego definiowania obszarów badań ewaluacyjnych, zgodnych z rzeczywistymi potrzebami informacyjnymi.

Podstawowymi dokumentami przygotowanymi w roku 2007, określającymi ramy systemu są: *Wytyczne w zakresie ewaluacji programów operacyjnych na lata 2007–2013*, *Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, *System ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013 oraz Narodowego Planu Rozwoju na lata 2004–2006*.

---

[6] Osób zajmujących się nie tylko ewaluacją, lecz również monitoringiem, kontrolą i programowaniem, czyli mogących najwięcej wnieść do całego procesu badawczego.

3. SYSTEM INSTYTUCJONALNY EWALUACJI POLSCE W ROKU 2008

Funkcję koordynacyjną w systemie pełni Krajowa Jednostka Oceny, która oprócz koordynacji działań odpowiada również za realizację procesu ewaluacji na poziomie NPR i NSRO. Strategiczne obszary tematyczne (MRR, 2007), w ramach których inicjowane są badania ewaluacyjne, to:

- rozwój społeczno-ekonomiczny Polski,
- rozwój i modernizacja infrastruktury,
- konkurencyjność i innowacyjność gospodarki,
- rozwój zasobów ludzkich,
- rozwój regionalny i terytorialny,
- budowa potencjału administracji publicznej i realizacja zasady *good governance*.

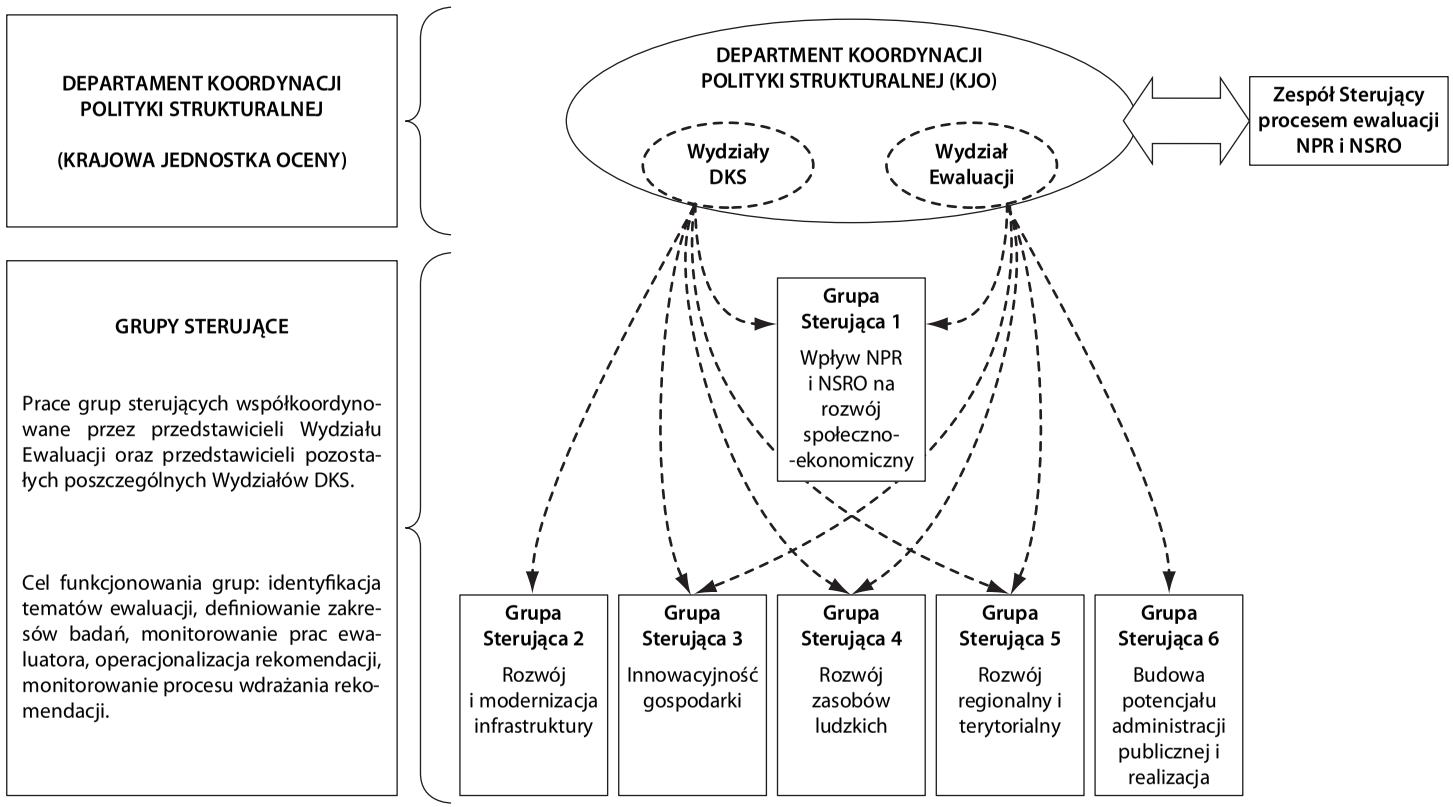
Obszary te wynikają bezpośrednio z celu głównego i celów szczegółowych NSRO, tak by ewaluacja w maksymalnym stopniu była podporządkowana realizacji strategii.

System ewaluacji na poziomie NPR i NSRO jest w sferze instytucjonalnej oparty na sześciu tematycznych grupach sterujących, odpowiadających wymienionym obszarom tematycznym (rysunek 1). W skład grup wchodzą przedstawiciele Wydziału Ewaluacji, poszczególnych wydziałów merytorycznych Departamentu Koordynacji Polityki Strukturalnej w MRR, Instytucji Zarządzających oraz innych instytucji zaangażowanych w zarządzanie i wdrażanie interwencji będących przedmiotem badań, a także eksperci zewnętrzni.

Prace grup sterujących zainicjowane zostały w 2. połowie 2007 roku. Pierwszym istotnym efektem prac było zdefiniowanie priorytetowych tematów badań ewaluacyjnych na poziomie NPR i NSRO, które zawarte zostały w projekcie *Okresowego planu ewaluacji NPR i NSRO na rok 2008*. Ponadto przedmiotem prac tych grup były badania inicjowane przez KJO. Dodatkowo, w celach koordynacyjnych, powołany został Zespół Sterujący procesem ewaluacji NPR i NSRO, który jest ciałem funkcjonującym na poziomie dyrektorów Instytucji Zarządzających. Jego rolą jest zapewnienie spójności procesu ewaluacji NPR i NSRO oraz wyznaczanie ogólnych, strategicznych kierunków procesu ewaluacji. Ponadto Zespół pozwala na silniejsze włączenie w cały proces ewaluacji tzw. *decision makers,* którzy w największym stopniu są w stanie wykorzystać wnioski i rekomendacje płynące z badań ewaluacyjnych.

Kluczowymi podmiotami w systemie ewaluacji – obok Krajowej Jednostki – są jednostki ewaluacyjne w ramach poszczególnych programów operacyjnych, odpowiedzialne za organizację procesu ewaluacji na swoim poziomie wdrażania. W ślad za decentralizacją systemu wdrażania, proces ewaluacji w okresie 2007–2013 jest realizowany również na poziomie regionalnym – w każdym województwie powołana została

Rysunek 1. Sposób organizacji grup sterujących dla systemu ewaluacji NPR 2004–2006 oraz NSRO 2007–2013



Jednostka Ewaluacyjna, odpowiedzialna za ewaluację regionalnego programu operacyjnego.

Istotnym elementem systemu są Plany ewaluacji. Zgodnie z *Wytycznymi w zakresie ewaluacji programów operacyjnych* Plany ewaluacji sporządzane są na poziomie NPR/NSRO oraz na poziomie programów operacyjnych (jak również na niższych poziomach wdrażania w przypadku delegowania kompetencji). Plany mają charakter strategiczny (*Plany ewaluacji na okres 2007–2013* dotyczące całego okresu programowania wyznaczają ogólne obszary ewaluacji oraz sposób organizacji systemu na danym poziomie wdrażania) oraz operacyjny (*Okresowe plany ewaluacji* wskazujące działania planowane do realizacji w danym roku).

Plan ewaluacji NPR i NSRO, *Plany ewaluacji programów operacyjnych* oraz *Plan ewaluacji kwestii horyzontalnych w ramach RPO* sporządzone zostały do końca 2007 roku. W styczniu 2008 roku zakończyły się prace nad projektami *Okresowych planów ewaluacji na rok 2008*.

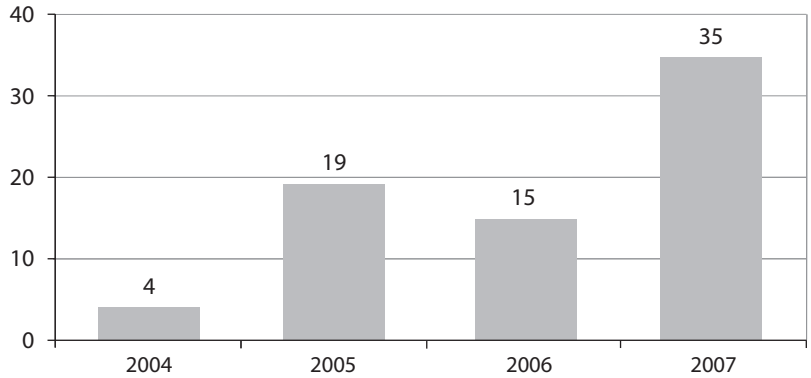
Realizacja ww. planów ewaluacji będzie podlegać monitorowaniu przez KJO. Instrumentem monitorowania planów ewaluacji jest *Informacja roczna w zakresie ewaluacji* przekazywana do KJO przez IZ do końca marca każdego roku.

Ukształtowany w ten sposób system stwarza szansę na prawdziwie efektywne wykorzystanie narzędzia, jakim jest ewaluacja *on-going*, czyli dostarczenie odpowiednich informacji, we właściwym czasie osobom, które mogą je wykorzystać w procesie zarządzania polityką spójności. Wydaje się, że tak działający system ewaluacji umożliwi zidentyfikowanie kwestii problemowych, które pojawią się w przyszłości, pozwalając jednocześnie na udzielenie odpowiedzi bazującej na doświadczeniach, które zostały zebrane w przeszłości[7].

4. BADANIA EWALUACYJNE ZREALIZOWANE W OKRESIE 2004–2007

W okresie 2004–2007 zrealizowanych zostało **135 badań ewaluacyjnych** (122 badania w ramach scharakteryzowanych poniżej obszarów tematycznych oraz 13 strategicznych ocen oddziaływania na środowisko realizowanych w ramach procesu ewaluacji *ex ante*), z czego prawie połowę (62) stanowią badania o charakterze *ex ante*. Badania realizowane były przede wszystkim w ramach poszczególnych programów operacyjnych (ok. 80% wszystkich badań) oraz na poziomie NPR i NSRO (ok. 20% badań). Na rysunku 2 przedstawiono rozkład czasowy zrealizowanych

[7] Realizacja badania trwa na ogół kilka miesięcy, w celu otrzymania wyników w momencie, gdy będą one najbardziej użyteczne – konieczne jest z wyprzedzeniem definiowanie tematów badawczych (na forum grup sterujących – w ramach okresowych planów ewaluacji). Jednocześnie przy udzieleniu odpowiedzi na „wirtualne" jeszcze kwestie problemowe musimy bazować na doświadczeniach, jakie zdobyliśmy w przeszłości – stąd idea połączenia dwóch okresów programowania 2004–2006 i 2007–2013.

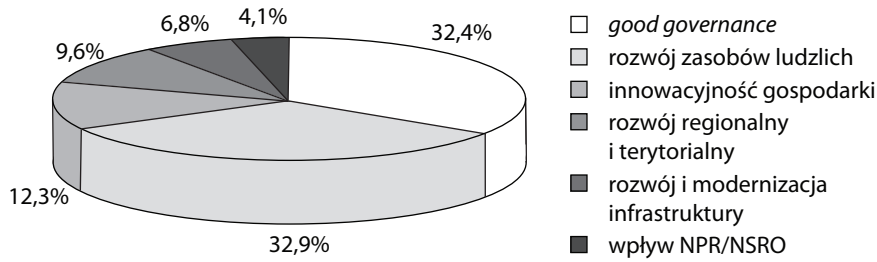


Rysunek 2. Liczba badań ewaluacyjnych typu *on-going* zrealizowanych w kolejnych latach

Źródło: Krajowa Jednostka Oceny, Departament Koordynacji Polityki Strukturalnej MRR (2008), *Proces ewaluacji polityki spójności w Polsce. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość*, Warszawa.

badań *on-going*. Widzimy wyraźnie, że liczba realizowanych badań *on-going*[8] systematycznie rośnie (przejściowe zmniejszenie liczby zakończonych badań w roku 2006 ma charakter pozorny i wynika z silniejszego zaangażowania administracji w badania *ex ante* dokumentów programowych na lata 2007–2013).

Proces ewaluacji polityki spójności w latach 2004–2007 związany był ze strategicznymi obszarami tematycznymi, wynikającymi bezpośrednio z celów NSRO. Strukturę tematyczną dotychczas przeprowadzonych badań *on-going* przedstawia rysunek 3.



Rysunek 3. Zakończone badania *on-going* według obszarów tematycznych

Źródło: Krajowa Jednostka Oceny, Departament Koordynacji Polityki Strukturalnej MRR (2008), *Proces ewaluacji polityki spójności w Polsce. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość*, Warszawa.

[8] Na rysunku 2 przedstawiono strukturę zrealizowanych badań ewaluacyjnych *on-going*, gdyż z uwagi na brak formalnego obowiązku realizacji tego typu ewaluacji, wydają się bardziej właściwym (niż bezwzględnie wymagane badania typu *ex ante*) wskaźnikiem roli, jaką odgrywa ewaluacja w procesie realizacji polityki spójności.

W ramach obszaru ***Wpływ NPR i NSRO na rozwój społeczno-ekonomiczny*** zrealizowanych zostało 9 badań ewaluacyjnych. W większości dotyczyły one oddziaływania makroekonomicznego oraz koncentrowały się na wpływie polityki spójności na kształtowanie się głównych wskaźników na poziomie NPR i NSRO oraz innych wybranych wskaźników makroekonomicznych. Celem badań ewaluacyjnych realizowanych w tym obszarze była ocena efektów polityki spójności traktowanej całościowo. Badania zrealizowane w tym obszarze wskazują, że polityka spójności przynosi widoczne efekty w sferze gospodarczej i społecznej. Duża rola procesu ewaluacji w niniejszym obszarze przewidywana jest w kontekście dyskusji nad przyszłym kształtem polityki spójności.

W okresie 2004–2007 zrealizowano 6 badań w ramach obszaru ***Rozwój i modernizacja infrastruktury***. Jest to relatywnie słabiej przebadany obszar w ramach całego procesu ewaluacji. Wynika to przede wszystkim ze specyfiki inwestycji infrastrukturalnych generujących głównie efekty długoterminowe oraz z braku wystarczających narzędzi oceny. Badania w niniejszym obszarze miały głównie charakter diagnostyczny (obejmowały analizę struktury i spójności realizowanych inwestycji) oraz prognostyczny (podejmowały próbę oceny przewidywanego wpływu inwestycji). Głównymi wyzwaniami w ramach niniejszego obszaru jest konieczność ukierunkowania procesu na badania w zakresie systemu wdrażania inwestycji infrastrukturalnych oraz wypracowanie odpowiednich narzędzi pomiaru oceny efektów interwencji, zarówno na poziomie programów operacyjnych i osi priorytetowych, jak i na poziomie poszczególnych projektów.

W ramach obszaru ***Konkurencyjność i innowacyjność gospodarki*** zrealizowanych zostało 11 badań ewaluacyjnych. Badania dotyczyły efektów interwencji przyczyniających się do poprawy konkurencyjności gospodarki w sposób bezpośredni (wsparcie bezpośrednie dla przedsiębiorstw) oraz pośredni (tworzenia korzystnych warunków rozwoju innowacyjności i konkurencyjności). Analiza badań w niniejszym obszarze wskazuje na konieczność ukierunkowania procesu na badania o charakterze kompleksowym i horyzontalnym oraz konieczność analizy przełożenia wsparcia dla instytucji otoczenia biznesu na wzrost konkurencyjności i innowacyjności przedsiębiorstw. Coraz bardziej widoczna jest potrzeba pogłębionej oceny efektywności interwencji w niniejszym obszarze, zwłaszcza w kontekście możliwości występowania efektu *deadweight*.

Obszar tematyczny ***Rozwój zasobów ludzkich*** jest najlepiej przebadany w ramach całego procesu ewaluacji polityki spójności w Polsce. W obszarze zrealizowanych zostało 25 badań ewaluacyjnych. Badania dotyczyły oceny efektów dla beneficjentów ostatecznych interwencji, oceny poszczególnych działań w ramach programów operacyjnych oraz oceny instytucji realizujących projekty wraz z identyfikacją barier we wdrażaniu. Ponadto zrealizowane zostały badania o charakterze

horyzontalnym, w tym badanie analizujące wpływ interwencji na zatrudnienie. Wysoka jakość procesu ewaluacji w niniejszym obszarze wynika przede wszystkim z dużego potencjału ewaluacyjnego instytucji zaangażowanych (zwłaszcza w ramach Instytucji Zarządzającej SPO RZL), a także ze specyfiki obszaru (występowanie efektów w krótkim okresie). Wyniki badań wskazują na istotny wpływ funduszy strukturalnych na rozwój zasobów ludzkich, w tym na zwiększenie zatrudnienia.

W ramach obszaru ***Rozwój regionalny i terytorialny*** zrealizowane zostały 42 badania, z czego większość (35) miała charakter ewaluacji *ex ante* (w tym ewaluacja *ex ante* 16 Regionalnych Programów Operacyjnych). Najistotniejszymi badaniami w obszarze były ewaluacje o charakterze horyzontalnym, zwłaszcza ocena wpływu funduszy strukturalnych na rozwój społeczno-ekonomiczny na poziomie podregionów. Analiza obszaru wskazuje na potrzebę zintensyfikowania prac nad ewaluacjami bieżącymi, odpowiadającymi specyfice obszaru oraz ukierunkowania badań na kwestie związane ze spójnością terytorialną i rozwojem przestrzennym.

W obszarze ***Budowa potencjału administracji publicznej i realizacja zasady „good governance"*** zrealizowanych zostało 29 badań ewaluacyjnych. Duża liczba badań ukierunkowana była na identyfikację barier systemowych oraz barier po stronie beneficjentów. Analiza obszaru wskazuje m.in. na celowość intensyfikacji prac nad realizacją badań dotyczących kwestii systemowych związanych z uruchamianiem programów operacyjnych na lata 2007–2013 oraz badań odnoszących się do efektów realizacji polityki spójności w Polsce dla administracji jako całości (również tej niezwiązanej bezpośrednio z wdrażaniem polityki spójności). Dużym wyzwaniem jest również konieczność rozwoju metodyki badania efektów realizacji zasady *good governance*. Wyniki ewaluacji w niniejszym obszarze pozwoliły na identyfikację i usunięcie szeregu barier oraz problemów we wdrażaniu funduszy strukturalnych.

Istotnym aspektem procesu ewaluacji, oprócz realizacji badań ewaluacyjnych, są działania związane z budową **potencjału ewaluacyjnego**. Analiza potencjału pozwala na identyfikację mocnych i słabych stron systemu, w ramach poszczególnych obszarów tematycznych. Największym potencjałem ewaluacyjnym w całym systemie dysponują instytucje zaangażowane w proces ewaluacji w obszarze rozwój zasobów ludzkich, w tym zwłaszcza Instytucja Zarządzająca SPO RZL oraz jednostka ewaluacyjna Polskiej Agencji Rozwoju Przedsiębiorczości. Duży potencjał w obszarze wynika z kompetencji zespołów w instytucjach oraz świadomości dyrekcji w zakresie znaczenia działań ewaluacyjnych. Dużym potencjałem dysponują również instytucje w ramach obszarów *Wpływ NPR i NSRO na rozwój społeczno-ekonomiczny* oraz *Konkurencyjność i innowacyjność gospodarki*. Istotnym wyzwaniem dla całego systemu ewaluacji w Polsce jest budowa potencjału ewaluacyjnego na poziomie regionalnym oraz wzmocnienie kadrowe instytucji zaangażowanych

w ewaluację największego programu operacyjnego *Infrastruktura i Środowisko*.

Analiza scharakteryzowanego powyżej procesu ewaluacji w okresie 2004–2007 pozwala na sformułowanie następujących ogólnych wniosków:

- **Szybko rośnie liczba i wielkość realizowanych badań ewaluacyjnych dotyczących polityki spójności.** Biorąc pod uwagę tylko liczbę realizowanych badań *on-going*, wzrosła ona z 4 w 2004 roku do 35 w 2007 roku. Świadczy to niewątpliwie o rosnącej roli tego narzędzia w całym procesie wdrażania programów współfinansowanych z środków UE. Pozytywnym zjawiskiem jest również rosnąca wielkość i wartość realizowanych badań (po początkowych trudnościach ze zlecaniem zamówień eksperckich, powyżej progu wymagającego stosowania Ustawy Prawo Zamówień Publicznych, jednostki ewaluacyjne nauczyły się przeprowadzać takie postępowania). Dodatkowym pozytywnym efektem jest włączenie na stałe kryteriów jakościowych do oceny ofert (na ogół stanowią one około 50% punktów, jakie może uzyskać potencjalny wykonawca)[9].
- **Systematycznie rośnie jakość opracowywanych badań.** W opinii osób bezpośrednio związanych ze zlecaniem i odbieraniem ewaluacji[10], wyraźnie widoczny jest pozytywny trend w ocenie jakości badań. Jakość raportów systematycznie wzrasta, a coraz więcej raportów uzyskuje ocenę dobrą lub bardzo dobrą. W roku 2007 ponad 75% raportów uzyskało taką właśnie ocenę, podczas gdy w roku 2006 niespełna połowa badań była oceniona aż tak wysoko. Fakt, że w ślad za rosnącą liczbą realizowanych badań, rośnie również ich jakość, jest bardzo dobrym prognostykiem. Należy jednak mieć na uwadze, że w 2008 roku w zlecanie i odbieranie badań zaangażowane zostanie wiele nowych osób, niemających doświadczenia w tym zakresie, co będzie stanowiło niewątpliwe wyzwanie dla stałego podnoszenia jakości realizowanych badań.
- **W latach 2004–2007 istniała ograniczona liczba podmiotów zlecających ewaluację.** Przeszło połowa wszystkich ewaluacji *on-going* w zakresie polityki spójności została zlecona z inicjatywy tylko dwóch podmiotów – Krajowej Jednostki Oceny i Jednostki

---

[9] Włączenie kryteriów jakościowych stanowi pewnego rodzaju nową jakość funkcjonowania administracji publicznej.

[10] Ocena została dokonana na podstawie ankiety przeprowadzonej wśród jednostek ewaluacyjnych na temat jakości zlecanych badań. Jednostki ewaluacyjne oceniały jakość raportów końcowych ze zlecanych badań w skali od 1 do 5 (gdzie: 1 – bardzo zła, 2 – mierna, 3 – dostateczna, 4 – dobra, 5 – bardzo dobra) oraz jakość współpracy z wykonawcami w skali od 1 do 4 (gdzie: 1 – bardzo zła, 2 – zła, 3 – dobra, 4 – bardzo dobra). Więcej na ten temat zob. KJO (2008), *Proces ewaluacji polityki spójności w Polsce. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość*.

Ewaluacyjnej programu SPO RZL. Taka koncentracja badań była związana głównie z wysoką świadomością co do możliwości wykorzystania narzędzia, jakim jest ewaluacja, przez kierownictwo departamentów, w których jednostki te są zlokalizowane. Liczba podmiotów zlecających ewaluacje systematycznie rośnie. Aktywnymi aktorami były również Jednostki Ewaluacyjne SPO WKP, a od roku 2006 Jednostka Ewaluacyjna ZPORR. Również w tych przypadkach kluczową rolę odegrało pozytywne nastawienie osób decyzyjnych w tych departamentach. Nasuwa się wniosek, że swoistą *target group* dla wszystkich działań ewaluacyjnych (w zakresie promocji narzędzia, adresowania rekomendacji i włączania w cały proces ewaluacji) jest właśnie kadra kierownicza na poziomie dyrektorów[11]. Przy opracowywaniu *Systemu ewaluacji NPR i NSRO* kwestia ta została uwzględniona.

- **Systematycznie rosła podaż usług ewaluacyjnych, a współpraca z wykonawcami badań była coraz wyżej oceniana**. Obserwowalne jest zjawisko rosnącego rynku podmiotów świadczących usługi w zakresie ewaluacji. Zdecydowanie przeważają tu firmy konsultingowe. Głębsza analiza pokazuje jednak, że w ramach różnych zespołów badawczych w wielu przypadkach pojawiają się ci sami eksperci. Świadczy to o mniejszym – niż wynika to tylko z analizy nazw firm – rozwoju rynku ewaluacyjnego w Polsce. Z przykrością należy zauważyć natomiast małą aktywność uczelni wyższych i instytutów badawczych na tym polu. Dalszego poszerzenia liczby podmiotów realizujących badania należy się spodziewać wraz z uruchomieniem realizacji ewaluacji na szczeblu regionalnym (będzie tu więcej małych badań, gdzie skłonność do podjęcia ryzyka i powierzenia realizacji ewaluacji mniej doświadczonym wykonawcom jest wyższa). Jednocześnie zauważalny jest wzrost profesjonalizmu podmiotów rynku ewaluacyjnego. Świadczy o tym nie tylko wzrost jakości raportów z badań, wysoko jest również oceniana współpraca z ewaluatorami zewnętrznymi. W blisko 85% przypadków współpraca ta była oceniana dobrze i bardzo dobrze.

Brak sprawnie działającego systemu monitoringu rzeczowego stanowi istotną przeszkodę dla działań ewaluacyjnych realizowanych w okresie 2004–2007. Brak dobrze funkcjonującego systemu SIMIK 2004–2006 powoduje zwiększenie kosztów realizowanych badań, wydłużenie czasu ich przygotowania (konieczność pozyskania przez wynajętych ekspertów w ramach każdego badania danych, które powinny być agregowane przez IZ). Dla procesu ewaluacji barierę stanowił również brak jednego, spójnego systemu wskaźników, który byłby

[11] Dodatkowym argumentem w tym względzie jest fakt, że dyrektorzy, w odróżnieniu od kierownictwa poziomu politycznego, najczęściej nie zmieniają się po każdych wyborach, przez co nagromadzona wiedza nie jest tracona.

wykorzystywany do monitorowania efektów realizacji polityki spójności jako całości, czyli tzw. wskaźników kluczowych. Brak właściwie funkcjonującego systemu monitoringu powoduje również utrudnienia związane z zastosowaniem bardziej wiarygodnych metod badawczych (np. problemy z doborem losowym reprezentatywnej próby badawczej). Słabości te, które niewątpliwie przyczyniały się do obniżenia jakości realizowanych badań[12], mają zostać wyeliminowane wraz z uruchomieniem systemu SIMIK 2007–2013.

5. EWALUACJA W LATACH 2008–2015

Na realizację działań ewaluacyjnych w okresie 2007–2015 w całym systemie zaplanowanych zostało **ok. 190 mln PLN** (w tym ok. 73 mln PLN na poziomie regionalnym oraz ok. 117 mln PLN na poziomie centralnym).

Tylko na rok 2008 zaplanowano realizację **ponad 100 badań ewaluacyjnych** (w porównaniu do 135 przeprowadzonych łącznie do roku 2007) na wszystkich poziomach wdrażania. Najwięcej ewaluacji dotyczyć będzie kwestii związanych z systemem zarządzania i wdrażania funduszy, systemem wyboru projektów oraz ogólnie pojętym potencjałem instytucjonalnym. Stosunkowo duża liczba badań ewaluacyjnych dotyczyć będzie działań informacyjno-promocyjnych.

Począwszy od jesieni 2008 roku ruszy cały cykl seminariów poświęconych podsumowaniu rezultatów osiągniętych dzięki realizacji Narodowego Planu Rozwoju 2004–2006. Wielu tematów do dyskusji dostarczą przeprowadzone badania ewaluacyjne. Celem cyklu tych dyskusji strategicznych, w którym powinny wziąć udział jak najszersze środowiska profesjonalne, eksperckie i akademickie, będzie próba wyciągnięcia wniosków nie tylko co do sposobu konstruowania poszczególnych instrumentów i organizowania procesu wdrażania polityki spójności (proces ten *de facto* postępuje na bieżąco), lecz także skoncentrowania się na efektywności całych programów i polityk krajowych. Które z założeń strategicznych okazały się fałszywe? Które cele należy przeformułować? Jak zapewnić większą skuteczność i efektywność w kluczowych sferach, takich jak polityka naukowa, informatyzacja, realizacja projektów transportowych, edukacja, rozwój regionalny itp.?

Wnioski z debat będą mogły mieć charakter systemowy i jako takie będą podstawą do sformułowania propozycji zmian w sposobie kształtowania celów strategicznych rozwoju, poszczególnych polityk oraz zmian prawnych i organizacyjnych na wszystkich poziomach zarządzania w administracji publicznej w tym jej strategicznej koordynacji.

---

[12] Często pojawia się zarzut pod adresem realizowanych ewaluacji, że mają one bardziej charakter monitoringu rzeczowego, a nie właściwego badania ewaluacyjnego.

6. WYZWANIA DLA PROCESU EWALUACJI W POLSCE

Na obecnym etapie rozwoju systemu ewaluacyjnego w Polsce najważniejszym wyzwaniem jest systematyczne wykorzystywanie (przez decydentów, a więc także polityków!) wyników ewaluacji dla poprawienia skuteczności i efektywności działań publicznych – nie tylko w obszarze polityki spójności wspieranej przez fundusze UE, lecz także w ramach całości działań finansowanych ze środków publicznych w Polsce. Znalezienie odpowiedniego sposobu odpowiedzi na to wyzwanie jest trudne – głównie ze względu na zakładaną konieczność wprowadzenia zmian systemowych w sposobie funkcjonowania wszystkich instytucji państwa i zmian w sposobie postrzegania roli administracji. Innymi słowy – przejście od zbiurokratyzowanego modelu „administrowania" do aktywnego zarządzania powierzonymi zasobami i środkami publicznymi dla realizacji długofalowych celów rozwoju, a nie krótkookresowych celów politycznych.

Inne pomniejsze wyzwania obejmują takie zagadnienia jak:

- dostarczenie argumentów w trakcie dyskusji nad przyszłym kształtem polityki spójności;
- wprowadzenie systemu okresowych pomiarów wpływu funduszy UE na gospodarkę,
- doskonalenie systemu wskaźników, w tym wprowadzenie wskaźników rezultatów dla wszystkich działań publicznych (budżet zadaniowy),
- wykorzystanie narzędzia, jakim jest ewaluacja, w procesie przygotowania i realizacji innych – niezwiązanych z polityką spójności – polityk krajowych;
- silniejsze niż dotychczas powiązanie ewaluacji z procesem zarządzania programami operacyjnymi;
- szybkie wykształcenie potencjału do zlecania i odbierania ewaluacji na poziomie regionalnym;
- dalszy rozwój metodologii badań ewaluacyjnych;
- aktywizacja środowisk akademicko-naukowych na rosnącym rynku usług ewaluacyjnych;
- realizacja badań na niższych poziomach wdrażania (w tym na poziomie projektów).

Polska – mimo stosunkowo niedługiej, czteroletniej obecności w strukturach UE – stała się obecnie jednym z liderów wśród państw członkowskich w zakresie tworzenia sprawnie funkcjonującego systemu ewaluacji. Ocenę taką przedstawiła Komisarz ds. Polityki Regionalnej, Danuta Huebner[13]. Słowa te świadczą o docenieniu wysiłku, jaki włożył nasz kraj w tworzenie systemu ewaluacji. Powinniśmy je jednak traktować jako swoisty kredyt zaufania, patrząc bowiem na dotychczasowe osiągnięcia

[13] Wystąpienie na konferencji *Europejska polityka spójności jako czynnik wzrostu i wyrównywania różnic rozwojowych* w dniu 18 lipca 2008 roku w Warszawie.

w tym zakresie, można pokusić się o stwierdzenie, że najtrudniejsze dopiero przed nami. Zbudowanie systemu ewaluacji w oparciu o wymogi związane z regulacjami UE wydaje się znaczenie łatwiejsze niż zmierzenie się z dotychczasową praktyką działania systemu instytucjonalnego państwa. Od czegóż są jednak wyzwania?

BIBLIOGRAFIA

Kierzkowski T. (2002), *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, Warszawa.

Krajowa Jednostka Oceny, Departament Koordynacji Polityki Strukturalnej MRR, (2008), *Proces ewaluacji polityki spójności w Polsce. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość*, Warszawa.

Krajowa Jednostka Oceny, Departament Koordynacji Polityki Strukturalnej MRR (2008), *Baza badań ewaluacyjnych wg stanu na dzień 1 stycznia 2008 r.*, Warszawa.

Krajowa Jednostka Oceny, Departament Koordynacji Polityki Strukturalnej MgiP(2005), *Ewaluacja Narodowego Planu Rozwoju i programów operacyjnych. Poradnik*, MGiP, Warszawa.

MRR (2007), *Wytyczne nr 6 w zakresie ewaluacji programów operacyjnych na lata 2007–2013,* Warszawa.

MRR (2007), *System ewaluacji Narodowego Planu Rozwoju na lata 2004–2006 i Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, Warszawa.

MRR (2007), *Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, Warszawa.

Olejniczak K. (2002), *Towards the evaluation culture in Poland – experiences and prospects*, Aix en Province.

Olejniczak K. (2004) *Proces przeprowadzenia ewaluacji ex ante i on-going w ramach Narodowego Planu Rozwoju z uwzględnieniem etapów prac: planowanie, projektowanie i realizacja ewaluacji – wkład do procedur operacyjnych w Polsce, ekspertyza dla Ministerstwa Gospodarki i Pracy,* Warszawa.

Patton M.Q. (2002), *Qualitative Research & Evaluation Methods,* London, Sage Publications.

Rozporządzenie Rady (WE) nr 1083/2006 z dnia 11 lipca 2006 r. ustanawiające przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego oraz Funduszu Spójności i uchylającego rozporządzenie (WE) nr 1260/1999.

Rozporządzenie Rady (WE) nr 1260/1999 z dnia 21 czerwca 1999 r. ustanawiające przepisy ogólne w sprawie funduszy strukturalnych.

Shaw I.F., Greene J.C., Mark M.M. (2006), *Handbook of Evaluation. Policies, Programs and Practices,* London, Sage Publications.

www.ewaluacja.gov.pl

www.evalsed.info

CZĘŚĆ DRUGA

ZARYS METODOLOGII I ORGANIZACJI BADAŃ

ROZDZIAŁ 2.1.

Karol Olejniczak

EWALUACJA JAKO PROCES BADAWCZY

Niniejszy rozdział jest wprowadzeniem do części poświęconej metodologii i organizacji badań ewaluacyjnych. Przedstawiamy w nim dwie kwestie. Pierwszą jest rozszerzone interpretowanie procesu ewaluacji – nie tylko jako prowadzonego badania, ale szerzej – procesu rodzenia się zapotrzebowania na badanie, jego koncepcji, poszukiwania wykonawców, prowadzenia procesu badawczego, a wreszcie – wykorzystywania jego wyników przez bezpośrednich interesariuszy. Drugą omówioną kwestią są główne wybory i dylematy, przed którymi stają zlecający i wykonawcy ewaluacji. Te dwa zarysowane tematy są następnie pogłębiane przez kolejnych autorów, w kolejnych rozdziałach.

1. ETAPY PROCESU EWALUACJI

Przyjęta przez nas w rozdziale 1.1 definicja wyraźnie zaznacza, że ewaluacja jest procesem czerpiącym z badań społeczno-ekonomicznych.

W europejskich poradnikach badań ewaluacyjnych wyróżniane są standardowe etapy takiego procesu badawczego. Są to (European Commission, 2007):

- planowanie i strukturalizacja,
- zbieranie danych,
- analiza danych,
- ocena.

Ujęcie takie obejmuje jednak tylko wąską perspektywę trwającego już badania. Tymczasem w rzeczywistości badań ewaluacyjnych zlecanych przez instytucje publiczne, sam proces badawczy jest poprzedzony szeregiem działań, decyzji i wyborów, do których należy zaliczyć:

decyzję o podjęciu badania ewaluacyjnego, przygotowanie zlecenia na nie, przetarg, ostateczny wybór wykonawcy. Ewaluacja nie kończy się również w momencie oddania raportu przez wykonawcę. Strategia wykorzystania wyników, a także śledzenia efektów wprowadzania rekomendacji (ang. *follow up*), również powinna być uwzględniona przy całościowej wizji badania.

Dlatego też, analizując proces ewaluacji, należałoby przyjąć szerszą perspektywę, wychodzącą poza samo badanie ewaluacyjne. Takie rozszerzone ujęcie – złożone z trzech głównych etapów (i szeregu działań) – zaproponowano w tabeli 1.

Tabela 1. Etapy procesu ewaluacji

1.	**Planowanie i zlecanie ewaluacji**	
1.1	Konceptualizacja badania	Decyzja o podjęciu badania, określenie jego celów, ram czasowych, tematycznych, finansowych.
1.2	Przygotowanie zamówienia na badanie (operacjonalizacja)	Przygotowanie szczegółowego zlecenia na ewaluację (zwykle w formie Specyfikacji Istotnych Warunków Zamówienia – SIWZ).
1.3	Konkurs i wybór wykonawcy	Przygotowanie przez potencjalnych wykonawców ofert na badanie ewaluacyjne, a następnie wybór przez zlecającego jednej z propozycji badawczych.
2.	**Realizacja badania**	
2.1	Strukturalizacja	Opracowanie przez wykonawcę, we współpracy ze zlecającym, szczegółowego planu badawczego, dobór metod, opracowanie narzędzi badawczych.
2.2.	Gromadzenie danych	Zbieranie danych, prowadzenie obserwacji i badań przez wykonawcę ewaluacji.
2.3.	Analiza i ocena	Analiza zebranych danych i przeprowadzenie oceny ewaluandy na podstawie zebranych materiałów i danych.
2.4.	Przedstawienie wyników	Przedstawienie przez wykonawcę interesariuszom badania wyników i rekomendacji z ewaluacji.
3.	**Wykorzystanie wyników**	
3.1.	Wdrożenie strategii rozpowszechniania wyników	Realizacja przez zlecającego badania strategii rozpowszechniania wyników, obranej na początku badania.
3.2.	*Follow up* rekomendacji	Sprawdzenie stanu wdrożenia wniosków i rekomendacji z przeprowadzonego badania.

Źródło: opracowanie na podstawie European Commission, 1999: Vol. 1; Haber 2007; Olejniczak 2004.

To całościowe ujęcie zwraca uwagę na kilka kwestii. Po pierwsze, uwypukla wagę decyzji i wyborów podejmowanych przed formalnym zleceniem ewaluacji. To one w dużej mierze zdeterminują zarówno kształt badania, jak i jego użyteczność. Po drugie, wyraźnie podkreśla, iż praca koncepcyjna jest wykonana nie tylko przez wykonawcę – na etapie konkursu na badanie, a potem realizacji ewaluacji – ale także wcześniej przez zlecającego badanie, już na etapie planowania. Wreszcie po trzecie, podział ten sygnalizuje, iż ewaluacja nie kończy się w momencie oddania raportu. Wykorzystanie wyników winno być aktywnie i świadomie wspomagane przez zlecającego.

Całościową analizę i wyjaśnienie każdego z etapów tak szeroko rozumianego procesu ewaluacji Czytelnik znajdzie w rozdziałach 2.6 i 2.7, autorstwa A. Haber i B. Ledziona.

W tym rozdziale chcielibyśmy natomiast zwrócić uwagę na fakt, że wybory dokonane na etapie konceptualizacji, a potem tworzenia ofert, determinują kształt, przebieg badania, a w rezultacie także jego wiarygodność i użyteczność. Tym wyborom (czy dylematom wyboru) poświęcamy kolejny podpunkt.

2. EWALUACJA – SZTUKA WYBORÓW

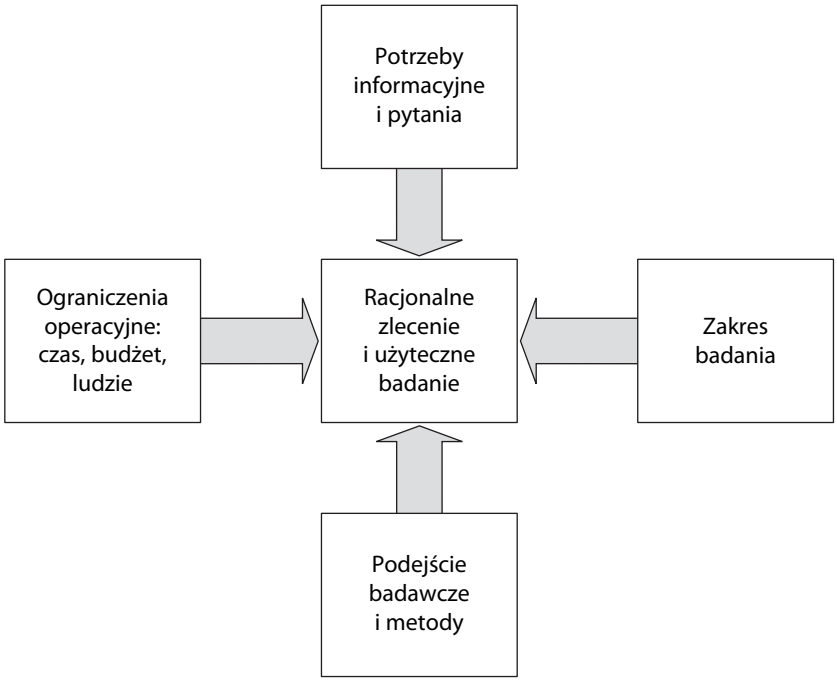
Podobnie jak w badaniach nauk społecznych, szeroko rozumiany proces ewaluacji zaczyna się od dyskusji nad postawieniem pytań badawczych, doborem strategii i planu badawczego (podejścia i metod badawczych) czy skalą badania (porównaj np. Bryman 2004: 3–58). Jednak specyfiką ewaluacji są cztery kwestie.

Po pierwsze, są to wysoce utylitarne potrzeby informacyjne. W odróżnieniu od badań naukowych, kierowanych ciekawością świata i chęcią poznania szerszych mechanizmów (weryfikowania ogólnych teorii), ewaluacje wynikają z wyjątkowo praktycznych potrzeb konkretnej grupy interesariuszy (np. wyjaśnienie przyczyn opóźnień programu, zidentyfikowanie efektów i potwierdzenie relacji przyczynowo-skutkowej między interwencją a jej otoczeniem etc.).

Po drugie, ewaluacja – w odróżnieniu od badań *stricte* naukowych, przykładających wagę do jednorodności i spójności warsztatowej danej dziedziny, dopuszcza, a nawet zaleca, większą dowolność metod, łączenie podejść, pewien „eklektyzm" metodologiczny.

Po trzecie, dokonywane wybory są silnie determinowane przez takie kwestie operacyjne jak dostępny budżet, a przede wszystkim czas dany na badanie (czy raczej oczekiwany termin przedstawienia wyników).

Wreszcie, po czwarte, wiele z decyzji określających ramy późniejszego badania podejmowanych jest nie przez samych badaczy, a przez jego interesariuszy. To z jednej strony wymaga świadomości i wiedzy tychże interesariuszy, z drugiej zaś – otwartości badaczy na dialog na temat



Rysunek 1. Kluczowe elementy strukturalizowania ewaluacji

Źródło: Ferry, Olejniczak 2008, w druku.

metod, podejść i kształtu badania, które mają prowadzić oraz potrzeb, które mają zaspokoić.

Jak pokazują najnowsze badania wykorzystania ewaluacji w Polsce (Ferry, Olejniczak 2008, w druku), można zidentyfikować cztery kluczowe elementy, wymagające wyważenia podczas planowania i strukturalizowania badania ewaluacyjnego (zob. rysunek 1).

Pierwszym elementem są potrzeby informacyjne, czyli to, co zlecający i główni interesariusze danej interwencji chcą otrzymać i dowiedzieć się z jej ewaluacji. Na poziomie strategicznym jest to określenie, jakie funkcje ma spełnić dane badanie (por. rozdział 1.1). Należy podkreślić omawiany już konflikt między funkcjami wspierania bieżącego zarządzania a funkcjami rozliczeniowymi. Na poziomie operacyjnym oznacza to sformułowanie konkretnych pytań, na które badanie ma przynieść odpowiedź, a także – gdy lista pytań jest długa i szeroka – wskazanie pytań priorytetowych[1]. W formułowaniu i grupowaniu pytań pomagają

[1] Warto zauważyć, że refleksja powinna obejmować nie tylko to, „jakie pytania zadać", ale przede wszystkim, „czy uprawnione i racjonalne jest ich zadawanie w danym momencie" (np. pytanie o rzeczywiste efekty może być przedwczesne dla słabo zaawansowanego programu). W przypadku dużych programów i planów dużych badań ewaluacyjnych w praktyce administracji anglosaskich pojawiały się (głównie w latach 70.) tzw. studia wykonalności ewaluacji (ang. *evaluability assessments* –

kryteria ewaluacyjne[2]. Teoretycznie wyróżnić możemy cztery typy stawianych pytań (European Commission, 2007):

- Pytania opisowe – starające się scharakteryzować istniejącą sytuację i prowadzoną interwencję (np. „Jak zmieniła się sytuacja na rynku pracy?”);
- Pytania o charakterze predyktywnym – starające się przewidzieć planowane zmiany, ich kontekst czy ewentualne procesy uboczne (np. „Czy wspieranie firm uczestniczących w sieci współpracy nie zakłóci konkurencyjności w regionie?”);
- Pytania przyczynowo-skutkowe – weryfikujące (logicznie i empirycznie) relacje między podjętymi działaniami a ich efektami, a także zmianami w adresowanym układzie przestrzennym czy społecznym (np. „Czy zarejestrowany wzrost eksportu jest bezpośrednim wynikiem kampanii promocyjnej lokalnych gron przedsiębiorczości?”);
- Pytania o charakterze normatywnym – oceniające daną interwencję według wybranych kryteriów i standardów (w przypadku programów funduszy strukturalnych są to omawiane w rozdziale 1.1 kryteria ewaluacyjne).

Oczywiście kluczowe są pytania normatywne. Bez nich ewaluacja traci swój charakter.

Elementem drugim jest decyzja na temat zakresu badania. Chodzi zarówno o zakres przedmiotowy (np. cały program albo tylko jego wybrane komponenty, wszystkie kwestie zarządzania albo tylko wybrane procesy), przestrzenny (np. wybrane obszary objęte programem, region, kraj), czasowy (wybrane lata funkcjonowania interwencji, cała edycja), jak i tematyczny (np. wybrany aspekt zarządzania interwencją). Praktyka ewaluacji (prezentowana również w niniejszym podręczniku, w rozdziałach dotyczących studiów przypadków) pokazuje, że ewaluacje procesów mogą mieć wąski zakres i krótki czas, podczas gdy ewaluacje efektów wymagają zarówno dużo dłuższego czasu, szeroko zakreślonego przedmiotu badania, jak i znaczącego zakresu przestrzennego. Wynika to z tego, że badania procesów są nastawione na szybką identyfikację ogniska problemu – procesu kluczowego, aspektu zarządzania,

---

EA). Składały się one z krótkiej analizy stanu interwencji, dostępności danych oraz potrzeb informacyjnych interesariuszy i ich gotowości do poddania się ewaluacji. Celem ich było więc zracjonalizowanie decyzji o badaniu. Miały zapobiegać sytuacjom rozpoczynania dużych, kosztownych i czasochłonnych badań o niskiej szansie na wiarygodne wnioski (np. z powodu zbyt wczesnego stadium rozwoju programu, zbyt skąpej ilości danych czy nawet niemierzalności danego fenomenu) (Mathison 2005: 137–139).

[2] W Polsce coraz częściej wykorzystywane jest drzewo pytań ewaluacyjnych jako podstawa do dyskusji o minimalnym zakresie pytań badawczych (Olejniczak 2007: 376). Zaletą grupowania pytań według kryteriów ewaluacyjnych jest to, że akcentują one element oceny (zmuszają więc do wyjścia poza pytania opisowe czy predyktywne).

elementu systemu. Krótkie, punktowe analizy kilku (z reguły przeciwstawnych) przypadków, są wystarczające do diagnozy problemów i szybkiego sformułowania działań zaradczych. Natomiast ocena całościowych efektów, aby była wiarygodna i przekonująca, musi opierać się na dłuższej i szerokiej obserwacji, a nie przyczynkowych studiach pojedynczych przypadków.

Elementem trzecim jest wybrane podejście badawcze, metodologia i techniki badania. Wybory podejść ewaluacyjnych są w dużej mierze determinowane funkcją ewaluacji (por. rozdział 1.1). Dobierane następnie metody i techniki mieszczą się oczywiście w ramach obranego podejścia, ale warto zauważyć, że w praktyce ewaluacji są zwykle przypisywane do każdego z pytań ewaluacyjnych, a następnie agregowane w jeden pakiet metodologiczny (szerzej o tym w rozdziałach 2.6 i 2.7, a także w częściach 3–5 niniejszej książki).

Ostatnim, czwartym elementem opisywanego układu, są warunki brzegowe:

- presja czasowa (moment, na który potrzebne są wyniki badania – często jest to wręcz „potrzeba chwili", szczególnie w przypadku ewaluacji mających rozwiązywać problemy zarządzania);
- dostępne zasoby ludzkie (a więc wymagani eksperci, skład i wielkość zespołu, który miałby wykonać badanie);
- zasoby finansowe, czyli budżet przeznaczony na daną ewaluację.

Te warunki brzegowe racjonalizują i ograniczają zwykle zbyt duże ambicje poznawcze i metodologiczne zarówno zlecających, jak i badaczy. Tak więc dla każdego badania ewaluacyjnego rozgrywa się swoista gra logiczna i poszukiwanie równowagi między powyższymi czterema elementami – listą nurtujących pytań i kwestii, ambicjami i możliwościami metodologicznymi, skalą badania oraz dostępnymi zasobami.

W rezultacie tej gry może zostać ograniczona lub poszerzona początkowa liczba pytań badawczych, zakres badania, warsztat metodologiczny (jednak przy zachowaniu jego rzetelności i bez naruszenia zasady triangulacji), zwiększony lub pomniejszony budżet projektu badawczego, czy czas przeznaczony na badanie.

Warto zwrócić uwagę, że wyważanie tych czterech elementów nie dokonuje się tylko w jednym momencie – przy projektowaniu badania – ale jest wykonywane trzykrotnie, przez różnych uczestników całościowo rozumianego procesu ewaluacji:

- przez zlecającego na wstępie – w fazie konceptualizacji i operacjonalizacji zlecenia na badanie ewaluacyjne;
- przez potencjalnych wykonawców w momencie przygotowywania ofert;
- przez wybranego wykonawcę, we współpracy ze zlecającym – w pierwszej fazie badania ewaluacyjnego (w fazie strukturalizacji).

Znalezienie równowagi między czterema elementami układu i dokonanie jasnych wyborów wprost rzutuje na jakość analizy, a co za tym

idzie – determinuje użyteczność ewaluacji. Brak równowagi (np. długa lista ogólnych pytań, połączona z brakiem danych, szerokim zakresem badania i niskim budżetem) skutkuje przyczynkowymi, fragmentarycznymi badaniami o niskiej jakości, wiarygodności i niskiej przydatności.

W kolejnych rozdziałach niniejszej części podręcznika odnosimy się właśnie do omówionych wyborów związanych z procesem ewaluacji.

W rozdziale 2.2. autorstwa J. Górniaka i K. Keler omawiana jest kwestia wskaźników. Są one rdzeniem myślenia o interwencjach publicznych w kategoriach zarządzania opartego na wynikach. Zmuszają więc do racjonalnego i konkretnego formułowania pytań ewaluacyjnych, a następnie ich weryfikacji.

Metody badawcze omawiamy w rozdziałach 2.3.–2.5. Objętość i forma niniejszej publikacji nie pozwalają nam na szczegółowe omawianie podejść ewaluacyjnych[3]. Świadomie uniknęliśmy też formy poradnikowej – katalogu metod[4]. Skoncentrowaliśmy się natomiast na całościowym i krytycznym (w kontekście badań ewaluacyjnych) przedstawieniu trzech grup metod – makroekonomicznych, ilościowych i jakościowych.

Wreszcie w rozdziałach 2.6 i 2.7 prezentujemy grę wszystkich czterech czynników i konsekwencje wyborów z perspektywy zlecającego i z perspektywy wykonawcy.

BIBLIOGRAFIA

Bryman A. (2004), *Social research methods*, Oxford, Oxford University Press.

European Commission (1999), *MEANS collection: evaluating socio-economic programmes*. DG-Regio, Office for Official Publications of the European Communities, Luxembourg.

European Commission (2007), Evalsed – the resource for the evaluation of socio-economic development, 12.04.2008, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/index_en.htm

Ferry M., Olejniczak K. (2008, w druku), *The use of evaluation in management of EU programmes – evidences from Poland*, Ernst & Young, Program „Sprawne Państwo", www.sprawnepanstwo.pl

Haber A. (2007), Proces ewaluacji *ex post* – konceptualizacja, operacjonalizacja, realizacja badania ewaluacyjnego, w: Haber A. (red.) *Ewaluacja ex post. Teoria i praktyka badawcza*, Warszawa, Polska Agencja Rozwoju Przedsiębiorczości.

Mathison S. (ed.) (2005), *Encyclopedia of evaluation*, Thousand Oaks, Calif., London, Sage.

[3] Zainteresowany Czytelnik może znaleźć wyczerpujący przegląd podejść ewaluacyjnych (nazywanych „modelami") w: Stufflebeam, Shinkfield 2007.

[4] Czytelników poszukujących takich informacji odsyłamy do: European Commission 1999, vol. 3; 2007: Sourcebook; Mathison 2005.

Olejniczak K. (2004), *Proces przeprowadzania ewaluacji ex ante i on-going w ramach NPR z uwzględnieniem etapów prac: planowanie, projektowanie i realizacja ewaluacji – wkład do procedur operacyjnych ewaluacji w Polsce*, Krajowa Jednostka Oceny, http://www.funduszestrukturalne.gov.pl/EWALUACJA+funduszy+w+Polsce/Dokumenty+ewaluacyjne/

Olejniczak K. (2007), *Ewaluacja jako narzędzie zarządzania w sektorze publicznym*, w: Gorzelak G., Tucholska A. (red.), *Rozwój, region, przestrzeń*, Warszawa, Ministerstwo Rozwoju Regionalnego.

Stufflebeam D.L., Shinkfield A.J. (2007), *Evaluation theory, models, and applications*., San Francisco, Calif., Jossey-Bass, Chichester, John Wiley [distributor].

ROZDZIAŁ 2.2.

Jarosław Górniak
Karolina Keler

ROLA SYSTEMÓW WSKAŹNIKÓW W EWALUACJI

Celem niniejszego rozdziału jest omówienie kwestii istotnych z punktu widzenia konstrukcji, doboru i użycia wskaźników w ewaluacji. Tekst składa się z dwóch zasadniczych części. W pierwszej nakreślone zostały najważniejsze pojęcia oraz kwestie fundamentalne dla budowy systemów wskaźników. W drugiej omówione zostały zagadnienia użycia gotowych systemów wskaźników w ewaluacji.

Część pierwszą rozpoczynamy od nakreślenia funkcji wskaźników w ewaluacji. Przedstawiamy je w kontekście analizy zamierzonych i niezamierzonych efektów interwencji. Następnie omówione zostały typy wskaźników oraz pożytki i ryzyko związane z ich stosowaniem. W części drugiej proponujemy sposób podejścia do budowy systemu wskaźników dla projektu czy programu, który pozwala uniknąć pułapek trywialności, bezużyteczności i błędności wskaźników. Omawiamy kryteria oceny, które pozwalają sprawdzić, czy przyjęte przez nas wskaźniki będą przydatne, rzetelne i trafne. Omawiamy też role wskaźników w kontekście relacji między monitoringiem ewaluacją oraz badania efektów interwencji. W artykule podkreślona jest rola analizy przyczynowej w ewaluacji programów. Całość zamyka dyskusja nad potencjalnymi problemami związanymi z użyciem systemów wskaźników.

1. WPROWADZENIE

Ewaluacja – pośród swoich rozmaitych funkcji – służy udokumentowanej ocenie, jakie efekty przyniósł program oraz czy uzyskane efekty są

rzeczywiście skutkiem podjętych działań. Cele programów wskazują zwykle na pożądaną zmianę w pewnym obszarze rzeczywistości (lub niedopuszczenie do zmiany wywoływanej przez rozmaite czynniki). Po to, by ewaluacja była użyteczna, jej wykonawcy powinni zmierzyć się z odpowiedzią na pytanie, czy zawarte (czasem ukryte) w programie założenia co do związków przyczynowych między zaplanowanymi działaniami a oczekiwaniami efektów tych działań, zawartymi w celach, były słuszne (testujemy teorię programu). W procesie ewaluacji interesuje nas również to, jak wdrażane były projekty działań, czy przyjęte założenia co do sposobu realizacji programu były właściwe (testujemy teorię wdrożenia) i identyfikujemy zarówno przeszkody, które utrudniały wdrożenie, jak i dobre praktyki w tym względzie[1]. Ewaluacja jest powiązana z monitoringiem, którego celem jest bieżące śledzenie tego, czy podejmowane są zaplanowane działania i czy przynoszą one spodziewane wyniki, a zasoby są użytkowane w sposób zaplanowany i efektywny. Podstawą ewaluacji i monitoringu jest empiryczne uchwycenie zarówno zmian związanych z programem, jak i właściwości procesu wdrożenia – temu służą wskaźniki. Wskaźniki są określane w kontekście ewaluacji i monitoringu często jako mierniki służące ocenie stopnia, w jakim podjęte działania przybliżyły osiągnięcie zaplanowanych wartości związanych z przyjętymi celami. Jednak stosowalność wskaźników wykracza poza pomiar efektów, podobnie jak przedmiotem ewaluacji jest nie tylko ocena stopnia wykonania celów. Wskaźnik to mierzalna cecha, która, pojedynczo lub w zestawie z innymi takimi cechami, dostarcza świadectwa wystąpienia określonego stanu rzeczy w sposób użyteczny pod względem menedżerskim i/lub naukowym. W programach publicznych wskaźniki przyjęte przez planistów są przede wszystkim powiązane z założonymi celami, lub – ogólniej – pożądanymi i zamierzonymi efektami. W niniejszym tekście pokażemy, że w procesie ewaluacji programów system wskaźników powinien pozwalać na uchwycenie także efektów niepożądanych i niezamierzonych.

Wskaźniki pełnią istotne funkcje na różnych etapach cyklu polityk publicznych, a także w poszczególnych fazach procesu ewaluacji. Na etapie planowania interwencji wskaźniki służą wewnętrznej integracji i koordynacji projektowanego przedsięwzięcia. Umożliwiają precyzyjne, wymierne i spójne formułowanie celów oraz powiązanie zaplanowanych efektów z nakładami i działaniami; dobrą weryfikacją metodologicznej jakości programu jest sprawdzenie możliwości jego ewaluacji w przyszłości (*evaluability*). W trakcie realizacji interwencji oraz po jej ukończeniu systemy wskaźników umożliwiają określenie efektów, nakładów i materialnych przejawów podejmowanych działań, przez co staje się

[1] Podstawy koncepcji ewaluacji opartej na teorii programu zawdzięczamy pracom Carol Hirschon Weiss i Huey-Tsyh Chena; por. m.in. Birickmeyer, Weiss 2000; Chen 2004.

możliwa ocena skuteczności i efektywności interwencji[2]. Przedmiotem pomiaru z użyciem wskaźników są zatem:

- nakłady – poniesione koszty, bezpośrednie i pośrednie, finansowe i rzeczowe;
- produkty – efekty zdefiniowane w postaci celów operacyjnych, będące bezpośrednią konsekwencją wykonania założonych działań, np. zrealizowane szkolenie, przeszkolenie osoby, wypłacone stypendium czy dotacja, raport z badań, strona internetowa itp.
- rezultaty – efekty zdefiniowane w postaci celów szczegółowych, które osiągnięto w grupie docelowej, wiążące się z jakąś zmianą w tej grupie, np. określone kwalifikacje, nabyte przez beneficjentów ostatecznych, zdobycie i utrzymanie zatrudnienia po czasie realizacji interwencji, zwiększona jakość usług oferowanych przez instytucje publiczne.
- oddziaływania/wpływ – efekty zdefiniowane w postaci celów ogólnych, określających, w jaki sposób interwencja wpłynie na rozwój społeczny lub rozwiązanie problemu czy kwestii społecznej, np. zwiększony kapitał ludzki, mniejsza skala wykluczenia społecznego.

2. EFEKTY ZAMIERZONE I NIEZAMIERZONE

Interwencje publiczne dotyczą złożonej rzeczywistości, w której występują liczne interakcje. Dlatego stawiane cele operacyjne, ujęte w precyzyjnej, wymiernej postaci, udaje się zwykle osiągnąć tylko w przybliżeniu, czasem nie udaje się osiągnąć, a bywa, że efekty są niezgodne z oczekiwaniami. Ponadto, podejmowane działania prowadzą nie tylko do efektów ujętych w celach, lecz także mogą mieć inne konsekwencje, ważniejsze i mniej ważne. Biorąc to wszystko pod uwagę, efekty interwencji publicznych możemy podzielić na: zaplanowane cele, użyteczne efekty dodatkowe i niezamierzone konsekwencje negatywne (por. tabela 1). Wykluczamy tu

Tabela 1. Efekty interwencji publicznych

	Pożądane	Niepożądane
Zamierzone	Cele	–
Niezamierzone	Użyteczne efekty dodatkowe	Negatywne konsekwencje

Źródło: opracowanie własne.

[2] Mianem interwencji publicznych określać będziemy tu wszelkie polityki publiczne, programy i projekty, czyli celowe działania podmiotów publicznych lub jednostek działających przy wykorzystaniu ich środków, na zlecenie lub w uzgodnieniu z nimi, zmierzające do rozwiązania jakiegoś problemu społecznego lub do ogólnie rozumianego rozwoju społecznego, jeśli nie będzie zachodziła potrzeba rozróżnienia pomiędzy tymi kategoriami.

z rozważań celową działalność szkodliwą, którą można określić mianem sabotażu społecznego.

Każdy rodzaj efektów może być ujęty jako zmiana (lub jej brak) obserwowana na poziomie właściwie dobranych wskaźników. W rzeczywistości zmiany te mogą być wywołane – i zwykle są – nie tylko przez interwencję, lecz także przez wiele innych oddziaływań. Chcąc poznać stopień, w jakim to interwencja przyczyniła się do ich wystąpienia, należy prowadzić przyczynową analizę oddziaływań, czyli zmierzać do ustalenia faktycznego, bezpośredniego efektu interwencji (tzw. efektu netto, o którym piszemy w części 6. rozdziału). Zastosowanie klasycznego schematu eksperymentalnego w przypadku polityk publicznych jest trudne, a najczęściej nie jest w ogóle możliwe, dlatego też stosuje się rozmaite schematy quasi-eksperymentalne, oparte na porównaniu grupy docelowej z odpowiednio dobraną próbą kontrolną. Analiza przyczynowa jest uważana przez niektórych teoretyków za esencję ewaluacji (por. Langbein, Felbinger 2006: 9–14).

Efekty niezamierzone mogą mieć charakter pozytywny lub negatywny[3] zarówno wobec grup objętych działaniami interwencji, jak i wśród innych osób. Pozytywne skutki działań są wartością dodaną działań i są ujęte w kryterium użyteczności, stosowanym powszechnie w ewaluacji. Negatywne efekty powinny być przedmiotem troski i w miarę możliwości także przedmiotem ewaluacji. Przejawiają się one, chociaż nie wyczerpują, w takich dobrze znanych efektach, jak efekt wypierania i efekt substytucji (przykłady efektów niepożądanych znajdują się w tabeli 2). Ponadto,

Tabela 2. Przykłady efektów niepożądanych

Efekt niepożądany	Charakterysyka efektu
wypierania	udzielanie wsparcia jednemu przedsiębiorstwu daje mu przewagę konkurencyjną, niewynikającą z aktywności rynkowej i prowadzi do wyparcia z rynku innych przedsiębiorstw, co skutkuje m.in. redukcją zatrudnienia, niwelującą wzrost uzyskany w wyniku interwencji
substytucji	w przypadku subsydiowania zatrudnienia może wystąpić zjawisko zastępowania dotychczas zatrudnionych pracowników nowymi, których płace są subsydiowane
zdarzenia niezależnego (*deadweight*)	udzielenie wsparcia osobom bezrobotnym, które nawet bez pomocy powróciłyby na rynek pracy (występuje, gdy źle została zdefiniowana grupa docelowa, gdy beneficjentami zostały osoby w rzeczywistości niezagrożone)

Źródło: opracowanie własne.

[3] Przykład analizy efektów zamierzonych i niezamierzonych można znaleźć w Murray 1988.

często ma miejsce tzw. efekt zdarzenia niezależnego (*deadweight*), który oznacza taki zakres zmiany, który wystąpiłby bez ponoszenia kosztów na realizowane działania. Tutaj efektem niepożądanym jest niemożność efektywnego przeznaczenia wydanych środków publicznych na inne cele. Uchwyceniu tych wszystkich efektów powinien służyć właściwy dobór wskaźników, połączony z odpowiednim metodologicznym planem badania ewaluacyjnego.

Interwencja może się okazać nieskuteczna. Ryzyko niepowodzenia powinno być uwzględniane, ale też niepowodzenia powinny być obowiązkowo analizowane w procesie ewaluacji, a wyciągnięte z tego wnioski powinny być uwzględniane w procesie planowania kolejnych interwencji lub modyfikacji aktualnie wdrażanych. Ucząca się instytucja potrafi wyciągnąć wnioski z takich lekcji, opartych na ewaluacji. Wobec możliwości wystąpienia niezamierzonych efektów ubocznych, zwłaszcza negatywnych, projektanci przedsięwzięć powinni brać pod uwagę konsekwencje i kierować się fundamentalną zasadą – interwencja ma przede wszystkim nie szkodzić.

Reasumując, jeżeli ewaluacja ma się przyczynić do (szeroko rozumianego) rozwoju, powinna eliminować działania nieskuteczne, czyli takie, które nie prowadzą do celów, oraz minimalizować możliwość pojawienia się niepożądanych skutków ubocznych. Systemy wskaźników przyjęte w ewaluacji powinny pozwalać na pomiar tych zjawisk. Dzięki badaniom ewaluacyjnym można udzielić odpowiedzi na fundamentalne pytania: dlaczego efekty negatywne się pojawiły i jak można im zapobiec oraz czy podjęte działania, przy danych kosztach, rzeczywiście wywołują zaplanowane efekty w stopniu zgodnym z oczekiwaniami. Dlatego też system wskaźników powinien obejmować nie tylko te aspekty zjawisk, które chcemy intencjonalnie zmienić, czyli tzw. wskaźniki programowe. Systemy te powinny zawierać również tzw. wskaźniki kontekstowe (por. Komisja Europejska 2006: 7), pozwalające sprawdzić, jakie zmiany zaszły w czasie wdrożenia programu w szerszym kontekście społeczno-ekonomicznym, oraz wskaźniki dotyczące efektów niezakładanych, ale z którymi mogą wiązać się istotne negatywne skutki.

3. WSKAŹNIKI W OCENIE EFEKTÓW

Namysł nad wskaźnikami powinien towarzyszyć interwencjom od momentu rozpoczęcia ich programowania, gdyż dzięki nim możliwe jest właściwe określenie celów operacyjnych, pomiar efektów i w konsekwencji ocena programu. Aby zrozumieć rolę wskaźników w tym zakresie, należy wyjaśnić, co rozumiemy pod pojęciem „wskaźnik" oraz przedstawić najważniejsze typy wskaźników.

Wskaźniki służą pomiarowi. W klasycznym ujęciu „wskaźnik zdarzenia (własności) *Z* to takie zdarzenie (taka własność) *W*, że stwierdzenie

jego (jej) istnienia, pojawienia się lub stopnia intensywności bądź faktycznie, jest wykorzystane jako przesłanka, bądź zasadnie nadaje się na przesłankę wnioskowania, iż w określonych przypadkach z pewnością, z określonym prawdopodobieństwem lub przynajmniej wyższym niż przeciętnie, wystąpiło zdarzenie (własność) *Z*" (Nowak 1985: 165). Wskaźniki wykorzystywane w ewaluacji ujmują w sposób ilościowy (czyli ujmują „stopień intensywności") przede wszystkim różne wymiary zjawisk, na które ma oddziaływać dana interwencja. Jeżeli efektem interwencji ma być jakaś zmiana, wówczas musimy jasno określić, jakie cechy określonych grup, w określonym miejscu i czasie mają ulec zmianie pod wpływem określonych instrumentów, a te cechy podlegające zmianie reprezentują wskaźniki. Zmiana wartości wskaźnika będzie nas informować o tym, że zaszła zmiana w zakresie wyrażanym przez wskaźnik.

Przedmiotem ewaluacji jest zawsze interwencja dokonywana przy pomocy przyjętych środków w określonej sytuacji, a więc musimy ująć przy pomocy wskaźników także same działania (instrumenty interwencji) oraz cechy kontekstu: grupy docelowej, miejsca i czasu. Określone rodzaje interwencji mogą mieć skuteczność ograniczoną tylko do pewnego rodzaju grup docelowych, znajdujących się w określonej sytuacji i czasie, dlatego należy uwzględnić możliwość występowania rozmaitych interakcji w planowaniu ewaluacji (np. czy badając skutek wprowadzenia nakazu jazdy z włączonymi światłami, należy badać zmianę prawdopodobieństwa wypadku nie tylko w ciągu całego roku, ale też np. w podziale na pory roku, bo może się okazać, że pora roku warunkuje wystąpienie efektów; badając skuteczność szkoleń należy ją analizować w podziale na różne kategorie beneficjentów, gdyż może występować interakcja pomiędzy ich cechami, np. wykształceniem a skutecznością działań). Uchwycenie interakcji wymaga zatem użycia nie tylko wskaźników efektów programu, lecz także wskaźników środków interwencji (są nimi faktycznie produkty) oraz cech beneficjentów i kontekstu.

Na ogół wykorzystywane statystyki powinny być wyliczane tak, by umożliwiać porównania między grupami czy okresami. Dlatego zazwyczaj używamy miar względnych, a nie bezwzględnych: np. lepiej jest śledzić stopę zatrudnienia obliczaną, powiedzmy, jako procent mieszkańców danego obszaru w wieku 18–75 lat, którzy są zatrudnieni (miara względna), niż po prostu liczbę osób zatrudnionych na poszczególnych obszarach (miara bezwzględna). Podobnie, w przypadku oceny efektu projektu czy programu **zazwyczaj** podzielimy liczbę beneficjentów, wśród których rezultaty zostały osiągnięte (zaszła planowana zmiana) do liczby wszystkich osób, które skorzystały ze wsparcia. Piszemy: „zazwyczaj", gdyż definicja wskaźnika zawsze musi być dostosowana do potrzeb informacyjnych.

Wskaźniki mogą mieć też bardziej złożony charakter. Dobrym przykładem, ilustrującym taki rodzaj wskaźników, jest znany z listy tzw. wskaźników z Laeken wskaźnik zagrożenia ubóstwem (zob. tabela 3). Bardziej

Tabela 3. Przykłady wskaźników w podziale na „typy"

Rodzaj wskaźników	Opis	Przykłady
proste	Sumowanie, wyliczanie wartości procentowych (udział, proporcja), uśrednianie itp.	liczba mieszkańców w wieku 18–75 lat na danym obszarze, którzy są zatrudnieni; procent osób zatrudnionych w wieku 18–75 lat na poszczególnych obszarach wśród osób wszystkich osób dorosłych w tej grupie wiekowej.
złożone	Uwzględnienie w konstrukcji wskaźnika opartego na jednej zmiennej wiodącej kilku informacji, w szczególności z różnych poziomów agregacji.	„odsetek osób o niskich (60% mediany) dochodach" (*at-risk-of-poverty rate*). Jest to odsetek osób, których wartość dochodu (rozumianego jako ekwiwalentny dochód do dyspozycji, tzn. przypisywany osobie całkowity dochód gospodarstwa podzielony przez „ekwiwalentną wielkość" gospodarstwa domowego) jest niższa niż 60% wartości mediany dochodów ekwiwalentnych w badanej populacji[4]. Wskaźnik ten wskazać ma na obszary bądź kategorie osób najbardziej zagrożone ubóstwem.
złożone indeksy	Wskaźniki złożone z wskaźników.	Przykładem może być indeks ubóstwa HPI-2 (Human Poverty Index). Indeks ten jest miarą złożoną z czterech wskaźników: prawdopodobieństwo (w momencie urodzenia) przedwczesnej śmierci (przed 60. rokiem życia; w %); odsetek osób dorosłych będących analfabetami funkcjonalnymi (w %); odsetek osób żyjących poniżej linii ubóstwa (w %) oraz stopa długotrwałego bezrobocia (>12 miesięcy; w %). Wartość indeksu dla danego kraju uzyskiwana jest przez wyliczenie pierwiastka trzeciego stopnia ze średniej ważonej z powyżej podanych wskaźników podstawowych, po uprzednim podniesieniu ich wartości do trzeciej potęgi. Indeks ma pozwalać wskazać najlepsze kraje pod względem minimalizacji nasilenia zjawiska ubóstwa[5].

Uwaga: nie jest to ścisła typologia; użyta została, by pokazać, że statystyczne „skomplikowanie" sposobu liczenia wartości wskaźnika może mieć swoje plusy, gdyż lepiej oddaje złożoną rzeczywistość. Należy jednak zwrócić uwagę na to, czy za złożonymi matematycznymi kalkulacjami nie kryje się w istocie nadmierne i arbitralne uproszczenie rzeczywistości (tak jak to ma niekiedy miejsce w przypadku indeksów).

Źródło: opracowanie własne.

---

[4] Szczegółowy opis algorytmu wyliczania wartości tego wskaźnika można znaleźć w publikacji: Frączek et al. (2007a: 143–149) lub w oryginale, na przykład w European Commission 2005: 23 i dalsze.

[5] Por. Frączek et al. 2007a: 190–202 lub http://hdr.undp.org/en/media/hdr_20072008_tech_note_1.pdf oraz http://hdr.undp.org/docs/statistics/indices/Calculating_Indices_EXCEL.xls

złożone algorytmy mają zapewnić adekwatność wskaźnika. Z drugiej strony jednak, nierzadko wymagają od autora gruntownej wiedzy metodologicznej oraz znajomości sposobów badania danego zjawiska.

Jeszcze innym typem wskaźników są złożone indeksy, tworzone na bazie informacji jednostkowych lub zagregowanych. Pełne omówienie zagadnienia wykracza poza ramy tego rozdziału[6]. Indeksy złożone tworzy się na podstawie wskaźników takich jak omawiane powyżej. Najprostsze spotykane rozwiązanie to zsumowanie wartości liczbowych poszczególnych wskaźników składowych lub wyliczenie wartości przeciętnej wskaźników dla poszczególnych przypadków (np. średniej arytmetycznej czy ważonej; po ich uprzedniej normalizacji). Przykładem może być indeks ubóstwa HPI-2 (Human Poverty Index), omówiony w tabeli 3. Ten przykład, podany w dość ogólnej postaci, przytoczyliśmy po to, by zilustrować możliwość konstruowania syntetycznych miar efektów programu, szczególnie wtedy, gdy cel programu ma wiele aspektów i chcemy je połączyć w jedną miarę, dającą przejrzyste wyniki.

Należy jednak pamiętać, że najlepiej jest, gdy indeksy ujmują jakiś wspólny wymiar przejawiający się przez syntezę informacji zawartych w wielu szczegółowych wskaźnikach. Jeśli tak się nie dzieje, to efekty uzyskiwane na poszczególnych wymiarach składowych mogą się wzajemnie znosić. Jeśli ze względów statystycznych lub interpretacyjnych złożony indeks nastręcza trudności, lepiej jest stosować zestawy wskaźników, które ujmują te różne wymiary badanego zjawiska.

4. BUDOWA SYSTEMU WSKAŹNIKÓW W RAMACH PLANU EWALUACJI

Wskaźniki pozwalają wykonać diagnozę sytuacji, w której interwencja ma być podjęta, jest realizowana lub została zakończona. Mają pomóc w odpowiedzi na klasyczne, metaforyczne pytania[7]:

- **W jakim miejscu jesteśmy?** Wskaźniki pozwalają ustalić sytuację w punkcie wyjścia i określić problemy oraz ich dotkliwość. Część wskaźników użytych do diagnozy wejdzie w skład zestawu służącego monitoringowi i ewaluacji efektów, część będzie dostarczała informacji o kontekście interwencji pozwalając dostrzec efekty niezamierzone.
- **W jakim kierunku chcemy pójść?** Wskaźniki pomagają w ustaleniu priorytetów i zdefiniowaniu celów, a zwłaszcza ujęciu ich w kategoriach docelowych wartości na wskaźnikach przyjętych do ewaluacji celów.

---

[6] Więcej na temat problematyki budowy indeksów, a zwłaszcza tzw. indeksów złożonych stosowanych na omawianym gruncie, można znaleźć np. w: Booysen 2002; Salzman 2003; Giovannini et al. 2005.

[7] W tych kategoriach komentowane jest użycie wskaźników m.in. w: Global Water Partnership Technical Brief 3.

- **Czy obraliśmy właściwą drogę?** Śledzenie wartości wskaźników w procesie monitoringu, a także wykorzystywanie ich w ewaluacji śródokresowej, pozwala określić, czy interwencja przebiega właściwie, jak są wykorzystywane środki i czy są przyczyniają się do osiągnięcia zaplanowanych efektów.
- **Czy jeszcze tam jesteśmy?** Wskaźniki pozwalają na rozliczenie z efektów interwencji oraz, przez ukazanie efektów, uzyskanie poparcia społecznego.

Opracowanie zestawu wskaźników dla określonej interwencji (polityki publicznej, programu czy projektu) wymaga zaangażowania zasobów, co najmniej ludzkich. Opiera się na fazach postępowania zbliżonych do tych, które podejmuje się w fazie konceptualizacji i operacjonalizacji w procesie badań.

Etap pierwszy związany jest z **identyfikacją problemów i ustanawianiem celów oraz możliwych niezamierzonych efektów interwencji**. Kluczowym zadaniem w tej fazie jest zapewnienie trafności i kompletności zestawu wskaźników: powinny one ujmować kluczowe aspekty rzeczywistości, związane z danym obszarem problemowym. Punktem wyjścia jest zidentyfikowanie kluczowych wymiarów analizowanych kwestii społecznych, a także celów stawianych przed interwencją przez decydentów (z udziałem istotnych interesariuszy). Pojęcia służące do ujęcia celów muszą być jasno zdefiniowane. Następnie przekładamy zidentyfikowane wymiary na sposób ich empirycznego uchwycenia. Zwykle przygotowuje się większą liczbę propozycji, z których wybiera się następnie te, które najlepiej spełniają wymagania stawiane na dalszych etapach. Bardzo ważnym kryterium selekcji na tym etapie jest możliwość dobrego uchwycenia przy pomocy wybranych wskaźników wartości docelowych, które dobrze reprezentują w kategoriach ilościowych cele interwencji. Ostatecznie nie można zbudować systemu wskaźników, które są narzędziami pomiaru efektów, bez jasnych definicji celów i bez dobrego rozróżnienia, co jest tylko produktem działań, a co rzeczywistym rezultatem czy oddziaływaniem. Definiowanie celów i dobór wskaźników jest procesem iteracyjnym: krok po kroku cele precyzowane są w kategoriach wskaźników i dookreślane są koncepcje wskaźników w odniesieniu do coraz bardziej klarownych celów. Ostatecznie system wskaźników efektów powinien pozwalać na ocenę stopnia osiągnięcia celów w trakcie i po zakończeniu interwencji – dawać podstawy do ewaluacji. Wskaźniki odnoszące się do celów powinny cechować się jasną interpretacją normatywną, tzn. zmiana wartości powinna dawać podstawę do pozytywnej lub negatywnej oceny. System wskaźników powinien ujmować także niezamierzone efekty działań. Budując taki system, powinniśmy zatem zadać też pytanie o to, na co i na kogo program może mieć jeszcze wpływ (na innych ludzi, na inne obszary, na inne przedsiębiorstwa itp.). Wskaźniki powinny być zatem tak dobrane, by w pełni i adekwatnie ujmowały kluczowe aspekty sytuacji społecznej,

na które interwencja może oddziaływać – celowo lub w sposób niezamierzony.

Dobór wskaźników powinien być prowadzony we współpracy z kluczowymi interesariuszami. To ogólny postulat, który niekiedy może być kłopotliwy w realizacji. Partycypacja kluczowych interesariuszy w fazie konceptualizacji ustanawiania celów ma prowadzić do lepszego zrozumienia, akceptacji i łatwiejszego wdrożenia interwencji oraz uwzględnienia dodatkowej informacji specyficznej dla różnych środowisk i osób. Niebagatelnym postulatem jest to, by wskaźniki były zrozumiałe dla interesariuszy i szerszej opinii publicznej. Czasami wymaga to pewnej kompetencji statystycznej, co jest trudnym do zrealizowania postulatem, a więc chodzi o to, by intuicyjna interpretacja wskaźników nie wprowadzała odbiorców w błąd.

Kolejny etap to **weryfikacja wskaźników z punktu widzenia ich praktycznego wykorzystania.** Najbardziej pragmatyczną kwestią jest **dostępność danych**, gdyż nawet najlepszy merytorycznie system wskaźników nie zda się na nic, jeżeli nie będziemy mieć adekwatnych danych. Jeśli dane muszą zostać wygenerowane w procesie badań, należy oszacować koszt ich pozyskania i podjąć decyzję o tym, czy zakwalifikować do systemu poszczególne wskaźniki. Informacje są ważne dla realizacji i oceny interwencji, ale – trawestując słynne powiedzenie – my nie znamy pojęcia informacji za wszelką cenę. Wartość informacji to różnica pomiędzy korzyściami, jakie przynosi, a kosztem jej uzyskania. Najtaniej jest wtedy, gdy wskaźniki opierają się na danych gromadzonych w toku obowiązkowej sprawozdawczości lub danych zastanych, zbieranych przez rozmaite instytucje dla innych celów (np. Główny Urząd Statystyczny).

Następnie należy dokonać **weryfikacji wskaźników z punktu widzenia ich statystycznej wiarygodności**. Należy ustalić metody gromadzenia danych, które nie są dostępne ze źródeł zastanych i zweryfikować jakość statystyczną tych ostatnich. Dane wykorzystywane do zasilania wskaźników powinny w miarę możliwości spełniać poniższe warunki (por. Atkinson et al. 2002):

- wskaźniki powinny być oparte na wiarygodnych danych niepoddanych 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[8];

[8] Pomiar obciążony jest błędem systematycznym jeżeli zawyża lub zaniża wartości, które są mierzone; średnia błędów takich pomiarów nie sumuje się do 0, jak w przypadku błędów losowych, przypadkowych (zob. np. Ferguson, Takane 2007: 489).

- powinno się unikać wskaźników, które podlegają silnie wpływowi nieprzewidywalnych (np. pogoda) i niemożliwych do wyjaśnienia czynników;
- należy zachować ostrożność wobec wskaźników, których wartość w znacznym stopniu nie zależy od polityk publicznych, lecz od czynników niezależnych, np. cyklu koniunkturalnego;
- dane powinny być aktualne.

Wreszcie, należy ocenić wskaźniki z punktu widzenia tego, czy zmiany ich wartości właściwie oddają zmiany pod wpływem interwencji:

- wskaźnik powinien zmieniać wartość, jeśli interwencja przynosi efekty, natomiast nie powinien podlegać zbyt łatwo zmianom pod wpływem błędów i zakłóceń, czyli niesystematycznych oddziaływań różnych zdarzeń niemających związków z interwencją;
- wskaźniki powinny dać się powiązać z działaniami;
- wskaźniki nie powinny pozwalać na manifestowanie pozornych efektów (np. proporcję podmiotów objętych świadczeniem można łatwo zwiększyć, przyznając minimalne świadczenia na szeroką skalę);
- należy sprawdzić, czy zmienność wartości wskaźnika w czasie (np. sezonowe) i przestrzeni (np. pomiędzy regionami) oddaje wiarygodnie zmiany w relacji do celów, czy też odzwierciedla zmiany niezależne.

Sprawdzenie powyższych cech danych może wymagać wykonania pilotażowych badań wskaźników. Dane mogą mieć różną jakość. Czasami dobieramy wskaźniki na podstawie dostępnych, a nie najlepszych możliwych sposobów pomiaru danego aspektu problemu. Wybór wskaźników do systemu polega w tym przypadku na skorygowaniu definicji potrzeb informacyjnych i dostosowaniu ich do danych, które możemy otrzymać w ramach zakładanego budżetu. W ramach wspomnianych kluczowych kwestii dwa ostatnie problemy są dla nas najważniejsze – dostępność i zarazem koszt danych oraz jakość i wiarygodność tych danych.

Na koniec raz jeszcze warto sprawdzić, czy wskaźniki korespondują właściwie z celami (planowanymi wartościami efektów interwencji), czy są rozumiane i akceptowane przez decydentów, ekspertów, interesariuszy i opinię publiczną oraz czy dostarczają informacji użytecznej w procesie podejmowania decyzji lub ocenie skutków decyzji podjętych w przeszłości.

W praktycznej pracy ewaluatora pomocne są listy kryteriów oceny wskaźników, które ułatwiają ich selekcję i weryfikację. Szczególnie popularne, zwłaszcza w kontekście ewaluacji programów Unii Europejskiej, są rozmaite rozwinięcia popularnego akronimu SMART, według którego dobry wskaźnik powinien być (Sandhu-Rojon 2003: 11–12, por. European Commission 2001: 13, European Commission 2002: 54):

- (***S**pecific*) Konkretny, dokładny:
  - Czy wiadomo dokładnie, co będzie mierzone? Czy został określony odpowiedni poziom dezagregacji?

  - Czy pozwala uchwycić różnice pomiędzy obszarami czy kategoriami ludzi?
  - Czy wskaźnik jest na tyle dokładny, by zmierzyć postęp w kierunku rezultatu?
- (***M**easurable*) 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?
- (***A**ttainable*) Osiągalny:
  - Jakie zmiany są oczekiwane jako rezultaty udzielonej pomocy?
  - Czy rezultat(y) są realistyczne? Dla tego niezbędne jest wiarygodne połączenie pomiędzy produktami (*outputs*) a rezultatami (*outcomes*).
- (***R**elevant*) Istotny:
  - Czy wskaźnik oddaje istotę pożądanego rezultatu?
  - Czy jest adekwatny w stosunku do planowanych produktów i rezultatów? Aby określić znaczenie wskaźników, decydenci mogą zidentyfikować grupę docelową i jej potrzeby, oczekiwania i kryteria.
  - Czy wskaźnik jest w sposób przekonujący powiązany z daną sferą działalności?
- (***T**rackable*) Możliwy do śledzenia:
  - Czy w rzeczywistości dostępne są dane po rozsądnych kosztach i niezwiązane z większymi trudnościami?
  - Czy znamy źródła danych? Decydenci winni ustalić realistyczne zasady, mechanizmy i obowiązki w zakresie gromadzenia danych?
  - Czy wskaźnik monitoruje realizację planu?

Łatwo zauważymy, że te listy kryteriów są użytecznym rozwinięciem podstawowych wymagań, stawianych w metodologii narzędziom pomiarowym, które muszą być trafne, rzetelne i obiektywne. Inne kryteria są konsekwencją szczególnych potrzeb, wynikających z praktyki zarządzania projektowego. Niektóre zestawy analogicznych kryteriów zwracają uwagę na dodatkowe aspekty, np.: dostępność danych historycznych dla potrzeb analizy trendów i spodziewana dostępność danych w przyszłości, zdolność dostarczania wczesnych ostrzeżeń w stosunku do problemów objętych programem, czy też możliwość ustalenia wartości referencyjnych i docelowych (por. np. Indicators For Evaluation 1996). Wiele kryteriów oceny da się odczytać także z przedstawionego przez nas w punkcie 4. opisu etapów pracy nad konstrukcją zestawu wskaźników. Warto zwłaszcza zwrócić uwagę na trafność wskaźników

oraz statystyczną jakość i aktualność danych. Zespół pracujący nad zestawem wskaźników, które Unia Europejska przyjęła na szczycie w Laeken dla potrzeb Otwartej Metody Koordynacji polityk społecznych, mocno przywoływał te zasady w zestawie 10 kryteriów, które proponował przyjąć jako standard konstrukcji wskaźników efektów polityk publicznych i ich zestawów. Zasada trafności została ujęta przezeń jako postulat zdolności ujmowania istoty zjawiska oraz jasnej normatywnej interpretacji, natomiast zasada statystycznej wiarygodności wskazywała na konieczność posługiwania się danymi z badań spełniających najwyższe metodologiczne standardy (por. Atkinson et al. 2002)[9]. W praktyce trzeba czasami poświęcić niektóre z tych kryteriów na ołtarzu ograniczeń budżetowych lub limitu czasu (np. aktualność). Mimo to, warto weryfikować nasze zestawy wskaźników, korzystając z takich pomocnych list.

5. ROLA MONITORINGU I SYSTEMU SPRAWOZDAWCZOŚCI W BUDOWIE SYSTEMU WSKAŹNIKÓW

Plan ewaluacji, w tym system wskaźników, należy skoordynować z planowanym systemem monitoringu i sprawozdawczości. Pozwala to znacznie zaoszczędzić koszty. O wzajemnie uzupełniającej się roli monitoringu i ewaluacji pisze Komisja Europejska (2006). Obraz, jaki wyłania się z cytowanej publikacji, został przez nas uzupełniony o kilka uwag i istotnych kwestii z punktu widzenia budowy systemu wskaźników.

W całym dokumencie roboczym Komisji Europejskiej (2006) poświęconemu wskaźnikom, mowa jest jednocześnie o „monitoringu i ewaluacji" i o „wskaźnikach w monitoringu i ewaluacji" choć *explicite* podaje się rozróżnienie: „Monitorowanie i ewaluacja są dwoma ściśle połączonymi koncepcjami, jednakże w celu zaprojektowania właściwego systemu wskaźników należy wyraźnie zaznaczyć różnice" (ibidem: 11). Podano, że monitoring jest „badaniem osiągniętych wyników porównaniu do oczekiwań" (ibidem), a ewaluacja „wiąże się z analizą informacji pochodzących z monitoringu i innych źródeł, mając na celu określenie i wytłumaczenie efektów interwencji" (ibidem). Jeżeli przyjmiemy, że „osiągnie wyniki" i „efekt interwencji" to jedno i to samo, wówczas różnica polega tylko na tym, że monitoring dotyczy raczej produktów i może (przy pewnym wysiłku) rezultatów[10], a ewaluacja – choć uwzględnia wskaźniki produktów i rezultatów – raczej oddziaływań. Zatem w tym ujęciu monitoring obejmuje w obrębie programu

[9] Syntetyczne ujęcie tych kryteriów prezentujemy w artykule Górniak, Keler 2007.

[10] W innym miejscu nazywa je się „wskaźnikami monitoringowymi" (por. Komisja Europejska 2006: 26).

tylko zakładane do zrealizowania działania, ujęte w formie dokonanej (produktów), a w przypadku ewaluacji nacisk jest na efekty dalszego rzędu.

Warto jednak wskazać na kilka wątpliwości wobec powyższego ujęcia, przede wszystkim dotyczące analizy rezultatów interwencji. Oczywiście wskazane definicje ukazują płynne definiowanie monitoringu i ewaluacji, dlatego nie dziwi fakt, że mowa jest o „monitoringu i ewaluacji" jednocześnie. Niebezpiecznie jednak w całym rozrachunku zagubione zostają rezultaty, choć ich wskaźniki zostały uznane w innym miejscu w dokumencie za „podstawowy instrument zarządzania programem" (ibidem: 12), dosyć sceptycznie przedstawiona została też wizja możliwości ujęcia ich w monitoringu. Na rezultaty jest miejsce w ewaluacji, ale z kolei w tym przedsięwzięciu najważniejsze jest oddziaływanie. Tymczasem, aby ewaluacja miała sens, musimy powiedzieć, czy coś się zmieniło w życiu beneficjentów ostatecznych projektów w ramach programu, zanim przejdziemy do oceny oddziaływań. Jeżeli nie zachodzą zmiany w grupie docelowej beneficjentów działań, to nie można przypisać badanej interwencji wpływu na obserwowane zmiany w szerszym układzie społecznym, czyli oddziaływania. Możliwość analizy efektów jest mocno ograniczona, gdy zawczasu nie pomyślimy o tym, co jest tak naprawdę rezultatem działań, i nie uwzględnimy wskaźników rezultatu w planie ewaluacji.

Aby monitoring, w rozumieniu przyjętym powyżej, był możliwy, konieczne jest zbudowanie efektywnego systemu sprawozdawczości, w ramach którego przy minimalnym wysiłku gromadzone byłyby dane konieczne do wyliczenia wartości zakładanych wskaźników. Planując ewaluację, a w tym system wskaźników, podczas budowania i oceny planu sprawozdawczości, można zrezygnować z niektórych wskaźników, których nie można oprzeć na danych dostępnych na bieżąco, na rzecz wskaźników zastępczych (tzw. *proxies*). Może się zdarzyć, że zdobycie pewnych danych nie mieści się w ramach sprawozdawczości i wymagać będzie dodatkowych nakładów na badania. Kryterium minimalizacji kosztów jest jednym z ważniejszych przy budowie systemu wskaźników. Jeżeli tylko jesteśmy w stanie określić wskaźniki zastępcze dla tych, które pierwotnie zostały przyjęte w ramach systemu i ustalimy dostępność danych, warto rozważyć przyjęcie nieco gorszych, ale tanich wskaźników. Można ustalić kilka skorelowanych wskaźników zastępczych, by mieć kontrolę nad błędem, jaki popełniamy, stosując owe bardziej pośrednie sposoby pomiaru efektów.

Warto zwrócić krytycznie uwagę na jeszcze jeden element analizowanego dokumentu. Sugeruje on ograniczenie się w ewaluacji do wskaźników programowych i kontekstowych; inaczej mówiąc: do wskaźników ujmujących efekty zakładane oraz wskaźników ukazujących tło społeczno-gospodarcze interwencji. Natomiast w cytowanym dokumencie roboczym nie porusza się zagadnienia ujęcia przy pomocy wskaźników także

efektów niezamierzonych, choć sam problem jest dostrzegany[11]. Dodajmy przy tym: uchwycenie tych efektów jest trudne, a z racji tego, że są niezamierzone, zwykle nie antycypuje się ich przy pomocy wskaźników już na etapie tworzenia programu czy projektu. Często wiedzy na temat niezamierzonych efektów dostarczają badania jakościowe, które – gdy zaistnieje taka potrzeba – mogą zostać wykorzystane w procesie włączania do systemu dodatkowych wskaźników dla potrzeb ilościowego uchwycenia skali zjawiska i jego związków z innymi zmiennymi.

Sądzimy, że różnica między monitoringiem a ewaluacją polega jednak przede wszystkim na tym, że ten pierwszy ma przede wszystkim charakter opisowo-sprawozdawczy i informuje o postępach w realizacji projektu czy programu, podczas gdy ewaluacja nie poprzestaje na opisowym podsumowaniu programu lub jego etapu, bądź też antycypacji efektów (w przypadku ewaluacji *ex ante*), lecz stawia (raczej: powinna stawiać) pytania o związek przyczynowy pomiędzy podjętą interwencją a obserwowaną zmianą, w przypadku określonych osób, w określonej sytuacji i czasie. Jeszcze ambitniejsze jest stawianie pytań o prawomocność stosowania uzyskanych w tym zakresie odpowiedzi do innych osób, w innych okolicznościach i w przyszłości, w stosunku do której planuje się następne programy. Wskaźniki powinny dawać możliwość realizacji takich planów badań ewaluacyjnych, które starają się odpowiadać właśnie na pytania o związki przyczynowe[12].

6. BADANIE EFEKTÓW PRZY POMOCY SYSTEMÓW WSKAŹNIKÓW

System wskaźników powinien zatem umożliwić nie tylko odpowiedź na pytania opisowe (np. „jaki odsetek osób objętych szkoleniem uzyskał stałe zatrudnienie?”), lecz także badanie tzw. efektu netto, czyli odpowiedź na pytania o przyczynowy wpływ interwencji. Badanie efektu netto jest ciągle problemem, który ożywia dyskusje metodologiczne. Można jednak

[11] „W trakcie procesu tworzenia systemu wskaźników, należy pamiętać, że wskaźniki (zarówno z ich całym potencjałem, jak i ograniczeniami) są jednym z instrumentów skutecznego zarządzania. Dla przykładu, wskaźniki mogą odwrócić uwagę zarządzających programem od ważnych, lecz niezamierzonych oddziaływań, bądź od istotnych informacji o procesach. Co więcej, w przypadku pewnych przedsięwzięć (np. wsparcie B+R, innowacyjności, współpracy terytorialnej) może być szczególnie trudno zebrać wiarygodne informacje na temat rezultatów i oddziaływania. W takich przypadkach, zbieranie informacji związanych z zachodzącymi procesami, zwłaszcza dotyczących rozwoju potencjału i kompetencji osób zaangażowanych w realizację programu, może być instrumentem wspierającym w procesie zarządzania i raportowania” (Komisja Europejska 2006: 19).

[12] Taki sposób rozumienia ewaluacji należy do tradycji reprezentowanej przez takie postaci jak: Donald T. Campbell, Lee Cronbach, William R. Shadish, Thomas D. Cook, a także, w znacznej mierze Carol H. Weiss czy Peter Rossi, żeby wspomnieć tylko „klasyków gatunku”.

wskazać kilka dobrych praktyk w tym zakresie. Jedną z popularniejszych technik jest dobór prób kontrolnych metodą dopasowania według przewidywanych prawdopodobieństw (*propensity score matching*)[13]. Jest to złożone zagadnienie, wykraczające dalece poza ramy niniejszego tekstu. Tu warto podkreślić, że czynnikiem ułatwiającym badanie efektów przyczynowych jest dobre zintegrowanie zestawu wskaźników z systemem monitoringu i sprawozdawczości.

W szacowaniu oddziaływania/wpływu programu mogą pomóc wspomniane wskaźniki kontekstowe, które pokazują szerszy zakres zmian, jakie zachodzą w społeczeństwie (np. demograficzne czy ekonomiczne). Można interpretować zmianę uchwyconą przez porównanie wartości wskaźników oddziaływania przed dokonaniem interwencji i po jej ukończeniu jako wyraz oddziaływania programu. Jednak to przede wszystkim ocena na poziomie rezultatów daje realną szansę na przełożenie wniosków z ewaluacji na użyteczne rekomendacje dla przyszłych interwencji tego samego rodzaju, prowadzonych w zbliżonym kontekście instytucjonalnym. Przede wszystkim umożliwia wyeliminowanie z repertuaru programów publicznych działań ewidentnie nieskutecznych, co pozwala poprawić proces programowania. Warto pamiętać, że skuteczność na poziomie rezultatu nie gwarantuje osiągnięcia celów strategicznych odzwierciedlanych wskaźnikami oddziaływania/wpływu. Natomiast, brak skuteczności na poziomie rezultatów gwarantuje, że ewentualna zmiana wartości wskaźników oddziaływania jest efektem innych czynników niż program. To, czy rezultaty przerodzą się w szersze oddziaływania, zależy między innymi od stopnia, w jakim występują niezamierzone i niepożądane efekty uboczne.

7. OGRANICZENIA I ZAGROŻENIA ZWIĄZANE Z SYSTEMAMI WSKAŹNIKÓW

Zarówno konieczność budowy i użycia systemów wskaźników, jak i płynące z tego (potencjalne) korzyści nie powinny spowodować bezkrytyczności wśród ewaluatorów i osób planujących ewaluację. Niespełnienie założeń prezentowanych powyżej może prowadzić do wadliwych systemów, które przynieść mogą tylko negatywne konsekwencje.

Nie należy zapominać, że wskaźniki mają przede wszystkim odgrywać swoją rolę, mianowicie pozwolić ująć efekty interwencji. Wskaźniki dobrać można tylko wtedy, gdy wiemy, jakie są cele interwencji, i – najlepiej – gdy potrafimy wskazać obszary, na których mogą pojawić się efekty niezamierzone. Gdy efekty są niedookreślone, systemy wskaźników

[13] Na temat tej metody można przeczytać np. w Konarski, Kotnarowski (2007) czy Bryson, Dorsett, Purdon (2002).

mogą być tylko przykrywką (symulująca precyzję) dla nieprzemyślanych i źle zaplanowanych działań. Nierzadko namysł nad wskaźnikami pozwoli na lepsze dopracowanie celów, ale czerpać korzyści z tego procesu można tylko wówczas, gdy wskaźniki dobieramy jeszcze podczas okresu planowania, a nie wtedy, gdy żadnych zmian dokonać nie można.

Warto w tym miejscu mocno podkreślić różnicę między celami programu (*goals*) i ich operacyjnym ujęciem (*objectives*), a wartościami docelowymi wskaźników (*targets*). Tym, co jest najistotniejsze, jest osiąganie celów. To, czy cele zostały osiągnięte, mierzone jest przy pomocy wskaźników, których wartości docelowe możemy ustalić konkretnie lub wskazać tylko kierunek pożądanej zmiany. Nigdy jednak celem interwencji nie jest osiągnięcie wartości wskaźników, a zmiana społeczna, która jest lepiej lub gorzej mierzona dzięki wskaźnikom. Ustalenie wartości docelowej może mieć nawet negatywne konsekwencje, gdy „rozliczenie", „ocena" działania będzie wiązała się tylko ze sprawdzeniem, czy wartości docelowe zostały osiągnięte. Taka postawa będzie wymuszać na projektodawcach zaniżanie aspiracji w stosunku do wartości docelowych, do „bezpiecznych wartości", a tymczasem cele podejmowanych działań muszą być ambitne, jeżeli mają rzeczywiście przynieść jakieś zmiany. Z kolei w ewaluacji nie możemy koncentrować się tylko na wskaźnikach dotyczących celów, czyli efektów zamierzonych i pożądanych, lecz powinniśmy dokonać także analizy efektów niezamierzonych, na tyle, na ile jest to tylko możliwe. To miał na myśli Michael Scriven (1972), postulując ewaluację wolną od celów (*goal-free evaluation*), co raczej lepiej można nazwać ewaluacją wykraczającą poza cele.

8. ZAKOŃCZENIE

Praca nad doborem wskaźników do programów czy projektów ma bardzo istotne znaczenie. Wskaźniki bywają traktowane jako zło konieczne, czysto biurokratyczny wymysł. Tymczasem, jak pokazaliśmy, służą one dobrze samemu planowaniu interwencji, pozwalając sprecyzować i skwantyfikować cele, odnieść planowane i osiągane efekty do nakładów, a w trakcie realizacji przedsięwzięć kontrolować postępy i dbać, by program czy projekt nie zboczył z wytyczonej ścieżki. Pozwalają także badać związki przyczynowe pomiędzy podejmowanymi działaniami a ich efektami. Po to, by dobrze definiować wskaźniki, trzeba przede wszystkim dobrze rozumieć cele i kontekst interwencji. W przypadku bardziej zaawansowanych konstrukcji potrzebna jest wiedza metodologiczna i statystyczna. Duża rola wskaźników i stopniowa specjalizacja profesjonalnej aktywności w tym obszarze znalazły swoje odzwierciedlenie m.in. w fakcie powołania do życia w 2004 roku otwartej sieci pod nazwą: *The European Network of Indicator Designers*. Komisja Europejska zaplanowała odrębne działanie poświęcone wskaźnikom

w politykach publicznych (8.6) w ramach roboczego programu dla nauk społecznych i humanistycznych 7. Programu Ramowego. Ten wzrost znaczenia wskaźników jest skorelowany ze wzrostem znaczenia ewaluacji w doskonaleniu programów i kreowaniu przejrzystości i rozliczalności działań w sferze polityk publicznych. Jest też jednym z przejawów procesu racjonalizacji i rozwoju społeczeństwa opartego na wiedzy. Dlatego wskaźników nie należy traktować jako zła koniecznego, lecz jako okazję do tego, by nasze przedsięwzięcia były lepiej zarządzane, usprawniane i coraz skuteczniejsze. A w procesie ewaluacji nie należy rezygnować z jej kluczowej funkcji, jaką jest pomiar efektów i rozliczenie programów. Nie można też zapomnieć o istotnej roli wskaźników w realizacji tej funkcji. Jak to zostało podkreślone w innym miejscu, przy okazji krytyki podejścia, jakie propagują Guba i Lincoln – tzw. ewaluacji czwartej generacji – podejście do ewaluacji, które rezygnuje z aspiracji do obiektywizmu i pomiaru efektów, oparte jest na idealizacji społecznego kontekstu ewaluacji i nie docenia pożytków płynących z odrzucanych przez siebie, „niższych" generacji. Zapewne bardziej nadaje się do ewaluacji poszczególnych projektów o mniejszej skali, a zwłaszcza samoewaluacji, natomiast jej przydatność staje się ograniczona w przypadku złożonych programów i polityk, w tym regionalnych. Nie sposób omawiać tu rozlicznych publikacji proponentów tego nowego nurtu, ale przypominają one spór zarysowany przez Mickiewicza w *Romantyczności* – pomiędzy zorientowanymi na działanie romantykami a „mędrca szkiełkiem i okiem". Szkoła konstruktywistyczna traktuje ewaluację jako *action research*, wręcz jako formę działania, które jest stałym ulepszaniem prowadzonych projektów w kooperacji, najlepiej harmonijnej, wszystkich zainteresowanych stron. Na pewno z jednej strony wartością tego podejścia jest zwrócenie uwagi na dobre strony pragmatycznie traktowanej partycypacji oraz na walory wynikające z jakościowego rozeznania wszystkich ważnych aspektów wdrożenia programów. Z drugiej strony, zaciera się w nim właściwa rola ewaluacji, którą jest weryfikacja trafności podjętych działań i wykorzystania zasobów z perspektywy osiągniętych efektów. Mieszają się funkcje ewaluatora z funkcjami konsultanta, a nawet (nieformalnego) przywódcy (Górniak 2007).

Metodyka ewaluacji przyjęta przez Komisję Europejską nie podąża za nowinkami „czwartej generacji", stojąc na fundamencie pragmatycznie zorientowanej ewaluacji, w której pomiar efektów przy pomocy wskaźników jest punktem wyjścia, choć nie końcem procesu ewaluacji. W procesie doskonalenia programów i eliminacji z repertuaru działań interwencji nieskutecznych trafny, rzetelny i obiektywny pomiar zjawisk jest ciągle „w modzie". Rośnie też znaczenie procedur badawczo-analitycznych, testujących przyczynowe związki między działaniami a efektami. Polityki publiczne zyskują stopniowo porządny naukowy, metodologiczny fundament.

BIBLIOGRAFIA

Atkinson T., Cantillon B., Marlier E., Nolan B. (2002), *Social Indicators. The EU Social Inclusion*, New York, Oxford University Press.

Birckmayer J.D., Weiss C.H. (2000), Theory-Based Evaluation in Practice. What Do We Learn?, *Evaluation Review* **24**, 407–431.

Booysen F. (2002), An Overview and Evaluation of Composite Indices of Development, *Social Indicators Research* **59**, 115–151.

Bryson A., Dorsett R., Purdon S. (2002), *The Use Of Propensity Score Matching In The, Evaluation Of Active Labour Market, Policies*, Working Paper Number 4, www.dwp.gov.uk/asd/asd5/WP4.pdf [30.06.2008].

Chen H.-T. (2004), *The Roots of Theory-Driven Evaluation*, w: Alkin M.C. (ed.) (2004), *Evaluation Roots: Tracing Theorists' Views and Influences*, London, New Delhi, Thousand Oaks, Sage.

European Commission (2001), *Ex Ante Evaluation: A Practical Guide for Preparing Proposals for Expenditure Programmes,* Brussels: European Commission, Budget, Own Resources, evaluation and financial programming Evaluation, http://ec.europa.eu/budget/library/documents/evaluation/guides/ex_ante_guide_2001_en.pdf [18.08.2008].

European Commission (2002), *Project Cycle Management Handbook*, EuropeAid Co-operation Office General Affairs Evaluation, http://www.stgm.org.tr/docs/1123450143PCM_Train_Handbook_EN-March2002.pdf [18.08.2008], wersja polska podręcznika: http://www.equal.gov.pl/NR/rdonlyres/3A81D042-E1A4-42A6-8A31-6D587DDCB8D9/0/equal_podr_zarz_cykl.pdf [18.08.2008].

European Commission (2005), *Continuity of indicators between end-ECHP and start-SILC. Algorithms to compute cross-sectional indicators of poverty and social inclusion adopted under the open method of coordination*, Eurostat, Directorate D: Single Market, Employment and Social statistics, Unit D-2: Living conditions and social protection, http://www.insee.fr/en/insee-statistique-publique/colloques/pauvrete/pdf/Eurostat_social_indicators_EN.pdf [18.08.2008].

Ferguson G.A., Takane Y. (2007), *Analiza statystyczna w psychologii i pedagogice*, Warszawa, WN PWN.

Frączek M., Górniak J., Keler K., Laurisz N., Małodzińska-Strzebońska A., Mazur S., Perek-Białas J., Worek B. (2007a), *Wybrane wskaźniki wykluczenia społecznego i sytuacji na rynku pracy. Definicje, sposób liczenia, zastosowanie*, Kraków, Małopolska Szkoła Administracji Publicznej Akademii Ekonomicznej w Krakowie (dostępne także w postaci elektronicznej: http://www.rynekpracy.msap.pl/ftp/Produkty/wskazniki_publ.pdf).

Frączek M., Górniak J., Keler K., Laurisz N., Małodzińska-Strzebońska A., Mazur S., Perek-Białas J., Worek B. (2007b), *Diagnoza rynku pracy i wykluczenia społecznego w Małopolsce. Refleksja nad wykorzystaniem wskaźników*, Kraków: Małopolska Szkoła Administracji Publicznej Akademii Ekonomicznej w Krakowie (dostępne także w postaci elektronicznej: http://www.rynekpracy.msap.pl/ftp/Produkty/diagnoza%20rynku%20net.pdf).

Giovannini E., Nardo M., Saisana M., Saltelli A., Tarantola S., Hoffman A. (2005), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, http://www.olis.oecd.org/olis/2005doc.nsf/LinkTo/NT00002E-4E/$FILE/JT00188147.PDF [30.06.2008].

Global Water Partnership Technical Brief 3 (b.d.) *Monitoring and evaluation indicators for IWRM strategies and plans*, http://www.gwpforum.org/gwp/library/Tec_brief_3_Monitoring.pdf [30.06.2008].

Górniak J. (2007), Ewaluacja w cyklu polityk publicznych, w: Mazur S. (red.), *Ewaluacja funduszy strukturalnych. Perspektywa regionalna*, Kraków, MSAP, Akademia Ekonomiczna w Krakowie.

Górniak J., Keler K. (2007), Wskaźniki w ewaluacji *ex post* programów publicznych, w: Haber A. (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Warszawa, Polska Agencja Rozwoju Przedsiębiorstw.

Górniak J., Keler K. (2008), Ewaluacja i wskaźniki w cyklu planowania i realizacji polityk publicznych, *Zarządzanie publiczne* **1(3)**, 109–124.

Indicators For Evaluation Task Force Of The International Joint Commission (1996), Indicators to Evaluate Progress under the Great Lakes Water Quality Agreement, http://www.ijc.org/php/publications/html/ietf.html [30.06.2008].

Komisja Europejska (2006), *Nowy okres programowania 2007–2013. Przewodnik po metodach ewaluacji: Wskaźniki monitoringu i ewaluacji*, Dokument roboczy nr 2, http://www.funduszestrukturalne.gov.pl/NR/rdonlyres/C0B38DF6-E998-47CF-BA74-A1316B8B794E/35538/wskazniki170707.pdf [30.06.2008].

Konarski R., Kotnarowski M. (2007), Zastosowanie Metody *Propensity Score matching* w ewaluacji *ex-post*, w: *Ewaluacjia ex post. Teoria i praktyka badawcza* pod redakcją A. Haber, Warszawa, Polska Agencja Rozwoju Przedsiębiorstw.

Langbein L., Felbinger C.L. (2006), *Public Program Evaluation. A Statistical Guide*, Armonk, London, M.E. Sharpe.

Murray Ch. (1988), *In Pursuit of Happiness and Good Government*, New York, London, Toronto, Sydney, Tokyo, Simon and Schuster.

Nowak S. (1985), *Metodologia badań społecznych*, Warszawa, PWN.

Purdon S., Lessof C., Woodfield K., Bryson C. (2001), *Research Methods for Policy Evaluation, Department for Work and Pensions*, Research Working Paper No 2, www.dwp.gov.uk/asd/asd5/WP2.pdf [30.06.2008].

Salzman J. (2003), Methodological Choices Encountered In the Construction of Composite Indices of Economic and Social Well-Being, Center for the Study of Living Standards, http://www.csls.ca/events/cea2003/salzman-typolcea2003.pdf [30.06.2008].

Sandhu-Rojon R. (2003), *Selecting Indicators for impact evaluation*, New York, UNDP.

Scriven M. (1972), *Pros and Cons About Goal-Free Evaluation*, w: Thomas L.G. (ed.), *Philosophical Redirection of Educational Research: The Seventy-First Yearbook of the National Society for the Study of Education*, Chicago, University of Chicago Press.

ROZDZIAŁ 2.3.

Sławomir Mandes

METODY JAKOŚCIOWE W EWALUACJI

Każda próba prezentacji metod jakościowych napotyka już na wstępie trudności w zdefiniowaniu, czym są metody jakościowe. Tylko pozornie problem ten wydaje się mało istotny. Utarło się, że metody badawcze w naukach społecznych dzieli się na ilościowe i jakościowe. Te pierwsze opierają się na spójnej i ścisłej teorii statystycznej oraz nieco mniej spójnych zasadach konstruowania podstawowego narzędzia tej metody, jakim jest ankieta. Metody jakościowe nie dają się już scharakteryzować w tak prosty i jednoznaczny sposób. Obejmują one w swoim zakresie takie metody jak: analiza treści, wywiady, obserwacja uczestnicząca i wiele innych. Tradycyjny podział miał do niedawna proste uzasadnienie: zasadniczo dane zebrane za pomocą metod jakościowych nie są policzalne. Jednak od pewnego czasu to się zmienia i w coraz większym stopniu wykorzystywane są elementy metodologii ilościowej do różnych metod jakościowych.

Sposób zbierania danych w naukach społecznych za pomocą metod jakościowych jest starszy od metod ilościowych. Klasyczne dzieła Herberta Spencera czy Jamesa Fresera powstały na podstawie pamiętników, listów, artykułów, biurokratycznych dokumentów i informacji od podróżników. Znajdziemy w nich typowe wykorzystanie analizy dokumentów o dużym rygorze metodologicznym.

Duży wkład w rozwój tych metod mieli Polacy. Florian Znaniecki wraz Williamem Thomasem w *Chłopie polskim* (1976) opierają się na analizie listów. Bronisław Malinowski stworzył nowoczesną antropologię, pracując metodą badania terenowego, co dziś uznalibyśmy za obserwację uczestniczącą, połączoną z wywiadami pogłębionymi.

Charakterystyczne dla wskazanych przykładów jest to, że autorzy ci, prowadząc swoje badania, koncentrowali się na problemach i nie myśleli

o kwestiach metodologicznych. Zastanawiali się oczywiście nad tym, jak zbierają materiały, czy są one obiektywnym odbiciem rzeczywistości, czy zdobyta ten sposób wiedza wytrzyma konstruktywną krytykę. Nie zastanawiali się jednak, czy ich „metoda" daje się uzasadnić z epistemologicznego punktu widzenia. Namysł nad prawomocnością naukową metod i pojawienie się czegoś, co współcześnie nazywa się metodologią, jest efektem rozwoju w pierwszej połowie XX wieku kierunku filozoficznego zwanego pozytywizmem logicznym. W dużym uproszczeniu, jego reprezentanci twierdzili, że kryterium weryfikującym prawdziwość naszej wiedzy o świecie jest metoda, która powinna mieć charakter naukowy. Inaczej mówiąc, jeśli jakaś metoda nie daje się obronić z punktu widzenia logiki naukowej, nie ma dobrego uzasadnienia epistemologicznego, wówczas zebrane przez nią dane i oparta na nich wiedza są bezwartościowe.

Metody, które przeszły taki test w naukach społecznych, zaczęto rozwijać w latach 40. Ich prekursorem był wybitny socjolog i metodolog Paul Lazersfeld. Z punktu widzenia pozytywizmu metody ilościowe są dobrze ugruntowane, ponieważ opierają się na teorii wnioskowania statystycznego. Umożliwiają badanie i wnioskowanie o zjawiskach społecznych w kategoriach policzalnych, zgodnie z logiką naukową. Pozostałe metody mają charakter nienaukowy, pozwalają zebrać wprawdzie jakąś wiedzę, ale jej status jest prowizoryczny i do czasu weryfikacji właściwymi metodami nie można jej uznać za prawdę naukową.

Pozytywizm dokonał rewolucji w naukach społecznych (i nie tylko), odwracając wcześniejszy sposób uprawiania nauki. Nie liczyło się teraz „co" (empiryczna treść badania), ale „jak" (zgodność wykorzystanych metod z epistemologicznymi kryteriami). Tradycyjny podział na metody ilościowe i jakościowe jest czymś więcej niż wygodnym, konwencjonalnym – z punktu widzenia praktyki nauczania – rozróżnieniem (por. Becker 1996). Pomiędzy oboma rodzajami metod istnieje zasadnicza różnica, mająca istotne znaczenie z punktu widzenia ich wykorzystania. Można oczywiście spytać, po co zaprzątać sobie głowę rozważaniami nad epistemologicznymi różnicami pomiędzy metodami. Jest to ważne ze względu na współistnienie obu rodzajów metod. Od dłuższego czasu w naukach społecznych w ogóle, a w ewaluacji szczególnie, stosuje się równocześnie oba rodzaje metod. Zjawisko to jest odzwierciedleniem głębokich zmian w porównaniu do czasów, kiedy metody jakościowe były uprawiane przez marginalną i postrzeganą jako ekscentryczną grupę badaczy, którzy uparcie odmawiali uznania pozytywistycznego kanonu uprawiania nauk społecznych[1]. Były to czasy, kiedy oba obozy praktycznie się nie przenikały, a opowiedzenie się za jedną z metod było jednocześnie deklaracją postrzegania świata społecznego. Od tego czasu wiele się zmieniło.

---

[1] W polskiej socjologii metody jakościowe dopiero w drugiej połowie lat 80. zostały formalnie uznane za metodę badawczą.

Współcześnie powszechnie postuluje się zasadę triangulacji, czyli wykorzystywania różnych metod, tak aby uzyskać obraz danego zjawiska z różnych perspektyw (por. Mason 2006). Jest to postulat słuszny i godny propagowania. Warto jednak jeszcze wytłumaczyć, na czym polegają i skąd się biorą owe różne perspektywy. Pozwoli lepiej zrozumieć, czym są metody jakościowe jako odrębna grupa technik badawczych wykorzystywanych w ewaluacji. Dlatego artykuł ten zacznie się od omówienia specyfiki metod jakościowych i usytuowania ich w kontekście badań ilościowych. Następnie omówię poszczególne metody, aby pokazać, na czym polegają różnice pomiędzy nimi, jakie są ich wady oraz zalety. Na tej podstawie sformułuję ogólną zasadę określającą zakres wykorzystania tych metod oraz interpretacji wyników uzyskanych za ich pomocą. W kolejnej części artykułu przedstawię sposób wykorzystania metod jakościowych w ewaluacji. Na koniec poruszę często pojawiający się w kontekście metod jakościowych problem doboru próby.

1. DO CZEGO SŁUŻĄ METODY JAKOŚCIOWE I CZYM RÓŻNIĄ SIĘ OD METOD ILOŚCIOWYCH

Użyteczność metod jakościowych była dość długo kwestionowana w ewaluacji. Poza wspomnianym na początku wpływem pozytywizmu na nauki społeczne, ważną rolę w przypadku ewaluacji odgrywał fakt, iż jedne z pierwszych badań w tym zakresie były przeprowadzone przez psychologa Edwarda Thorndike'a, który, aby uzyskać odpowiedź na pytanie, co się sprawdza w edukacji, sięgnął do sprawdzonych i znanych sobie ilościowych metod eksperymentalnych. Podejście to zyskało wielu naśladowców i przez długi czas ewaluacje miały prawie wyłącznie charakter sumatywny i quasi-eksperymentalny; koncentrowały się na wynikach danego programu, aby na podstawie założonych celów dokonać oceny skuteczności przez ilościowe wskaźniki zebrane w grupach, które brały udział w programie (grupa eksperymentalna) i takich, które były poza programem (grupa kontrolna). Pomijano natomiast kwestię *jakości* całego procesu, tzn. problem **znaczenia**, jaki dany proces miał dla uczestników.

Z punktu widzenia metod eksperymentalnych nic nie stało na przeszkodzie, aby uzupełnić badanie o element jakościowy – jeśli chcemy jeszcze wiedzieć, co myślą uczestnicy danego programu, to możemy ich po prostu spytać. Podkreślmy, że z tego punktu widzenia nie ma to żadnego znaczenia dla oceny skuteczności programu. Nie jest istotne, czy uczestnicy danego projektu, zmierzającego na przykład do poprawy poziomu nauczania, byli z niego zadowoleni. Ważne jest, czy program osiągnął swój cel, na przykład – czy przyczynił się do poprawy ocen uczniów. W efekcie sprowadza się znaczenie, jaki proces nauczania

miał dla uczestników, do dodatkowego elementu mającego charakter ozdobnika. Postrzeganie ewaluacji jakościowej jako dodatku do ewaluacji ilościowej jest oparte na pewnym nieporozumieniu, któremu warto przyjrzeć się bliżej, albowiem praktyka pokazuje, że jest ono wciąż dosyć rozpowszechnione. Nieporozumienie to polega na myśleniu o jakościowych aspektach programu jako wtórnych elementach bez większego znaczenia dla jego przebiegu[2].

Przykładem, ilustrującym różnicę pomiędzy dającymi się zmierzyć ilościowo i jakościowo aspektami projektu, może być klasyczny eksperyment Rollo Maya. May miał za zadanie zidentyfikować czynniki odpowiedzialne za efektywność pracy w pewnej fabryce. Zabrał się do tego zgodnie z wszelkimi kanonami metodologicznymi: odizolował wybraną losowo grupę kobiet i zaczął manipulować różnymi czynnikami, takimi jak oświetlenie, temperatura, liczba przerw, długość. Zmieniając te i inne elementy środowiska pracy, wciąż nie mógł znaleźć odpowiedzi na postawione mu pytanie: co zrobić, aby robotnice pracowały wydajniej? Okazywało się bowiem, że wydajność wprawdzie zmieniała się, ale nie bardzo było wiadomo, co miało na to wpływ. Dopiero wykroczenie poza utarty sposób myślenia podsunęło Mayowi odpowiedź, która z punktu widzenia współczesnej teorii zarządzania jest banałem: zmiany były uzależnione od motywacji robotnic. Początkowy wzrost produktywności wynikał z tego, że robotnice poczuły się dowartościowane uwagą, jaką im poświęcano. Następnie, gdy sytuacja została oswojona, ich zapał do pracy osłabł.

Eksperyment Maya ilustruje bardzo dobrze pewne nieścisłości dotyczące zjawisk społecznych określanych jako jakościowe. Po pierwsze, dla wyjaśnienia przyczyn zachodzenia pewnych zjawisk, nie wystarczają dane ilościowe i trzeba odwołać się do danych, które można zebrać tylko metodami jakościowymi. Innymi słowy, jakościowe elementy procesu nie są dodatkiem, lecz podstawą dla wyjaśnienia, dlaczego ludzie postępują tak, a nie inaczej. Po trzecie, kontekst, w ramach którego dany proces ma miejsce, może być bardzo ważny dla wyjaśnienia tego, co zaszło.

Jakie ma to konsekwencje dla wykorzystania metod jakościowych w ewaluacji? Schematycznie rzecz ujmując, każdy program składa się z trzech elementów: założeń wyjściowych, realizowanych działań (tj.: szkolenia, warsztaty, konferencję itd.) oraz skutków (efektów zakładanych i niespodziewanych). Przed ewaluatorem stają dwa pytania, które można sformułować wobec każdego z trzech elementów programu: 1) co się zdarzyło? 2) dlaczego coś się zdarzyło? Pierwsze pytanie stawia przed ewaluatorem zadanie **opisania** założeń lub realizowanych działań lub

[2] Warto zauważyć, że na tym właśnie polegał pozytywizm w naukach społecznych. Eliminował on znaczenie, jakie dany proces społeczny miał dla jego uczestników jako nieistotne (subiektywne) dla wyjaśnienia jego przebiegu.

skutków lub wszystkich razem. Drugie ma prowadzić do **wyjaśnienia**, dlaczego miały one taką, a nie inną, rzeczywistą postać. Należy też pamiętać o kontekście społecznym, w którym dany program jest realizowany.

Posłużmy się przykładem programu przygotowanego dla bezrobotnych[3]. Na podstawie wcześniej zdobytej wiedzy założono, że bezrobotnym brakuje tzw. miękkich umiejętności, tj.: napisania CV, wyszukiwania pracy, autoprezentacji, odpowiedniej motywacji. Na podstawie tych założeń przygotowano program szkoleń, warsztatów i praktyk. Po zakończeniu programu monitorowano skutki, czyli poziom zatrudnienia wśród beneficjentów programu.

Najczęściej stawianym przed ewaluatorem zadaniem jest opisanie skutków. Aby to zrobić, wystarczy na reprezentatywnej próbie beneficjentów przeprowadzić badanie ankietowe. Metoda pozwoli na uogólnienie zebranych wyników na cały zbiór (populację) beneficjentów, dzięki czemu w postaci procentowej będzie można uzyskać odpowiedzi na różne, dające się wyrazić ilościowo, pytania, np.: ile osób znalazło zatrudnienie, a ile nie.

Zadanie ewaluatora może też polegać na opisaniu elementów, czyli pokazaniu, ile szkoleń się odbyło, czy bezrobotni biorący w nich udział są zadowoleni. Również w tym wypadku można posłużyć się metodologią ilościową, na przykład ankietą audytoryjną, która w sposób całościowy i systematyczny pozwoli zebrać dane potrzebne do takiego zadania. Pomocne w tym wypadku mogą okazać się wybrane metody jakościowe, takie jak wywiady pogłębione, aby uzyskać wiedzę pozwalającą na lepszą interpretację wyników ilościowych. Nie są one jednak konieczne i nie wniosą niczego istotnego.

Ewaluator może też chcieć się dowiedzieć, z czego wynikały przesłanki, na których program się opierał. Aby taką wiedzę zdobyć, w większości wypadków wystarczająca jest analiza dokumentów uzupełniona o wywiady.

Przed ewaluatorem może być jednak postawione ambitniejsze zadanie[4]. Zleceniodawca może chcieć wiedzieć **dlaczego** osiągnięte skutki różniły się od planowanych. Jakie były **przyczyny** pewnych zjawisk? Dlaczego, na przykład, liczba zatrudnionych beneficjentów była mniejsza niż zakładano? Dlaczego krótko po zatrudnieniu beneficjenci tracili pracę? Dlaczego pewne szkolenia były źle oceniane? Metodologicznie poprawna – z punktu widzenia metod ilościowych – procedura powinna

---

[3] Poniżej przedstawiona analiza ma charakter przykładowy i służy jedynie ilustracji ewentualnych sposobów wykorzystywania obu rodzajów metod.

[4] Przedstawione w następnych akapitach tekstu argumenty na rzecz metod jakościowych są nieco uproszczoną wersją toczących się obecnie dyskusji w naukach społecznych i w ewaluacji o których więcej można znaleźć w pracach: Maxwell 2004a, 2004b; Pawson; Tilley 1997; Shaw 1999.

wzorować się na modelu eksperymentalnym. Na samym początku powinno się losowo podzielić bezrobotnych na dwie grupy: jedną włączyć do programu, drugą pozostawić samą sobie. Przed rozpoczęciem całego programu obie grupy powinny wypełnić szczegółową ankietę, mierzącą różne zmienne, które ewentualnie mogłyby wpływać na ostateczny efekt. Na koniec programu, obie grupy ponownie podlegałyby temu samemu badaniu ankietowemu, następnie wyniki zostałyby porównane. Taka procedura niesie za sobą przynajmniej dwa problemy: po pierwsze, nie są tu zachowane wszystkie standardy badania eksperymentalnego, które pozwalałyby na wnioskowanie o zależnościach przyczynowych (por. Sułek 1979).

Według właściwej procedury oddziaływanie eksperymentalne (w tym wypadku: program) powinno mieć postać pojedynczego oddziaływania. W tym przypadku mamy do czynienia z szeregiem złożonych czynników, jakie każdy program w sobie zawiera. Oddziaływaniem może być na przykład jedno szkolenie. Poza tym eksperyment zakłada pełną kontrolę innych czynników, która wymaga izolacji badanych na czas trwania całej procedury. Eksperyment Mayo pokazuje sytuację, w której eksperyment nie uwzględniał wszystkich czynników kontekstowych mających znaczenie dla ostatecznego wyniku. W przykładzie z programem dla bezrobotnych, badacze z konieczności zmuszeni byliby pominąć fakt, że możliwość zatrudnienia zależy od sytuacji na rynku pracy, czyli od kontekstu społecznego, który nie jest uwzględniany w sytuacji eksperymentalnej. Z badania możemy uzyskać wynik, w którym w obu grupach badanych odsetek zatrudnionych jest ten sam. Możemy wówczas przypuszczać, że za ten stan rzeczy odpowiada sytuacja na rynku. Nie możemy jednak stwierdzić tego z całą pewnością. Może być również tak, że przyczyna leży w samym programie: niedostosowaniu do potrzeb, nieadekwatności użytych metod itd. Kolejnym problemem są wysokie koszty badań eksperymentalnych, niewspółmiernie do wiedzy, jaką za pomocą tej metody moglibyśmy uzyskać.

W szukaniu odpowiedzi na pytanie o determinanty danych zjawisk nie pomogą nam również badania na próbach. Częstym błędem popełnianym przy okazji interpretacji danych ilościowych jest przekonanie, że wykryte zależności statystyczne świadczą o rzeczywistych zależnościach i dają podstawę do odpowiedzi na pytania zaczynających się od: „dlaczego". Najprościej jest to wyjaśnić na przykładzie. Załóżmy, że analiza pokazała istnienie korelacji pomiędzy wykształceniem, satysfakcją ze szkoleń a zatrudnieniem. Oznacza to, że im wyższe wykształcenie, tym częściej osoba znajdowała zatrudnienie, oraz im wyższa satysfakcja ze szkolenia, tym częściej osoba znajdowała zatrudnienie. Obie zależności są zdroworozsądkowe i łatwe do wytłumaczenia na podstawie codziennej wiedzy. Błędne jest jednak przekonanie, że są one dowodem na istnienie realnych zależności przyczynowych. Korelację (i inne metody

analizy statystycznej oparte na analizie wariancji) pokazują jedynie **współwystępowanie** danych zmiennych.

W praktyce badawczej mamy najczęściej do czynienia z zależnościami, o których wiemy, że są powiązane relacją przyczynowo-skutkową. Wiedza ta jednak pochodzi nie z wykrytych zależności, ale codziennego doświadczenia. Łatwo można znaleźć zależności, o których nikt nie pomyśli, że mają związek przycznowo-skutkowy. Na przykład: liczba bocianów jest związana z dzietnością. Nie oznacza to jednak, że wzrost liczby bocianów pociągać za sobą będzie większy przyrost naturalny. Nie ma tu żadnej zależności, jedynie współwystępowanie. Korelacja taka daje się łatwo wytłumaczyć większą dzietnością na wsi, która ma wiele różnych przyczyn niezależnych od liczby bocianów. Wspomniany przykład z wykształceniem wyraża pewną rzeczywistą zależność edukacji i szansy znalezienia pracy. Mówimy, że wyższe wykształcenie zwiększa szansę na znalezienie pracy. To jednak nie wykształcenie o tym przesądza. Gdyby tak było, nie powinno być od tej reguły wyjątków. Szansę na znalezienie pracy zależą od wielu czynników społecznych i psychologicznych, które są ze sobą powiązane w niezwykle skomplikowany sposób. Regularność obserwowana w danych ilościowych nie wyraża wcale rzeczywistych przyczyn zachodzenia danych zjawisk. Aby poznać przyczyny, musimy sięgnąć do metod jakościowych.

Różnicę pomiędzy oboma rodzajami metod wyraża dobrze rozróżnienie wprowadzone przez Lawrence'a Mohra (1995, 1996, 1999) na teorie wariancji (*variance thory*) i teorie procesu (*process theory*). Teorie wariancji opierają się na zmiennych i korelacji pomiędzy nimi. Polegają na analizie zmienności poszczególnych zmiennych w zależności od zmienności innych. Wzorcowym modelem takiej analizy jest liniowa analiza regresji. Teorie wariancji bazują na metodologii ilościowej, wnioskowaniu statystycznym i modelu eksperymentu.

Teorie procesu natomiast interesują się zdarzeniami (*events*) i procesami, które łączą je w pewne całości. Celem teorii procesu jest wyjaśnienie zależności przyczynowych pomiędzy poszczególnymi zdarzeniami, które prowadzą do zachodzenia pewnych ciągów zdarzeń. Teorie tego typu ujmują proces jako coś specyficznego i dlatego są sceptyczne wobec podejścia ilościowego. Bazując na pogłębionych studiach przypadków oraz niewielkich próbach, dążą do rekonstrukcji rzeczywistych przyczyn przebiegu zdarzeń.

Z punktu widzenia takiego podziału, ewaluowany program jest pewnym zdarzeniem o względnie niepowtarzalnym charakterze. Szukając odpowiedzi na pytanie o przyczynę takiego, a nie innego przebiegu danego programu, ewaluator powinien bazować na metodach ilościowych, dających mu wgląd w rzeczywiste przyczyny tworzących go zdarzeń. Tylko one dają wgląd w rzeczywiste zależności pomiędzy różnymi zdarzeniami.

2. CO EWALUATOR MA DO WYBORU? KRÓTKI PRZEGLĄD WAD I ZALET WYBRANYCH METOD JAKOŚCIOWYCH

Metody jakościowe są narzędziami badania rzeczywistości społecznej. Jak każde narzędzie, metody badawcze można wykorzystać źle lub dobrze. Z gruntu błędne są przekonania, że jakaś metoda jest lepsza lub gorsza od innej. Nie tyle metody są złe lub dobre, to ich zastosowania mogą prowadzić nas do trafnych lub błędnych wniosków o rzeczywistości.

Jedna z trudności badań polega na dobraniu odpowiedniej metodologii do przedmiotu badania i odpowiedniego postawienia pytań, na które chcemy odpowiedzieć za pomocą danej metody. Na pewnym poziomie są to wybory oczywiste. Im jednak problem badawczy jest bardziej złożony, im mniej mamy wiedzy na jego temat, tym wybór jest trudniejszy, oparty raczej na doświadczeniu i intuicji. Studiowanie podręczników nie zastąpi wiedzy praktycznej, ugruntowanej na bazie doświadczenia (por. Oakeshott 2008). Czytanie literatury metodologicznej ma jednak ten pożytek, że zmniejsza prawdopodobieństwo popełniania błędów, które inni już popełnili i wyciągnęli z tego wnioski[5]. Niech nie odstrasza nas na wstępie liczba metod oraz kwestii z nimi związanych. Wizyta w sklepie z narzędziami pokazuje, że jest wiele rodzajów kombinerek; podobnie jest z metodami badawczymi: ich wielość wynika z wymyślania coraz bardziej wyrafinowanych sposobów mierzenia się z coraz bardziej wąskimi problemami badawczymi. Nie ma uniwersalnych metod badawczych, takich, które można by zawsze i wszędzie zastosować. Poza ryzykiem błędnego zastosowania oznacza to, że nawet najlepiej użyte narzędzie daje obraz fragmentu rzeczywistości społecznej. Aby zilustrować złożoność kwestii, która się kryje za tym stwierdzeniem, można odwołać się do anegdoty o niewidomych, którzy mieli opowiedzieć, czym jest słoń: jeden chwycił za nogę, drugi za trąbę, a trzeci za ogon. Każdy z nich powiedział, co czuje, uznając to za wyczerpujący opis słonia. Metody badawcze są narzędziami poznania rzeczywistości społecznej, należy jednak pamiętać, że umożliwiają nam poznanie tylko pewnego jej aspektu, którego nigdy nie powinniśmy brać za opis całości.

Metody badawcze dają nam wgląd w rzeczywistość społeczną z pewnej perspektywy, co oznacza, że kształtują w dużej mierze obraz, który uzyskujemy. To istotny problem dla filozofii nauki, którym się tutaj nie będziemy zajmować (por. Crab, Benton 2003). Ważne jest jednak, aby podczas interpretacji wyników pamiętać o metodzie, którą wykorzystano do

[5] Dobre podstawowe wprowadzenie w języku polskim w metody jakościowe i ilościowych zawiera praca Frankfort-Nachmias, Nachmias 2001. Wyłącznie metodom jakościowym są poświęcone dwie klasyczne pozycje: Milees i Huberman 2000, Silverman 2007 oraz Jasiński, Kowalski 2007. W odniesieniu do metod jakościowych w kontekście ewaluacji warto polecić pozycję: Shaw 1999; Patton 2002; Kuckartz, Dresing, Rädiker, Stefer 2008.

ich zebrania. Wypowiedzi zebrane podczas zogniskowanych grupowych wywiadów oznaczają co innego niż wypowiedzi zebrane podczas indywidualnego wywiadu pogłębionego, nawet jeśli jest to sama wypowiedź.

Przedstawione poniżej enumeratywne omówienie poszczególnych metod nie jest wyczerpujące. Poniższa prezentacja ogranicza się do trzech najważniejszych rodzajów metod badawczych. Większość nowych metod jest uszczegółowioną, dopasowaną do jakiegoś problemu, wersją jednej z omówionych poniżej metod.

2.1. Wywiad

Pozornie wywiad trudno nazwać metodą. Zadawanie pytań było od zawsze najlepszym sposobem zdobywania wiedzy. Czy jest jakaś różnica pomiędzy pytaniem o drogę napotkanego przechodnia a pytaniem uczestnika jakiegoś programu o to, czy na szkoleniu wyczerpująco omówiono dane zagadnienie? Pomimo powierzchownego podobieństwa, można wskazać trzy zasadnicze różnice.

Po pierwsze, kiedy pytamy o drogę, korzystamy z wiedzy innej osoby, bez względu na to, czy wiedza ta była prawdziwa, czy nie. W przypadku pytania o ocenę szkolenia nie zbieramy wiedzy o szkoleniu – interesuje nas **subiektywne nastawienie** do określonego przedmiotu. Ta subtelna różnica określa istotę metody badawczej, jaką jest wywiad. Wywiad służy do rekonstrukcji opinii ludzi na temat danego procesu, ich doświadczeń z uczestnictwa w nim, przeżyć, które w trakcie tego procesu im towarzyszyły i motywów, które nimi kierowały. Wywiad nie daje nam wiedzy o tym, jakie dane wydarzenie na prawdę jest, jaki miało przebieg.

Łatwo ze sobą te dwa aspekty pomieszać. W trakcie wywiadu rozmowa zawsze dotyczy konkretnego wydarzenia, na przykład programu, w którym osoba badana uczestniczyła w jakiejś roli. Pytania są formułowane tak, aby badany wyraził swoją opinię, na przykład na temat materiałów szkoleniowych, trenerów, o warsztatów. Analizując wypowiedzi trzeba zawsze pamiętać o tym, że nie uzyskujemy w ten sposób oceny materiałów, trenerów i warsztatów, a jedynie wiedzę na temat tego, jak były one odbierane przez badanych. Jeśli zadaniem ewaluatora jest ocena poszczególnych elementów programu, to opinie zebrane wśród beneficjentów mogą być pomocne, ale zdecydowanie niewystarczające do sformułowania zobiektywizowanej oceny.

Po drugie, w porównaniu do zwykłej rozmowy, wywiad prowadzi się według pewnych reguł, różnych w zależności od rodzaju wywiadu. Klasyczną formą wywiadu jest pogłębiony wywiad zogniskowany, zwany IDI. Wywiad taki realizuje się z jedną osobą, według określonego scenariusza (czyli zbioru pytań, które chcemy zadać badanemu). Kolejność pytań oraz sposób ich sformułowania nie muszą być koniecznie przestrzegane. Liczy się, by rozmowa była naturalna i aby poszczególne wątki

rozmowy wynikały płynnie z poprzednich. Scenariusz określa jednak zakres tematyczny rozmowy i w dużej mierze jej przebieg.

Przeciwieństwem tej formy jest wywiad nieukierunkowany, niespotykany raczej w ewaluacji, choć wykorzystywany w badaniach naukowych i marketingowych. Idea wywiadu nieukierunkowanego wychodzi z założeń teorii socjologicznych, które podkreślają spontaniczność życia społecznego oraz uprzywilejowaną rolę ludzkiej subiektywności. Główne założenie tej metody polega na minimalizowaniu wpływu badacza na respondenta. Pytania są zadawane w sposób maksymalnie ogólny i krótki, przy założeniu, że badany spontanicznie rozwinie narrację na dany temat.

Bardzo popularną formą wywiadu jest zogniskowany wywiad grupowy, zwany fokusem lub FGI[6]. Jest to metoda niezwykle popularna w każdym rodzaju badań (rynkowych, społecznych i ewaluacyjnych) i ma wiele wersji. Współcześnie funkcjonuje na samodzielnych prawach metody niełączonej zasadniczo z wywiadem indywidualnym. Idea wywiadu grupowego polega na tym, że grupa osób (od 6 do 10) prowadzi pomiędzy sobą rozmowę na określone tematy, podsuwane im przez prowadzącego (moderatora). Nie jest to zatem uproszczona i przyspieszona wersja wywiadów indywidualnych, w których prowadzący odpytuje po kolei poszczególne osoby. Moderator ma jedynie stymulować rozmowę i kontrolować jej przebieg. Dobre badanie fokusowe polega na tym, że ludzie rozmawiają ze sobą o danym problemie, formułując swoje opinie, broniąc ich i krytykując (w pozytywny sposób) opinie innych. Fokus jest metodą trudną i wymagającą od moderatora dużego doświadczenia.

Omówione pokrótce trzy metody różnią się znacznie metodologią i, jak widać, nie przypominają zwykłej rozmowy. Wywiady są poza tym praktycznie zawsze nagrywane. Stwarza to dość nietypową sytuację, w której osoba wypowiadająca się wie, że jej wypowiedź jest utrwalona i będzie dalej analizowana, co może wpływać na wypowiedzi. Stwarza to wiele problemów etycznych i metodologicznych. Zdarza się, że badany, powściągliwy w czasie wywiadu, uzupełnia swoje wypowiedzi o opinie istotne dla procesu ewaluacji po wyłączeniu urządzenia nagrywającego. Czy można wykorzystać taki materiał? Raczej nie. Z etycznego punktu widzenia byłoby to złamanie zasady poufności. Z metodologicznego – błędem jest wykorzystywanie informacji, na poparcie której brak nam materiałów. Opinia lub wiedza zdobyte poufnie i nieutrwalone są łatwe do podważenia. Ewaluator naraża się wówczas na zarzut braku profesjonalizmu lub stronniczości.

Wywiad jest niezastąpioną metodą zbierania wiedzy o rzeczywistych motywach, doświadczeniach, postawach, opiniach ludzi biorących udział

---

[6] Wyczerpujące omówienie tej metody w języku polskim można znaleźć w Maison 2001.

w jakichś zdarzeniach. Warto jednak zdawać sobie sprawę z wad i ograniczeń, jakimi obciążona jest ta metoda. Po pierwsze, dobór badanych do wywiadu, jest dość problematyczny, co stanowi problem wspólny dla różnych rodzajów badań jakościowych. Kolejny problem, o którym trzeba zawsze pamiętać, choć ma on naturę praktyczną, to czasochłonność i kosztowność tego typu badań. Wbrew pozorom szybciej można zrealizować badanie ilościowe na próbie 500 osób metodą ankiety realizowanej przez telefon niż zrobić na przykład 20 wywiadów. Co więcej, jeśli owe 20 osób mieszka w różnych, odległych od siebie miejscach, może to być nawet bardziej kosztowne. Trzecią wadą jest trudność i czasochłonność interpretacji. Analiza danych jakościowych wymaga wiedzy i doświadczenia. Bardzo rzadko spotkać można dobre raporty z wywiadów pogłębionych. Najczęściej ograniczają się one do przytoczenia różnych wypowiedzi zebranych w trakcie wywiadów, do których dodany jest komentarz autora. Raport taki trudno nazwać analizą. W najlepszym wypadku jest rzetelnym streszczeniem wypowiedzi kilkunastu osób[7]. Błąd, jaki się zazwyczaj popełnia, polega na tym, że wychodzi się od poszczególnych wypowiedzi, nie zaś od problemu badawczego. Kluczem do dobrego raportu jest precyzyjne postawienie problemu, adekwatne do danych, jakie wywiad pogłębiony może przynieść. Dobrze postawiony problem daje perspektywę analityczną, pozwalającą wyjść poza poszczególne wypowiedzi i zinterpretować ich sens.

Dotykamy tutaj kolejnego problemu związanego z wywiadami: chodzi o często podnoszoną groźbę subiektywnej analizy. Warto wspomnieć, że mitem jest, jakoby metody ilościowe były w pełni obiektywne i nienarażone na groźbę subiektywnego skrzywienia lub manipulacji (por. Best 2001; Huff, Geis 1993). Przekonanie, że metody ilościowe dają bardziej wiarygodne wyniki bierze się z kilku przyczyn. Po pierwsze, uzyskane z losowej próby wyniki można statystycznie uogólniać na całą populację, to znaczy, przebadanie zgodnie z pewnymi rygorami wybranej liczby osób pozwala założyć, że zebrane opinie są charakterystyczne dla całego zbioru, z którego wybieraliśmy próbę. Wnioskowanie takie zawsze jest jednak obciążone określonym błędem. Po drugie, liczby same w sobie nic nie mówią i wymagają komentarza. Po trzecie, w przypadku metod ilościowych działa magia liczb. Ze względów kulturowych (szczególnej roli nauk przyrodniczych w naszej kulturze) reprezentowanie czegoś za pomocą liczb jest uznawane za bardziej obiektywne, bardziej prawdziwe niż tekst, z którego składa się raport, odbierany jako subiektywny.

Każdy, kto prezentował wyniki badań jakościowych, wie, że łatwo narazić się na zarzut nieobiektywności, szczególnie, jeśli to, co ma do

---

[7] Nie ma tutaj miejsca na omówienie złożonych i trudnych zawiłości analizy danych jakościowych. Dla osób zainteresowanych pogłębieniem swojej wiedzy godny polecenia jest podręcznik Milesa i Hubermana (2000). Opisano tam aż 26 sposobów wyciągania wniosków i ich weryfikacji.

powiedzenia, jest w opozycji do opinii odbiorów. Dobrze przemyślany plan badań i umiejętnie dobrana próba są podstawą do formułowania wniosków z badań. Dlatego na wstępie pisania raportu warto poświęcić im wiele uwagi. Sama analiza zebranego materiału powinna być logiczna i systematyczna. Każdy kolejny krok analizy, prowadzący do wniosków, powinien być dobrze uzasadniony, aby nie było wątpliwości, że wyraża on rzeczywisty obraz sytuacji, a nie subiektywne poglądy badacza. To właśnie powoduje, że pisanie raportów z badań jakościowych jest przedsięwzięciem czasochłonnym i trudnym.

2.2. Wtórna analiza danych

Wtórna analiza danych jest pojęciem rozległym, obejmującym bardzo różne metody. W dosłownym znaczeniu, chodzi tutaj o analizę dokumentów zastanych. Jest to rodzaj badania, charakterystyczny dla pracy historyka, który bazuje przede wszystkim na różnorodnych dokumentach oficjalnych i prywatnych, analizując je pod kątem interesującego go tematu. W przypadku ewaluacji nie ma aż tak rozbudowanej metodologii krytycznej analizy danych zastanych, gdyż samo badanie traktowane jest, jako oczywiste. Każdy ewaluator swoją pracę zaczyna od przestudiowania dokumentów powstałych w trakcie realizacji programu. Nie pojawia się w tym kontekście typowy dla historyka problem wiarygodności źródeł, sposobu ich interpretacji czy relacji do innych materiałów.

Analiza danych zastanych nie kończy się jednak na analizie dokumentów programowych i biurokratycznych ewaluowanego programu. Ważnym elementem takiej analizy jest sięganie do innych badań, raportów, opracowań i analiz, które są w jakiś sposób związane z programem. Przystępując na przykład do ewaluacji projektu dla bezrobotnych, ewaluator powinien zapoznać się z materiałami na ten temat. W takim ujęciu, wtórna analiza danych oznacza zbieranie wiedzy na temat obszaru, którego dotyczy program tak, aby mieć kontekst społeczno-ekonomiczny, konieczny do interpretacji wyników ewaluacji programu. Wtórne dane mogą służyć nie tylko jako źródło dodatkowej wiedzy dla ewaluatora, lecz również stanowić istotny element procesu ewaluacji. Dostępnych jest zarówno wiele danych statystycznych, jak i danych z badań jakościowych, które mogą być wykorzystane do konstruowania wskaźników efektywności; mogą również stanowić punkt odniesienia do analizy procesu realizacji programu.

Odrębnym rodzajem analizy danych, który współcześnie funkcjonuje jako odrębna i wyspecjalizowana metoda, jest analiza treści, mająca na celu systematyczne badanie zawartości i formy przekazu. Metoda ta może być wykorzystana do analizy różnego rodzaju przekazów: materiałów prasowych, dokumentów, wywiadów, a nawet obrazów czy

fotografii. Istota analizy treści polega na systematycznym wyszukiwaniu interesujących badacza treści i analizowaniu kontekstów, w których one występują. Metoda ta rozwinęła się szczególnie intensywnie na początku lat 90. ubiegłego wieku, wraz z pojawieniem się programów komputerowych, umożliwiających analizowanie nieograniczonych praktycznie ilości materiałów zarówno w sposób jakościowy, jak i ilościowy.

Problemem wspólnym dla różnych rodzajów wtórnej analizy danych jest to, że najczęściej dane zastane jedynie w przybliżeniu odpowiadają potrzebom badacza: nawet jeśli znajdzie on interesujący raport na dany temat – raport ten dotyczyć może innego regionu lub pochodzić sprzed kilku lat. Często w danych występują – istotne z punktu widzenia ewaluatora – braki. Problem stanowić może również niejasny sposób zbierania danych, co podaje w wątpliwość ich wiarygodność. Ostatnia, choć wcale nie najmniej ważna kwestia, to utrudniony dostęp do danych – wiele instytucji gromadzi różne, interesujące dla ewaluatora, dane, jednak mogą one być dostępne jedynie odpłatnie.

2.3. Obserwacja

Ostatnią metodą, którą warto pokrótce omówić, jest obserwacja. Przez podręczniki metodologii traktowana jest nieco po macoszemu, co częściowo wynika z faktu, iż trudno jest w tym wypadku wymyślić coś unikalnego. Doświadczenie badawcze pokazuje jednak, że nie sposób przecenić znaczenie i wartość wspomnianej metody. Możliwość zobaczenia na własne oczy miejsca, w którym np. odbywają się szkolenia, osobistego uczestnictwa w szkoleniu, daje znajomość kontekstu do interpretacji danych zbieranych innymi metodami. Jest to trudno wymierna wiedza i niełatwo wskazać jej konkretne zalety, jednak każdy, kto analizował zgromadzone przez innych dane bez bezpośredniego doświadczenia, jakiego nabywa się w trakcie obserwacji uczestniczącej wie, że zrozumienie poszczególnych wypowiedzi bywa czasem niezwykle trudne. Zupełnie inaczej przedstawia się sytuacja, gdy ewaluator wie, o czym badani mówią, bo sam to zobaczył lub w czymś podobnym uczestniczył.

Sytuacja taka jest zupełnie zrozumiała, jeśli weźmie się pod uwagę fakt, iż wypowiedzi zbierane podczas rozmów są ze swojej natury oparte na pewnych oczywistościach, niewymagających wyjaśnienia: dwie rozmawiające ze sobą osoby nie muszą sobie wszystkiego wyjaśniać, wystarczy na coś wskazać, coś pośrednio wspomnieć. Cały kontekst rozmowy ginie w spisanym tekście, analizowanym w innymi miejscu i czasie.

Obserwacja w swoim podstawowym sensie jest sposobem zbierania informacji niejako przy okazji prowadzenia innych badań: ewaluator

przeprowadzający wywiad pogłębiony ma jednocześnie możliwość osobistego wglądu w to, jak wyglądają warunki, w których program jest realizowany, jak odnoszą się do siebie ludzie, z którymi rozmawia itd. Obserwacja może jednak przybrać bardziej sformalizowany charakter – dzieje się tak wtedy, gdy przybiera postać wizyt studyjnych. Często zdarza się, że wizyty te stanowią część systemu zarządzania projektem i odbywa je personel zarządzający. Ewaluator może wykorzystać taką sytuację do zebrania dodatkowych danych. Wystarczy wówczas przygotować proste narzędzie w postaci zbioru kilku pytań/kwestii, na które osoba wizytująca powinna zwrócić uwagę, zapisując swoje obserwacje. Taki materiał jest doskonałym uzupełnieniem obserwacji wyniesionych przez samego ewaluatora.

Przy okazji obserwacji uczestniczącej pojawia się wątpliwość, czy badaczowi uda się dotrzeć „za kulisy", to znaczy, czy uda mu się zobaczyć nie tylko oficjalną stronę projektu, ale również skrywaną w takich sytuacjach rzeczywistość. Aby działanie to zakończyło się sukcesem, niezbędne jest dobre przygotowanie badacza do przeprowadzanej wizyty (nie może zadawać naiwnych pytań, nie powinien zadowalać się powierzchownymi odpowiedziami itp.).

3. WYKORZYSTANIE METOD JAKOŚCIOWYCH W EWALUACJI

Omówione metody są wspólnym warsztatem badań społecznych, marketingowych i ewaluacyjnych. W zależności od tego, do czego są wykorzystywane, zmienia się sposób ich użycia. W badaniach ewaluacyjnych wypracowano kilka standardowych sposobów wykorzystania poszczególnych metod jakościowych – poniżej omówione zostaną typowe spośród nich (por. Patton 2002).

Efekty każdego programu można podzielić na krótko- i długoterminowe[8]. Efekty krótkoterminowe są łatwo mierzalne, występują bezpośrednio po zakończeniu programu lub nawet w czasie jego trwania i – co nie mniej ważne – dają się przedstawić w sposób ilościowy. Utartą praktyką większości programów jest zakładanie już na wstępie ilościowego poziomu efektów krótkoterminowych. Z wymienionych względów dominującą metodologią w tym przypadku są badania ilościowe. Skupiając całą uwagę na efektach krótkoterminowych, zapomina się jednak, iż celem programów jest **zmiana społeczna**, nie zaś np. przeszkolenie *N*-osób czy znalezienie tym osobom pracy na okres 6 miesięcy. W przypadku

[8] W języku polskim brakuje dobrych słów na wyrażenie stosowanej w praktyce, szczególnie amerykańskiej, podziału na *output* (efekty bezpośrednie realizacji danego programu) i *outcome* (efekty pośrednie). Szczegółowe omówienie ewaluacji efektów pośrednich i zwrotu, jaki się dokonał przy tej okazji w ewaluacji, można znaleźć w pracy wydanej przez stowarzyszenie United Way of America (1994). Praca ta wyznaczyła nowy standard w podejściu do ewaluacji w USA.

celów długoterminowych również posługujemy się wskaźnikami liczbowymi, mają one jednak tę wadę, że dają uproszczony obraz zjawisk społecznych, które chcemy uchwycić. Nie dowiemy się z nich, czy naprawdę coś się zmieniło w sposób trwały. Konieczne jest w tym przypadku posłużenie się metodami jakościowymi. Tak więc metody jakościowe są niezastąpione wówczas, gdy problemy analizujemy w szerszej perspektywie. Dobrze przygotowane badanie jakościowe pokaże, na czym polegał wpływ danego programu, które jego elementy w dłuższej perspektywie czasowej okazały się efektywne, które wpłynęły w sposób trwały na motywację jednostek, na ich sposób działania. Pokaże również, co okazało się nieskuteczne, jakie podejmowane w ramach programu działania nie zostały zapamiętane, a jakie odbierano negatywnie.

Aby osiągnąć dany efekt, program musi być zrealizowany – należy wykonać wiele czynności, podjąć wiele działań. Wszystko to składa się na pewien **proces** realizacji, prowadzący do określonych efektów. Efekty programu są jednak czymś innym niż proces do nich prowadzący. Wystarczy sobie uświadomić, że identyczne efekty można osiągnąć w różny sposób: niektórzy potrzebują więcej czasu, inni zużywają więcej materiałów, czasem pojawiają się konflikty... Odstępstw od planu nie można przewidzieć (gdyby było inaczej, stanowiłyby część planu). Chcąc badać to, co nieprzewidywalne, musimy posłużyć się maksymalnie elastycznymi metodami, tymi zaś z definicji są metody jakościowe. Swobodny wywiad pogłębiony czy grupowe wywiady zogniskowane pozwolą ewaluatorowi zebrać potrzebną wiedzę na temat tego, co dzieje się w danym programie.

Użycie metod jakościowych może być szczególnie istotne przy równoległej realizacji programu w złożonym otoczeniu – dla wyjaśnienia mechanizmów wdrażania programu (zob. ramka).

Program aktywizacji obywatelskiej był realizowany w kilkudziesięciu gminach; celem długofalowym programu była aktywizacja społeczna, mierzona wzrostem zainteresowania działalnością władz gminy, wyrażona takimi mierzalnymi kryteriami jak: liczba zapytań skierowanych do wójtów/burmistrzów, wzrost liczby odwiedzin u radnych itd. Przygotowując program, opracowano pulę różnych narzędzi aktywizacji, które miałyby być w nim wykorzystane. Wyniki programu okazały się bardzo zróżnicowane, jednak ze względu na to, że całą uwagę przy ewaluacji skoncentrowano na celach krótkoterminowych, trudno było znaleźć przyczynę owych rozbieżności. Dopiero pogłębiona ewaluacja *ex post* pokazała, jak wiele zależało od gminy, w której program był realizowany: od zaangażowania władz i liderów lokalnych, od ogólnych społeczno-ekonomicznych uwarunkowań.

Opisane doświadczenie pokazuje, że przy projektach, co do których można przypuszczać, że istotnym czynnikiem wpływającym na efekt końcowy będzie kontekst, w którym program będzie realizowany, wskazane jest uzupełnienie metod ilościowych, na przykład o pogłębione studium procesu implementacji metodą studiów przypadków.

Sama ilościowo ujęta analiza stosunku kosztów do zysków (i stopnia realizacji planu) jest często myląca. Weźmy za przykład program, w którym udało się zrealizować zakładane cele krótkoterminowe przy względnie niskich kosztach. Analiza ilościowa pokaże nam tylko sukces. Co jednak dzieje się w sytuacji, gdy realizacji programu towarzyszyły konflikty wewnątrz zespołu lub tarcia pomiędzy partnerami zaangażowanymi w jego prowadzenie? Co dzieje się w sytuacji, gdy beneficjenci programu mieli złe doświadczenia związane z jego realizacją?

Badania jakościowe są więc szczególnie wskazane jako narzędzie rekonstrukcji procesu realizacji programu, gdyż pozwalają opisać w sposób szczegółowy i zindywidualizowany zaangażowanie ludzi w dany program. Tylko w ten sposób będziemy mogli poznać motywy działania poszczególnych jednostek i wyjaśnić bieg zdarzeń.

Analizując proces realizacji programu, będziemy również mogli zidentyfikować czynniki odpowiedzialne za jego sukces bądź porażkę. Badanie dostarcza informacji szczególnie istotnych w przypadku programów powtarzalnych: jeśli jakiś program ma być powtórzony, warto wiedzieć, co przyczyniło się do jego ostatecznego kształtu, bez względu na to, czy odniósł sukces, czy nie. Czynniki przyczyniające się do sukcesu danego programu mogą być bardzo specyficzne, może to być na przykład zaangażowanie władz samorządowych lub współpraca z grupą lokalnych przedsiębiorców. Czynniki te nie muszą być obecne przy następnych edycjach programu, jednak bez wiedzy na temat ich znaczenia łatwo je przeoczyć, co prowadzić może do nieprzewidzianych sytuacji.

Na koniec należy wspomnieć o dominującym modelu postrzegania sposobów wykorzystania metod jakościowych w ewaluacji, sprowadzającym je do roli pomocniczej w stosunku do metod ilościowych. Na wstępie uwaga wyjaśniająca: ów sposób myślenia o metodach jakościowych nie jest wcale błędny, wręcz odwrotnie – to dobrze ugruntowane teoretycznie oraz poparte niezliczoną ilością badań podejście. Problem polega na tym, iż podejście to jest jednostronne, co utrudnia pełne wykorzystanie płynących z nich możliwości.

Wykorzystując metody jakościowe, jako uzupełnienie dla danych ilościowych, możemy zbudować narrację o programie. Wywiady pozwalają zrekonstruować historię, która da szerszy horyzont do interpretowania danych ilościowych. Przykładem może być autentyczna historia z realizacji programu pomocowego (zob. ramka na następnej stronie).

Metody jakościowe są pomocne w procesie przygotowania narzędzi ilościowych. Badania ilościowe, o czym dość często się zapomina,

W pewnej miejscowości ocena szkoleń była słabsza w porównaniu z innymi. Przez długi czas nie wiadomo było dlaczego, dopiero wywiady pogłębione, przeprowadzone z beneficjentami programu, pozwoliły zdobyć istotną informację: okazało się, że jakość dowożonego na szkolenia pożywienia postrzegana była jako „nie najlepsza", co odbierano jako wyraz lekceważącego stosunku do uczestników programu, co z kolei rzutowało w sposób istotny na ocenę całego przedsięwzięcia.

opierają się na wielu założeniach dotyczących badanej rzeczywistości. Przygotowując badanie ilościowe, osoba realizująca projekt powinna dysponować szeroką wiedzą na temat problemu badawczego (na przykład, przygotowując ankietę, badacz musi ustalić listę pytań, wyczerpujących temat badania i zrozumiałych dla respondentów). Jeśli ewaluator nie ma pełnej wiedzy na temat wydarzeń istotnych z punktu widzenia ogólnej oceny programu – wstępne badania jakościowe umożliwią mu zdobycie ogólnej orientacji w przebiegu programu, wskażą istotne problemy, pozwolą określić, które są mniej ważne. Tego typu wstępne badanie jest użyteczne także wówczas, gdy ewaluator porusza się po słabo rozpoznanym terenie, na przykład, gdy program realizowany jest po raz pierwszy lub po raz pierwszy realizowany w danym środowisku społecznym. Przeprowadzenie badania jakościowego przed realizacją ankiety wskazane jest również w sytuacji, gdy problem jest słabo poznany lub ewaluatorzy mają niewielkie doświadczenie w jego zakresie.

Pozornie mało istotnym problemem jest kwestia, czy pytania są sformułowane w sposób zrozumiały. Formułując pytanie w ankiecie, badacz musi mieć pewność, że respondenci zrozumieją je w taki sposób, w jaki on je rozumie. Szybkie badanie jakościowe, zrealizowane przy wykorzystaniu zogniskowanych wywiadów pogłębionych w grupie potencjalnych respondentów, pozwoli uniknąć wielu błędów i rozczarowań. Może być bowiem tak, iż duża liczba braków odpowiedzi na dane pytanie nie wynika z faktu, iż respondenci nie mieli zdania na temat, którego dotyczyło, mogli po prostu go nie zrozumieć, uznać za niejasne lub, co gorsza, obraźliwe. Weryfikacja zrozumiałości pytań ankiety jest szczególnie wskazana w przypadku badań na dużych próbach (czyli badań kosztownych) oraz realizowanych w grupach społecznych o niższych kompetencjach kulturowych.

Badania jakościowe są użyteczne także po zakończeniu badań ilościowych. Interpretując dane uzyskane z ankiety, niejednokrotnie napotkać można zaskakujące wyniki. Ewaluator może oczywiście opierać się na domysłach i intuicji, wyjaśniając przyczyny, dla których respondenci odpowiadali tak, a nie inaczej, jednak tego typu argumenty będą narażone

na krytykę ze strony odbiorców badań. Jeśli na przykład w jednej miejscowości dany trener jest oceniany dużo słabiej niż w pozostałych, to można przypuszczać, że miały tam miejsce jakieś zdarzenia istotne dla uzyskanej oceny. Tylko badanie jakościowe – wywiad z trenerem oraz uczestnikami szkolenia – może dać odpowiedź na pytanie, co się stało i dlaczego uzyskano takie, a nie inne wyniki.

4. WIARYGODNOŚĆ METOD JAKOŚCIOWYCH

W omówieniu metod jakościowych nie może zabraknąć istotnej kwestii, jaką jest problem wiarygodności. Jak wspomniano, metody jakościowe generalnie postrzegane są jako mniej wiarygodne od metod ilościowych. Taka opinia wydaje się nietrafna i niesprawiedliwa. Nie zmienia to jednak faktu, iż badacz jakościowy, a szczególnie ewaluator, powinien być bardzo wyczulony na ten problem, bowiem najczęstszym sposobem podważania wyników ewaluacji jest właśnie kwestionowanie obiektywności badań. Badacz jakościowy nie może pozwolić sobie na brak precyzji. Od projektu badania po końcowe wnioski, badanie jakościowe powinno opierać się na jasnych założenia, rzetelności badania i przejrzystej logice analizy.

Brak tu miejsca na dokładne omówienie wszystkich kwestii związanych z wiarygodnością, jednak na jedną kwestię warto zwrócić uwagę – problem doboru próby (por. Crouch, McKenzie 2006). Najczęściej pojawiające się pytanie w odniesieniu do badań jakościowych brzmi: jak dobrać ludzi do badania? Widać w tym wpływ metod ilościowych, opierających swą wiarygodność na reprezentatywności próby dla populacji.

W pewnym sensie, przytoczone pytanie wydaje się źle postawione. Jak wskazano, istotą badań jakościowych nie jest szukanie opisowych informacji reprezentatywnych dla całej populacji, lecz zrozumienie przyczyn zachodzenia pewnych zjawisk. Pamiętając o tym, możemy przeformułować powyższe pytanie w sposób następujący: jak wybrać osoby do badania, by uzyskać wiedzę konieczną do wyjaśnienia interesujących nas procesów? Michael Patton wymienia aż 15 strategii doboru jednostek do badania; ograniczę się tu do przytoczenia pięciu najważniejszych.

1. **Skupienie uwagi na przypadkach ekstremalnych**. Badając problem bezrobocia, możemy zrekrutować do badania osoby długotrwale bezrobotne, wychodząc założenia, że dyskusja z nimi pozwoli nam w sposób najwyraźniejszy znaleźć zależności i przyczyny odpowiadające za bezrobocie. Pewną wersją takiego sposobu doboru próby jest skupienie się na przypadkach przeciwstawnych: aktywny zawodowo/nieaktywny zawodowo, długotrwale bezrobotny/krótkotrwale bezrobotny itd. Prowadząc badania, następnie zaś analizując uzyskany materiał,

skupiamy się na porównaniu obu przypadków, aby wydobyć z nich szczególne wzory zachowań/motywów.

2. **Strategia próbowania wszystkiego po trochu**. Przyjmując tę metodę, na pierwszy rzut oka wydaje się to metodą optymalną, należy jednak pamiętać o niebezpieczeństwie zgubienia z pola widzenia problemu badawczego. Bardzo rzadko maksymalnie zakreślone pole badawcze można połączyć z głębią analizy. Skądinąd, to dość częsty błąd popełniany przez osoby niedoświadczone w realizacji badań jakościowych: mnożą liczbę jednostek do zbadania, nie zdając sobie sprawy, iż ceną za ilość jest powierzchowność uzyskanych wyników.

3. **Strategia dążenia do maksymalnej homogenizacji próby**. Ta strategia także nie jest idealna. Koncentrując się jedynie na kilku wybranych przypadkach, mamy możliwość zdobycia pogłębionej wiedzy na temat danego problemu. Niebezpieczeństwo polega jednak na tym, że łatwo przeoczyć pewne zjawiska, które mogą pojawiać się poza tą homogenicznie określoną grupą.

4. **Strategia kuli śniegowej**. Przyjmując tę strategię, badacze wychodzą z założenia, że próba będzie się konstruowała w pewnym sensie sama w trakcie badania: od samych badanych dowiadujemy się, kto jest ciekawym przedmiotem do dalszych badań. Metoda ta gwarantuje sporą elastyczność, ale i dowolność. Ze względu na wspomniany problem wiarygodności jakościowych badań ewaluacyjnych metoda ta nie wydaje się godna polecenia; wskazane jest odgórne określenie zasad doboru próby

5. **Dobór ze względu na określone kryteria**. Badacz z góry określa kryteria wyboru poszczególnych jednostek do badania. Kryteria te powinny w sposób logiczny wynikać z problemów badawczych. Metoda ta jest przeciwieństwem strategii kuli śniegowej i – jako taka – posiada przeciwstawne słabości. Przede wszystkim, nakłada nadzwyczaj sztywne ramy na badacza, uniemożliwiając mu dobranie do badania przypadków, które z jakichś przyczyn wydają się interesujące, jednak nie mieszczą się w przyjętych ramach.

Przedstawiając pięć metod doboru próby, celowo uwypuklono problemy z nimi związane tak, by podkreślić, iż nie istnieje jedyna słuszna (skuteczna) strategia. Na pytanie, którą z nich wybrać można odpowiedzieć tylko w jeden sposób: to zależy... od problemu badawczego, od środowiska, w którym badacz się porusza, a także – co czasem okazuje się kryterium przesądzającym o wyborze takiej a nie innej metody – od przeznaczonego na projekt budżetu.

5. ZAKOŃCZENIE

Tekst ten nie wyczerpuje złożoności problematyki badań jakościowych w ewaluacji. Jedynie praktyka połączona z ciągłymi studiami pozwoli

opanować trudny i rozległy obszar badań jakościowych[9]. Praktyka ewaluacji realizowanych w Polsce pokazuje, że coraz śmielej sięga się do metod jakościowych, wykorzystując je nie tylko do ilustracji badań ilościowych. Można mieć nadzieję, iż w przyszłości zapanuje większa równowaga pomiędzy obiema metodologiami, wyniki badań jakościowych zaś przyjmowane będą równie chętnie jak „twarde" dane ilościowe.

BIBLIOGRAFIA

Becker H. (1996), *The Epistemology of Qualitative Research*, w: R. Jessor, A. Colby, and R. Schweder, eds., *Essays on Ethnography and Human Development*, Chicago, University of Chicago Press.

Best J. (2001), *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*, Berkley, University of California Press.

Crab I., Benton T. (2003), *Filozofia nauk społecznych: od pozytywizmu do postmodernizmu*, przekł. L. Rasiński, Wrocław, Wyd. Dolnośląskiej Szkoły Wyższej Edukacji.

Crouch M., Mckenzie H. (2006), The logic of small samples in interview-based qualitative research, *Social Science Information* **45(4)**, 483–499.

Flick U., Kardorff E., Steinke I. (Hg.) (2000), *Qualitative Forschung. Ein Handbuch*, Hamburg, Rowholts.

Frankfort-Nachmias Ch., Nachmias D. (2001), *Metody badawcze w naukach społecznych*, przekł. E. Hornowska, Poznań, Zysk i S-ka.

Huff D., Geis I. (1993), *How to Lie With Statistics*, New York, W.W. Norton & Company.

Jasiński M., Kowalski M. (2007), *Fałszywa sprzeczność: metodologia jakościowa czy ilościowa?* w: Haber A., *Ewaluacja ex post. Teoria i praktyka badawcza*, Warszawa, PARP.

Kuckartz U., Dresing T., Rädiker S., Stefer C. (2008), *Qualitative Evaluation: Der Einstieg in die Praxis*, Opladen, Vs Verlag.

Maison D. (2001), *Zogniskowane wywiady grupowe : jakościowa metoda badań marketingowych*, Warszawa, WN PWN.

Mason J. (2006), Mixing methods in qualitatively driven way, *Qualitative Research* **6(1)**, 9–25.

Maxwell J.A. (2004a), Using Qualitative Methods for Causal Explanation, *Field Methods* **16(3), 243–264**.

Maxwell J.A. (2004b), Causal explanation, qualitative research, and scientific inquiry in education, *Educational Researcher* **33(2)**, 3–11.

Miles M.B., Huberman A.M. (2000), *Analiza danych jakościowych*, przekł. S. Zabielski, Białystok, Transhumana.

[9] Warto tutaj polecić interesujące czasopisma poświęcone ewaluacji oraz badaniom jakościowym, które dla osób mających doświadczenie stanowić mogą niezastąpione źródło wiedzy: „American Journal of Evaluation", „Evaluation and Program Planning", „Evaluation, Qualitative Research", „Qualitative Research in Action".

Mohr L.B. (1995), *Impact analysis for program evaluation*, 2nd ed. Thousand Oaks, Calif., Sage.

Mohr L.B. (1996), *The causes of human behavior: Implications for theory and method in the social sciences*, Ann Arbor, University of Michigan Press.

Mohr L.B. (1999), The Qualitative Method of Impact Analysis, *American Journal of Evaluation* **20(1)**, 68–84.

Oakeshott M. (2008), *O postępowaniu człowieka*, przeł. M. Szczubiałka, Warszawa, WN PWN.

Patton Q.M. (2002), *Qualitative Research and Evaluation Methods*, 3nd ed. Thousand Oaks, Calif., Sage.

Pawson R., Tilley N. (1997), *Realistic evaluation*, Thousand Oaks, Calif., Sage.

Shaw I. (1999), *Qualitative Evaluation (Introducing Qualitative Methods)*, Thousand Oaks, Calif., Sage.

Silverman D. (2007), *Interpretacja danych jakościowych*, Warszawa, WN PWN.

Sułek A. (1979), *Eksperyment w badaniach społecznych*, Warszawa, WN PWN.

Thomas W., Znaniecki F. (1976), *Chłop polski w Europie i Ameryce*, Warszawa, PWN.

United Way Of America (1994), *Measuring Program Outcomes: A Practical Approach*, New York.

ROZDZIAŁ 2.3.

Łukasz Widła

ZASTOSOWANIE METOD ILOŚCIOWYCH W EWALUACJI

1. WSTĘP

Oprócz szeroko stosowanych technik jakościowych, w ewaluacjach wykorzystuje się także techniki ilościowe. W odróżnieniu od tych pierwszych, cechujących się „miękkością" wyników, w przypadku technik ilościowych otrzymujemy wyniki „twarde", to znaczy opisane na skalach matematycznych, interpretowalne w sposób ścisły, jak również, pod pewnymi warunkami, porównywalne z innymi badaniami ilościowymi.

Jest to możliwe dzięki kilku założeniom, którym podlegają badania ilościowe, do których należy standaryzacja narzędzi badawczych oraz zachowanie statystycznego aktualnego stanu wiedzy.

Do przedmiotów badawczych zaliczamy próbę do badania oraz kwestionariusz z wystandaryzowanymi pytaniami. Od konstrukcji tych dwóch narzędzi zależy w dużej mierze jakość otrzymanych w efekcie badania danych, jak również zasadność wysnuwanych wniosków. W dalszych częściach tego rozdziału przedstawiono rodzaje prób i zasady budowy kwestionariusza, pozwalające na samodzielne zaprojektowanie badania ilościowego oraz sprawdzenie narzędzi badawczych.

Za pomocą kwestionariusza można przeprowadzić badanie technikami ilościowymi. Obecnie na rynku najbardziej rozpowszechnione są badania CAPI[1] oraz CATI, spotyka się jeszcze zanikającą technikę PAPI,

[1] Omówienia skrótów i terminów specjalistycznych Czytelnik znajdzie w dalszej części rozdziału; krótkie definicje wszystkich stosowanych terminów zostały ponadto zamieszczone w słowniczku zamieszczonym na końcu rozdziału.

zaś w niedalekiej przyszłości zostaną opracowane techniki badań internetowych (CAWI), które obecnie borykają się jeszcze z pewnymi ograniczeniami.

Niezależnie od techniki zbierania danych wyróżnia się także rodzaje badań. Do najczęściej występujących należą badania *ad hoc* i badania trackingowe. Jako dodatek opisano przypadek szczególny, jakim jest badanie panelowe, które jest (a właściwie powinno być) wykorzystywane w badaniach ewaluacyjnych, a które rządzi się swoimi, nieco odmiennymi prawami.

W wyniku badania ilościowego otrzymujemy zazwyczaj bazę danych z zapisanymi w niej zbadanymi przypadkami. Opisanie budowy bazy danych wraz z procesem analiz statystycznych przekracza niestety zakres tego rozdziału. Jako że autor nie chce pozostawiać uczucia niedosytu, na końcu rozdziału (przed słowniczkiem) zamieszczono kilka podstawowych komend pozwalających uzyskać tabelę frekwencji bądź tabelę krzyżową za pomocą pakietu SPSS for Windows (zainteresowanych odsyłam do doskonałego opracowania Górniak, Wachnicki 2000).

Mając bazę danych i posiłkując się kwestionariuszem, można zacząć wysnuwać wnioski z badania. W ostatniej części tego rozdziału opisano reguły pisania podsumowań, zasadność wysnuwania wniosków oraz problem ekstrapolacji wniosków z badania na całą populację poddaną badaniu.

Celem tego rozdziału jest przygotowanie Czytelnika do samodzielnego przygotowania badania ilościowego, jak również do weryfikacji *post hoc*, czy badanie ilościowe zostało przeprowadzone zgodnie z obowiązującym *state-of-art*. Rozdział ten składa się z pięciu głównych części:

- Próba.
- Rodzaj badania.
- Kwestionariusz.
- Raport.
- Podsumowanie.

Na samym końcu rozdziału umieszczono kilka podstawowych komend SPSS, umożliwiających uzyskanie podstawowych informacji z bazy danych oraz słowniczek pojęć wykorzystywanych w badaniach ilościowych, pozwalający na szybkie zorientowanie się w żargonie statystycznym.

2. PRÓBA

Zasadniczo próby dzielą się na dwa rodzaje: próby losowe i próby celowe. Spotykane są także inne rodzaje, na przykład losowo-warstwowe. Prócz tego, z pojęciem próby wiążą się takie pojęcia jak: wiązka próby czy booster (albo nadreprezentacja).

Losowanie i dobór próby nie są koniecznymi działaniami, zapewniającymi interpretowalność danych. Możliwe są sytuacje, w których badacz

lub inna osoba projektująca badanie, może pominąć etap dobierania próby. Z taką sytuacją mamy chociażby do czynienia wtedy, gdy badana populacja jest na tyle mała, że możemy przebadać wszystkie jej elementy. Dla przykładu, jeżeli badanie dotyczy ewaluacji niewielkiego projektu dotyczącego restrukturyzacji zatrudnienia w danym przedsiębiorstwie, należy zadać sobie pytanie, czy nie prościej byłoby przebadać wszystkie podmioty poddane projektowi, niż losować, bądź dobierać w inny sposób próbę. W takim przypadku mówimy o badaniu kompleksowym, nieopartym na wybranej próbce respondentów. Najczęściej jednak, tak w badaniach ewaluacyjnych, jak w badaniach rynku czy opinii, nie mamy możliwości przebadania wszystkich respondentów, jesteśmy więc skazani na dobieranie prób, które odzwierciedlą badaną populację (Noelle 1976).

2.1. Próba losowa, próba reprezentatywna

Próbą losową nazywamy taką próbę, dla której każdy element populacji ma jednakową szansę wejścia do konstruowanej próby, jak dowolny inny element populacji. W praktyce oznacza to, że z badanej populacji musimy wylosować, w sposób całkowicie poddany rachunkowi prawdopodobieństwa, elementy naszej próby. Zakłada się, że losowanie, w którym nie ma możliwości, by którykolwiek z elementów był faworyzowany, jest losowaniem, które zapewni nam reprezentatywność próby.

Pojęcie próby reprezentatywnej spotyka się w badaniach wyjątkowo często i, jak to bywa z nadużywanymi terminami, bardzo często traci się sens tego wyrażenia. W rozumieniu potocznym, próba reprezentatywna to taka próba, której wyniki mogą być ekstrapolowane na populację. Należy w tym miejscu podkreślić, że nie jest to rozumienie ścisłe. Dla statystyka próba reprezentatywna jest zawsze reprezentatywna ze względu na coś. Może być próba reprezentatywna ze względu na wiek i płeć, może być próba reprezentatywna ze względu na preferencje polityczne, stan majątkowy czy liczbę zatrudnianych pracowników. W rzeczywistości pojęcie próby reprezentatywnej oderwane jest od swoich desygnatów, stając się określeniem-wytrychem.

Z pojęciem próby wiąże się także problem błędu estymacji oraz poprawki populacyjnej. Każda wylosowana próba jest obarczona błędem estymacji (jedynym sposobem na ich uniknięcie jest przebadanie całej populacji) i dobrze zdawać sobie sprawę z tego, ile ten błąd wynosi, po to, by móc oszacować, czy jego poziom jest akceptowalny. Roger Penrose stwierdził, że każdy wzór matematyczny, zamieszczony w książce lub artykule, zmniejsza liczbę czytelników o połowę. Na szczęście oba zamieszczone tu wzory są wyjątkowo proste: błąd estymacji wyliczany jest ze wzoru: $\frac{0,98}{\sqrt{n}}$, gdzie n to liczebność wylosowanej próby, zaś poprawka populacyjna liczona jest ze wzoru $\sqrt{\frac{N-n}{N}}$, gdzie n oznacza liczebność

wylosowanej próby, zaś N – wielkość populacji. Ostatecznie, błąd oszacowania wynosić będzie tyle, ile wynosi iloczyn błędu estymacji oraz poprawki populacyjnej. Jak się łatwo zorientować, poprawka przestaje mieć praktyczne znaczenie w przypadku wielkich populacji, gdyż jej wartość wówczas będzie wynosić ~1.

Akceptowalny błąd zależy ściśle od preferencji badacza i od tego, na ile ważna jest dla niego możliwość dokładnego ekstrapolowania na populację. Dla przykładu, badania przeprowadzane na losowej próbie 1000 Polaków obarczone są błędem ~3.1%, co oznacza, że każdy wynik otrzymany w takim sondażu może być różny od rzeczywistości o ok. 3% w górę, bądź w dół. Aby otrzymać błąd na poziomie 1%, należałoby przeprowadzić badanie na 10 tys. osób.

Trzeba w tym miejscu zauważyć, że jednym z najczęściej spotykanych błędów jest wylosowanie próby, której błąd jest zbyt duży, by wyniki oddawały dobrze rzeczywistą sytuację. Zawsze należy zdać sobie sprawę z tego, jakim błędem oszacowania obarczone są wyniki badania i czy poziom takiego błędu nie powoduje, że badanie jest bezużyteczne. Dla przykładu, jeśli z 300 interesujących nas obiektów ($N = 300$) przebadamy jedynie 30 ($n = 30$), błąd oszacowania, którym obarczone będą wyniki wyniesie ~17%. Przy takim poziomie błędu praktycznie nie ma możliwości wyciągnięcia sensownych i niepodważalnych wniosków na temat populacji (Barnett 1982).

Czasami jednak mniej zależy nam na otrzymaniu wyników, które mogłyby być odzwierciedleniem sytuacji w populacji, zaś bardziej zależy nam na tym, by z badanej populacji wszystkie grupy elementów były równo reprezentowane.

Załóżmy, że mamy do zbadania projekt unijny, któremu zostały poddane rozmaite przedsiębiorstwa. Przedsiębiorstwa te zaklasyfikowano na podstawie jakiegoś kryterium (dochodowego, przestrzennego bądź pod względem liczby pracowników) jako małe, średnie i duże. Małych przedsiębiorstw jest 60%, średnich 25%, zaś dużych 15%.

Przy zastosowaniu próby losowej, widać wyraźnie, że małe przedsiębiorstwo ma czterokrotnie wyższe prawdopodobieństwo wejścia do wylosowanej próby niż przedsiębiorstwo duże (60% : 15% = 4). W przypadku losowania próby stu przedsiębiorstw, oczekiwalibyśmy, że otrzymamy próbę, w której będzie 60 przedsiębiorstw małych, 25 przedsiębiorstw średnich i tylko 15 przedsiębiorstw dużych. Owe 15 przedsiębiorstw dużych może nie wystarczyć do otrzymania wyników, które będą miarodajne dla ogółu takich firm. W takim przypadku należy wylosować próbę trzykrotnie: osobne próby dla każdego z rodzajów przedsiębiorstw, tak by poszczególne próby składowe były w miarę równoliczne (dla 100 przedsiębiorstw losujemy próbę małych, średnich i dużych osobno, otrzymując po 33 wywiady dla każdego z nich). Takie losowanie próby nie jest reprezentatywne na poziomie całej populacji, jest jednak wciąż reprezentatywne na poziomie poszczególnych podgrup przedsiębiorstw. Jest

tak dlatego, że w podgrupie dowolnej wielkości przedsiębiorstw każdy element miał takie samo prawdopodobieństwo znalezienia się w próbie, jak każdy inny. Opisany rodzaj losowania próby nazywamy losowo-warstwowym.

2.2. Próba celowa

Drugi, zupełnie odmienny sposób doboru próby, nazywa się doborem celowym. Dobór celowy zakłada pewną wiedzę o ewaluowanych podmiotach, podczas gdy dobór losowy może dotyczyć również obszarów, które są dla badacza nowe bądź niedostatecznie poznane. Doboru celowego próby dokonujemy w momencie, gdy mamy pewne założone oczekiwania co do wyglądu próby, a które mogą nie zostać spełnione w przypadku losowania próby (Babbie 2003).

Załóżmy, że mamy do czynienia z projektem, który obejmował pięć województw. Skądinąd wiemy, że dwa z nich określane są jako województwa o dobrej kondycji makroekonomicznej, dwa jako województwa o średniej kondycji i jedno o kondycji słabej. W tych województwach działaniu ewaluowanego programu poddanych zostało w sumie 200 jednostek. Z tych 200 jednostek powinniśmy przebadać 50.

W przypadku losowania próby oczekiwalibyśmy, że w każdym województwie wylosowanych zostanie 10 jednostek. Ze względu jednak na różne warunki ekonomiczne w badanych województwach, byłoby dobrze, gdyby województwo najsłabsze miało relatywnie najwyższy udział (załóżmy, że zależałoby nam, by zrobić tam 20 wywiadów). Na pozostałe województwa przypadłoby w sumie 30 wywiadów, czyli niespełna 8 na województwo. W przypadku tak niewielkiej liczby badanych jednostek można pokusić się o wybranie ewaluowanych podmiotów w taki sposób, który – zdaniem badacza – będzie najlepiej odzwierciedlał przekrojową sytuację wśród respondentów. Kluczowe w tym przypadku jest sformułowanie „zdaniem badacza", gdyż po pierwsze zakłada, że badacz ma jakiekolwiek zdanie w danym zakresie, a po drugie wie, które przedsiębiorstwa będą najbardziej reprezentatywne. Taki sposób dobierania próby nazywamy doborem celowym, gdyż to badacz, a nie metoda statystyczna, wyznacza cele, które zostaną poddane badaniu.

Warto zauważyć, że w przypadku losowania próby (oraz w przypadku próby losowo-warstwowej, pod warunkiem skonstruowania wag analitycznych), możliwe jest oszacowanie błędu, którym obarczone jest każde badanie oparte na statystyce, dzięki czemu można spróbować stwierdzić, jak rzecz się ma w całej populacji, która została poddana działaniu ewaluowanego programu, ale nie została poddana samemu badaniu. W przypadku próby celowej nie ma możliwości oszacowania standardowego błędu, a zatem otrzymane wyniki nie są interpretowalne w kontekście populacji (Babbie 2003).

3. RODZAJ BADANIA

W przypadku, gdy mamy już dobraną próbę, należy się zastanowić nad tym, jaki sposób badania pozwoli najlepiej założoną próbę zrealizować. Niektórzy badacze najpierw ustalają rodzaj badania, a następnie, biorąc go pod uwagę, dopiero konstruują próbę, mając na uwadze ograniczenia związane z każdym z rodzajów badania.

Tak jak zaznaczono we wstępie, obecnie stosowane są badania PAPI, CAPI, CATI oraz badania internetowe (CAWI). Poniżej omówiono każde z nich.

Badania PAPI to badania z wykorzystaniem papierowego kwestionariusza. W przypadku projektowania badania PAPI należy więc pamiętać o wcześniejszym wydruku odpowiedniej liczby kwestionariuszy oraz ewentualnie instrukcji ankieterskich. Od badań PAPI powoli odchodzi się na całym świecie, po pierwsze – ze względów ekonomicznych (inne sposoby badania stają się coraz bardziej rentowne), po drugie – ze względów ekologicznych oraz czasowych. W obecnej sytuacji wydrukowanie kilkudziesięciu czy kilkuset egzemplarzy kwestionariusza staje się coraz mniej opłacalne. Warto także zauważyć, że to właśnie kwestionariusze z badań PAPI najbardziej narażone są na ryzyko błędów ankieterskich: po pierwsze pomylić się może sam ankieter, wpisując dane do kwestionariusza, po drugie pomylić się może osoba przepisująca dane z kwestionariusza do bazy danych (koder).

Badania PAPI są bardzo dobrym rozwiązaniem, gdy ewaluacja ma niewielki zasięg i do jej przeprowadzenia nie korzystamy z agencji badawczej, zapewniającej sprzęt komputerowy i ankieterów. Wydruk kilkunastu, a nawet kilkudziesięciu kwestionariuszy i samodzielne przeprowadzanie ewaluacji mogą być najskuteczniejszym i najszybszym sposobem osiągnięcia celu. Badania PAPI dobrze sprawdzają się również w przypadku ankiet samowypełnialnych, gdy wydrukowany dokument przesyłamy (na przykład pocztą) do jednostek, które chcemy zbadać, licząc na to, że wypełnione ankiety wrócą do nas tą samą drogą. Kwestię przygotowania i postępowania z ankietami samowypełnialnymi omówiono bardziej szczegółowo w podrozdziale dotyczącym konstrukcji kwestionariuszy.

CATI (Computer Assisted Telephone Interview) i CAPI (Computer Assisted Personal Interview) są obecnie najbardziej rozpowszechnionymi rodzajami badań. Różnica między nimi jest znaczna. W przypadku badań CATI mamy do czynienia najczęściej ze studiem CATI, wyposażonym w komputery oraz centralę telefoniczną. Przy komputerach siedzą przeszkoleni ankieterzy, którzy posiłkując się kwestionariuszem w wersji elektronicznej (tzw. skryptem), obdzwaniają wszystkie jednostki z naszej dobranej uprzednio próby. W przypadku badań CAPI mamy do

czynienia z ankieterami wyposażonymi w laptopy, zawierające skrypt kwestionariusza, którzy osobiście odwiedzają osoby, które zdecydowaliśmy się przebadać i przeprowadzają z nimi wywiady (Babbie 2003; Sztabiński 1999, 2001).

Obie te metody mają oczywiście swoje ograniczenia. Metoda CATI zakłada, że każdy z ewaluowanych obiektów ma dostęp do telefonu (co w obecnych czasach raczej jest regułą), metoda CAPI wymaga fizycznych odwiedzin ankietera w siedzibie ewaluowanego podmiotu, bądź odwiedzin u ewaluowanej jednostki. Prócz tego, oczywiście, metody CATI i CAPI obarczone są kosztami – połączenia telefonicznego bądź dojazdu, nie mówiąc już o kosztach amortyzacji sprzętu – od których wolne są badania PAPI.

O ile więc zarówno CATI, jak i CAPI mogą być droższe od PAPI w przypadku niewielkich prób, o tyle zaczynają być opłacalne w przypadku większych badań. Są również, dzięki elektronicznym kwestionariuszom, mniej podatne na błędy – ankieter po prostu nie ma możliwości wpisania wartości przez skrypt nieprzewidzianych (na przykład w pytaniu „Ile procent zasobów przeznaczonych zostało na modernizację linii produkcyjnej?" nie ma możliwości wpisania „125%", co niestety czasem zdarza się w badaniach PAPI i jest źródłem problemów interpretacyjnych).

Ostatnim z dość powszechnych sposobów przeprowadzenia badania są badania internetowe (CAWI). Warto zauważyć jednak, że do tej pory nie udało się stworzyć żadnej metodologii, która umożliwiałaby przeprowadzenie badania internetowego przy pomocy kwestionariusza. W przypadku badań internetowych wciąż najbardziej rozpowszechnionym sposobem badania są ankiety samowypełnialne, które jednak mają dość poważne ograniczenia. Ta „ułomność" internetu bierze się z tego, że o ile w przypadku badań telefonicznych dzwonimy do konkretnej osoby i możemy z nią przeprowadzić wywiad w danej chwili, zaś w przypadku badania PAPI i CAPI ankieter nawiązuje fizyczny kontakt z respondentem, o tyle w przypadku badań internetowych nie ma możliwości przewidzenia, kiedy dany respondent jest akurat online. Możliwości nawiązania kontaktu i przeprowadzenia badania przez przeszkolonego ankietera są więc mocno ograniczone w przypadku badań internetowych. Bolączką ankiet internetowych jest także dość niski zwrot i brak możliwości przewidzenia profilu osób zwracających ankiety. W efekcie może się okazać, że nie uzyskano w ogóle ankiet od jednostek z jakiś względów najbardziej interesujących, czyli na przykład od tych, które mają najwięcej problemów bądź są najbardziej efektywne w danym zakresie (osoby, które są efektywne, zazwyczaj nie rozpraszają się na wypełnianie ankiet).

Samowypełnialne ankiety internetowe są jednak zdecydowanie najtańszą i najmniej czasochłonną z przedstawionych technik, gdyż nie

wymagają ani inwestycji sprzętowych, ani inwestycji ludzkich w postaci przeszkolonych ankieterów. Samowypełnialne ankiety internetowe dość dobrze sprawdzają się w przypadku bardzo dużych i mało skomplikowanych ewaluacji oraz w przypadku, gdy badana grupa jest dość homogeniczna (Sztabiński 2004).

Rodzaj badania nie determinuje jeszcze wszystkiego. Prócz tego możemy mieć do czynienia z badaniami ***ad hoc*** (czyli badaniami jednorazowymi, zarządzonymi naprędce) bądź badaniami **trackingowymi** (czyli badaniami wykonywanymi cyklicznie, w celu sprawdzenia zmian w czasie). Szczególnym przypadkiem badania trackingowego jest badanie panelowe.

Badaniem panelowym nazywamy badanie, które jest przeprowadzane cyklicznie na tej samej próbce osób bądź firm. W badaniu panelowym korzysta się zazwyczaj z tego samego kwestionariusza (rzadko dodając bądź odejmując pytania, unikając zaś modyfikacji pytań). Celem badania panelowego jest zbadanie zmian w czasie tej samej grupy osób bądź firm. Badania panelowe mają bardzo poważne ograniczenia, a ich przeprowadzenie wymaga czasu oraz konsekwencji.

Badanie panelowe powinno być przeprowadzane cyklicznie, na, w miarę możliwości, niezmienianym kwestionariuszu, wśród dokładnie tych samych osób. Zachowanie cyklu, kwestionariusza oraz powtarzalności próby jest niezbędne do tego, by móc właściwie wysnuwać wnioski. Badania panelowe znakomicie sprawdzają się w przypadku ewaluacji ciągłych, mających na celu zbadanie, w jaki sposób przebiega proces, na przykład aplikowania o środki czy wdrażania projektu. Badanie panelowe pozwala na wychwycenie „wąskich gardeł" i najczęściej pojawiających się problemów w realizacji zadań (Karpiński 2006).

Badania panelowe cieszą się niezasłużoną niesławą ze względu na to, że są „nudne" (stosuje się za każdym razem ten sam kwestionariusz) oraz czasochłonne (trzeba wykonać wiele pomiarów, by móc powiedzieć cokolwiek o procesie). W ich wyniku jednak otrzymujemy wyjątkowej jakości dane, niedostępne w najczęściej wykorzystywanych badaniach *ad hoc*.

4. KWESTIONARIUSZ

Kwestionariusz jest narzędziem badawczym, za pomocą którego zbierane są dane dotyczące badanych osób bądź instytucji (zob.: Noelle 1976; Sawiński 1992; Babbie 2003). Kwestionariusz jest konstruowany na podstawie celów badania, sformułowanych zazwyczaj przez zamawiającego badania w postaci briefu (autor jest zmuszony użyć tego określenia ze względu na brak polskiego odpowiednika). Brief zawiera wszystkie pytania badawcze, na które zamawiający chciałby uzyskać odpowiedź.

Zazwyczaj kwestionariusz składa się z kilku części: wstępu, kwestionariusza rekrutacyjnego, kwestionariusza właściwego oraz metryczki.

We wstępie zawarte są informacje na temat samego badania, zamawiającego badanie, przeprowadzającego badanie oraz szacowanej długości trwania wywiadu. W przypadku tematów drażliwych często dodaje się formułę tłumaczącą, dlaczego szczere i nieskrępowane odpowiedzi respondenta są szczególnie ważne. Czasami we wstępie pomija się informację na temat zamawiającego badanie, po to, by nie sugerować respondentom dalszych odpowiedzi.

Kwestionariusz rekrutacyjny jest stosowany w momencie, gdy istnieje szansa na to, że respondent, którego chcemy przebadać, nie spełnia założonych warunków badania. Zazwyczaj, w przypadku badań ewaluacyjnych, mamy do czynienia z wyjściową bazą danych jednostek, które były rzeczywiście poddane danemu projektowi. Minimalizuje to szansę takiego zdarzenia, w wyniku którego wywiad zostanie przeprowadzony z osobą fizyczną bądź przedstawicielem instytucji, która nie brała udziału w ewaluowanym projekcie. Niemniej jednak, warto jest na początku budowanego kwestionariusza zadać pytania bezpośrednio związane z badaniem, by zweryfikować, czy osoba, z którą przeprowadzany jest wywiad, na pewno ma odpowiednie kompetencje, by wraz z ankieterem kwestionariusz wypełnić. W przypadku rezygnacji z kwestionariusza rekrutacyjnego narażamy się na otrzymanie danych wątpliwej jakości: rzadko bowiem się zdarza, by ankieter sam przerwał wywiad, gdy na podstawie odpowiedzi respondenta doszedł do wniosku, że rozmawia z niewłaściwą osobą. Zazwyczaj taki bezwartościowy, z punktu widzenia badacza, wywiad jest z obopólnym trudem doprowadzany do końca, zaś jego wartość merytoryczna jest znikoma i w efekcie najczęściej należy go po prostu z ostatecznej bazy danych wyrzucić.

Kwestionariusz rekrutacyjny powinien składać się z pytań sprawdzających fakty związane z ewaluowanym projektem, takie jak: potwierdzenie uczestnictwa w projekcie, wysokość otrzymanych dotacji, okres, w którym zostały przyznane dotacje. Na pytania takie z łatwością uzyskamy odpowiedzi od jednostek poddanych ewaluacji, dzięki czemu będzie można już na wstępie „odsiać" tych respondentów, którzy bądź to z powodu słabej jakości bazy z której losowano próbę, bądź to z powodu błędów ludzkich (na przykład błędów ankieterskich) nie powinni w badaniu uczestniczyć.

Drugą częścią kwestionariusza jest kwestionariusz zasadniczy, zawierający wszystkie te pytania, które chcemy zadać respondentowi. Pytania te mogą dotyczyć zarówno jego wiedzy faktycznej, jak i przewidywań. Pytania zasadniczo dzielimy na zamknięte i otwarte. Jeżeli w trakcie konstruowania kwestionariusza zorientujemy się, że większość pytań ma charakter otwarty, należy zastanowić się nad zmianą sposobu badania i rozważyć przeprowadzenie badania o charakterze eksploracyjnym, co najczęściej oznacza po prostu badanie jakościowe.

Pytania zamknięte to pytania, na które respondent może odpowiedzieć wyłącznie w sposób przewidziany w kwestionariuszu (przykład 1). Pytania otwarte to pytania, na które respondent może odpowiedzieć swoimi słowami, w dowolny sposób (przykład 2). Zadaniem ankietera w przypadku pytania otwartego jest dokładne zapisanie wypowiedzi respondenta; w przypadku pytania zamkniętego jest zaznaczenie odpowiedzi wybranej przez respondenta.

Przykład 1

(Ankieter: Odczytać kategorie. Zaznaczyć odpowiedź respondenta. Możliwa tylko jedna odpowiedź)

A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?

1. zdecydowanie nie będziemy aplikować → zadać pytanie A2
2. raczej nie będziemy aplikować → zadać pytanie A2
3. nie wiem/trudno powiedzieć
4. raczej będziemy aplikować
5. zdecydowanie będziemy aplikować

Przykład 2

(Ankieter: Zadać tylko tym respondentom, którzy w A1 odpowiedzieli 1 lub 2. Zapisać dokładnie wypowiedź respondenta)

A2. Proszę powiedzieć, dlaczego nie będą Państwo aplikować o środki z funduszu X?

.....

.....

Jak widać, pytania składają się kilku części. Pokrótce omówimy wszystkie z nich.

Przykład 1

(Ankieter: Odczytać kategorie. Zaznaczyć odpowiedź respondenta. Możliwa tylko jedna odpowiedź)

A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?

1. zdecydowanie nie będziemy aplikować → zadać pytanie A2
2. raczej nie będziemy aplikować → zadać pytanie A2
3. nie wiem/trudno powiedzieć
4. raczej będziemy aplikować
5. zdecydowanie będziemy aplikować

Wyróżniono pole „do ankietera". Kwestionariusz służy nie tylko temu, by uzyskać odpowiedzi od respondenta; jest również narzędziem, za pomocą którego badacz komunikuje się z ankieterem, podając mu wytyczne dotyczące sposobu zadawania pytania.

Przykład 1

(Ankieter: Odczytać kategorie. Zaznaczyć odpowiedź respondenta. Możliwa tylko jedna odpowiedź)

A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?

1. zdecydowanie nie będziemy aplikować → zadać pytanie A2
2. raczej nie będziemy aplikować → zadać pytanie A2
3. nie wiem/trudno powiedzieć
4. raczej będziemy aplikować
5. zdecydowanie będziemy aplikować

Wyróżniono pole pytania. Jest to ta część, za pomocą której ankieter komunikuje się z respondentem. Zazwyczaj jest to jedyna część pytania, którą słyszy respondent. W tym przypadku jest inaczej, gdyż zawarliśmy w polu „do ankietera", by ten ostatni odczytał kategorie.

Przykład 1

(Ankieter: Odczytać kategorie. Zaznaczyć odpowiedź respondenta. Możliwa tylko jedna odpowiedź)

A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?

1. zdecydowanie nie będziemy aplikować → zadać pytanie A2
2. raczej nie będziemy aplikować → zadać pytanie A2
3. nie wiem/trudno powiedzieć
4. raczej będziemy aplikować
5. zdecydowanie będziemy aplikować

Wyróżniono pole odpowiedzi (tzw. kafeterię). Za pomocą takiej skali respondent będzie udzielać odpowiedzi na zadane pytanie. Warto zwrócić uwagę na tekst przy punkcie 1 i 2 („Zadać pytanie A2"). To również tekst do ankietera, mówiący mu, w jakich przypadkach zadać ma pytanie A2. Zastosowana skala, za pomocą której respondent odpowiadać będzie na pytanie, jest przykładowa. W dalszej części omówiono rodzaje najczęściej stosowanych skal wraz z ich zastosowaniami.

Przykład 2

(Ankieter: Zadać tylko tym respondentom, którzy w A1 odpowiedzieli 1 lub 2. Zapisać dokładnie wypowiedź respondenta)

A2. Proszę powiedzieć, dlaczego nie będą Państwo aplikować o środki z funduszu X?

.....

.....

Wyróżniono pole swobodnej wypowiedzi. W kwestionariuszach PAPI trzeba zostawić miejsce na zapisanie wypowiedzi respondenta; w przypadku badań CATI i CAPI wystarczy napisać, że jest to pytanie otwarte i przewidzieć w skrypcie możliwość zapisania przez ankietera dokładnej wypowiedzi respondenta.

Należy zwrócić uwagę, by pytania otwarte zadawać tylko w ostateczności, gdyż nie są one kwantyfikowalne i nie można użyć ich w analizie ilościowej. Odpowiedzi z pytania otwartego mają charakter subiektywny i trudno ocenić, jaka jest siła wypowiedzi poszczególnych respondentów. Innymi słowy, pytania otwarte są ostatecznością w badaniach ilościowych i zdecydowanie powinno się ich unikać. Jednym ze sposobów radzenia sobie z pytaniami otwartymi jest ich prekodowanie. Prekodowane pytanie otwarte przedstawia przykład 3.

Przykład 3

(Ankieter: Zadać tylko tym respondentom, którzy w A1 odpowiedzieli 1 lub 2. Nie czytać odpowiedzi. Zaklasyfikować odpowiedź respondenta)

A2. Proszę powiedzieć, dlaczego nie będą Państwo aplikować o środki z funduszu X?

1. nie będziemy mieć takiej potrzeby
2. nie ma komu napisać projektu
3. z powodu biurokracji
4. będziemy aplikować o inne fundusze
5. inne (jakie?)

.....

.....

Warto zwrócić uwagę na to, że nie zmieniła się treść pytania, w stosunku do przykładu 2, zmieniła się natomiast kafeteria odpowiedzi oraz instrukcja dla ankietera. Nadal jest miejsce na swobodną wypowiedź respondenta, jeśli udzieli on odpowiedzi, której ankieterowi nie uda się

zaklasyfikować do żadnego z kodów, zakładanych przez badacza. Konstruowanie pytań prekodowanych wymaga przynajmniej podstawowej wiedzy badacza na dany temat: należy wszak trafnie przewidzieć przynajmniej część możliwych odpowiedzi, których udzielą respondenci. Tak przygotowane pytanie wymaga szczególnej atencji: badacz *post factum*, już na poziomie bazy danych, powinien sprawdzić, czy jakaś odpowiedź, która została zapisana jako „inna", nie może zostać zapisana za pomocą zaproponowanego klucza kodowego.

Taki sposób radzenia sobie z pytaniami otwartymi po pierwsze przyspiesza kwestionariusz (zapisywanie dokładnych wypowiedzi zajmuje więcej czasu niż zaznaczanie odpowiedzi), po drugie zaś pozwala, przynajmniej w części przypadków, zawęzić, już na poziomie kwestionariusza, liczbę możliwych odpowiedzi.

W przypadku zapisywania pytań zamkniętych należy zastanowić się nad skalą, której użyjemy w treści pytania. Wybór skali zdeterminuje późniejsze analizy, które będzie można przeprowadzić na bazie danych. Przykładowe skale podano poniżej:

Skala 1

A1. Proszę powiedzieć czy słyszał(a) Pan/Pani o funduszu X?

1. tak
2. nie
3. *nie wiem/trudno powiedzieć*

W powyższym przykładzie skala ma charakter dychotomiczny. Możliwe jest również, że respondent nie przypomni sobie o funduszu X, o którym jest mowa w pytaniu. W takim przypadku badacz dopuszcza odpowiedź „nie wiem", ale ankieter jej nie odczytuje (dlatego jest zapisana kursywą).

Skala 2

A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?

1. zdecydowanie nie będziemy aplikować → *zadać pytanie A2*
2. raczej nie będziemy aplikować → *zadać pytanie A2*
3. nie wiem/trudno powiedzieć
4. raczej będziemy aplikować
5. zdecydowanie będziemy aplikować

To skala symetryczna, która została zastosowana w przykładzie 1. Warto zauważyć, że środkiem tej skali jest odpowiedź „nie wiem/trudno

powiedzieć", której zazwyczaj się nie odczytuje. W związku z tym częściej stosuje się inną, choć pokrewną skalę.

Skala 2b

A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?

1. zdecydowanie nie będziemy aplikować → *zadać pytanie A2*
2. raczej nie będziemy aplikować → *zadać pytanie A2*
3. raczej będziemy aplikować
4. zdecydowanie będziemy aplikować
5. *nie wiem/trudno powiedzieć*

Tutaj odpowiedź, która nic nam nie mówi, z punktu widzenia analitycznego, została wyrzucona poza nawias. Oznaczenie kursywą sygnalizuje, że tej odpowiedzi ankieter nie odczytuje. Zarówno skala 2, jak i 2b, są z punktu widzenia metodologii poprawne i mogą być stosowane zamiennie (Sztabiński 2002).

Jak Czytelnik zapewne zauważył, odpowiedź „nie wiem/trudno powiedzieć", jeśli nie została potraktowana jako środek skali, otrzymała kod „9". Wiąże się to ze zwyczajem, w którym wartości „puste" otrzymują kody odległe od reszty wskazań. W przypadku kafeterii dłuższych, przekraczających 10 pozycji, dla odpowiedzi typu: „nie wiem", „trudno powiedzieć", „odmowa odpowiedzi", „nie dotyczy", rezerwuje się kody 96–99. W przypadku jeszcze dłuższych kafeterii ostatnie – kody trzy- lub czterocyfrowe. Dzięki takiej numeracji można „fizycznie" oddzielić te elementy kafeterii, które są nieodczytywane przez ankietera od rzeczywistych odpowiedzi na pytanie. Stosowanie takich cezur znacznie ułatwia potem analizę danych.

Przy wybieraniu skal do kwestionariusza trzeba pamiętać o tym, by zachować między nimi spójność. Możliwe jest posługiwanie się skalami odwróconymi w stosunku do wyżej przedstawionych (w których na pierwszym miejscu jest najsilniejsze stwierdzenie pozytywne, zaś na ostatnim – najsilniejsze negatywne), jednak zastosowanie w jednym kwestionariuszu wymieszanych skal, ułożonych raz tak, raz inaczej, wprowadza tylko niepotrzebny chaos i zwiększa możliwość popełnienia błędu przez ankietera, przez respondenta, czy w końcu przez samego badacza. Należy więc już na początku podjąć decyzję o orientacji skal w pytaniach zamkniętych i konsekwentnie tę orientację stosować.

Przyjmuje się, że nieco lepsze są te skale, które zorientowane są rosnąco (od najbardziej negatywnego stwierdzenia, do najbardziej pozytywnego), ze względu na naturalne skojarzenia ze skalą szkolnych ocen. Odwrócone skale (od najbardziej pozytywnego do najbardziej negatywnego stwierdzenia) nie stanowią jednak błędu i, jeżeli są stosowane

konsekwentnie, swobodnie mogą być wykorzystywane do budowy kwestionariusza.

Odrębnym rodzajem skal są skale niewerbalne (skala 4), bądź quasi-niewerbalne. Przykładem skali quasi-niewerbalnej jest skala 3.

Skala 3

A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?

1. zdecydowanie nie będziemy aplikować → *zadać pytanie A2*
2. … → *zadać pytanie A2*
3. …
4. …
5. zdecydowanie będziemy aplikować

Jak widać, w tym przypadku, zostały zachowane tylko krańce skali, zaś odpowiedzi 2–4 zostały wykropkowane. Użycie takiej skali uprawnia do późniejszego policzenia z niej średniej; zakłada się, że liczenie średniej ze skal określonych słownie jest niepoprawne metodologicznie. Niektórzy badacze odrzucają możliwość liczenia średnich (i innych miar rozproszenia) nawet ze skal quasi-niewerbalnych.

Skala 4

(Ankieter: Zaznaczyć odpowiedź respondenta. Możliwa tylko jedna odpowiedź)

A1. Proszę, na skali od 1 do 10, ocenić swoje zaangażowanie w działalność zawodową, gdzie 1 oznacza bardzo niskie zaangażowanie, zaś 10 – bardzo wysokie.

1	2	3	4	5	6	7	8	9	10

Zaletą takiej skali jest po pierwsze wielość kategorii, za pomocą których respondent może dokonać autoewaluacji, a poza tym umożliwia stosowanie bardziej zaawansowanych statystyk, w rodzaju średniej, odchylenia standardowego czy współczynników korelacyjnych.

Ostatnią częścią kwestionariusza jest metryczka, umieszczana zazwyczaj na końcu kwestionariusza. W metryczce powinny znaleźć się wszystkie te pytania, które nie dotyczą bezpośrednio badania, a które zwykle służą zebraniu danych demograficznych. Do pytań metryczkowych należą w przypadku badań osób fizycznych: pytania o płeć, wiek, status zawodowy, stan cywilny, wielkość miejscowości zamieszkania, dochód, samoocenę sytuacji materialnej. W przypadku instytucji przydatne są informacje o samoocenie sytuacji przedsiębiorstwa, obrocie, przychodzie czy liczbie zatrudnionych pracowników.

Pytania metryczkowe pozwalają niejednokrotnie na skonstruowanie wag analitycznych, umożliwiających ekstrapolację wyników na populację, czy „wyprostowanie" próby.

Na koniec, gdy kwestionariusz jest już gotowy, trzeba zwrócić uwagę na trzy rzeczy. Po pierwsze kwestionariusz powinien być logicznie ułożony; zazwyczaj oznacza to takie ułożenie pytań, by przechodziły od ogółu do szczegółu. Przykład 4 obrazuje przykładowe właściwe ułożenie pytań.

Przykład 4

Czy słyszał(a) Pan(i) o Unii Europejskiej?

Czy słyszał(a) Pan(i) o funduszach strukturalnych w ramach Unii Europejskiej?

Czy słyszał(a) Pan(i) o funduszu strukturalnym X?

Czy korzystał(a) Pan(i) kiedykolwiek z funduszu strukturalnego X?

...

W tym przypadku mamy do czynienia z właściwym ułożeniem pytań kwestionariuszowych. Każde następne pytanie jest uszczegółowieniem pytania poprzedniego i trudno sobie wyobrazić, by ktoś, kto słyszał o funduszach strukturalnych UE, nie słyszał o samej UE i tak dalej. Od tej ogólnej reguły istnieje tylko jeden poważny wyjątek i dotyczy pytań, które mogą być dla respondenta szczególnie drażliwe: w takim przypadku warto rozważyć złamanie logiki kwestionariusza i umieszczenie takich pytań na samym końcu, tuż przed metryczką. Często się bowiem zdarza, że respondent nie odpowie na takie pytanie, a nawet odmówi w ogóle współpracy z ankieterem. W przypadku umieszczenia takiego pytania w środku, bądź zgoła na początku kwestionariusza, ryzykujemy utratę większości odpowiedzi respondenta; zalet umieszczenia takich pytań na końcu chyba nie trzeba bliżej omawiać. Do pytań drażliwych zaliczamy m.in.:

- pytania dotyczące osobistej sfery życia,
- pytania dotyczące układów personalnych i towarzyskich,
- pytania dotyczące spraw społecznie napiętnowanych, jak na przykład pedofilia czy korupcja,
- pytania o ocenę działań konkretnych jednostek.

Należy pamiętać o numeracji pytań. Jeżeli, dla przykładu, kwestionariusz zawiera pytania o: Unię Europejską jako taką, fundusze EFS oraz fundusze strukturalne, to w kwestionariuszu również powinny być trzy bloki pytań, np.: A1, A2, A3, ... (dla pytań o UE), B1, B2, B3, ... (dla

pytań o EFS) oraz C1, C2, C3, ... (dla pytań o fundusze strukturalne). Nie jest to oczywiście warunek konieczny, jednak po pierwsze umożliwia sprawną nawigację w obrębie samego kwestionariusza, co nie jest bez znaczenia, po drugie zaś, determinuje później strukturę bazy danych, co ma niebagatelne znaczenie dla późniejszych analiz[2].

Trzecią kwestią, zazwyczaj wprowadzaną w ostatniej chwili, jest kwestia filtrów. Niektóre pytania nie powinny być zadawane wszystkim respondentom. Pytanie, na przykład, o powody chęci skorzystania z funduszu X w przyszłości powinno być zadane tylko tym respondentom, którzy zadeklarowali chęć skorzystania z tego funduszu. Wówczas trzeba zaznaczyć, że, przykładowo, pytanie A2 jest zadawane tylko określonej grupie respondentów. Filtry do kwestionariusza wprowadza się w ostatniej chwili, ze względu na to, że czasami logika badania wymaga przesunięcia jakiegoś pytania w inne miejsce. Wówczas powinno ono także zmienić swój numer, a co za tym idzie, również filtr mu przypisany, zmieni swoją treść.

Podsumowując: kwestionariusz jest podstawowym narzędziem w badaniu ilościowym. Od jego konstrukcji zależy jakość otrzymanych danych. Kwestionariusz z pomieszanymi skalami, zdublowanymi pytaniami, dużą liczbą pytań otwartych oraz nielogicznie ułożony, da w rezultacie słabej jakości bazę danych, a co za tym idzie, wnioski wyciągnięte na podstawie otrzymanych wyników będą mało nośne, błędne lub wysnute na wyrost (Noelle 1976; Sawiński 1992; Frankfort-Nachmias; Nachmias 2001; Karpiński 2006).

5. RAPORT I WNIOSKOWANIE

W wyniku przeprowadzenia badania kwestionariuszowego na badanych podmiotach, otrzymujemy bazę danych (zazwyczaj w formie plików Excela bądź jakiegoś pakietu statystycznego, np. SPSS), z której liczone są następnie interesujące nas statystyki. Opis szczegółowych technik statystycznych (zwłaszcza tych, do których niezbędne jest posiadanie pakietu statystycznego) daleko wykracza poza ramy tego podręcznika, omówię więc tylko najczęściej używane wskaźniki służące skonstruowaniu raportu.

---

[2] Łatwo jest zbagatelizować koncepcję logicznego numerowania bloków pytań; problem pojawia się w momencie, gdy do badania przygotowano kilkanaście wersji kwestionariusza (czasem bywa ich nawet kilkadziesiąt). Ponadto wyniki z badania zazwyczaj pojawiają się wiele miesięcy po skonstruowaniu narzędzia badawczego. Autor wielokrotnie zetknął się z przypadkiem, gdy osoba, która pisała kwestionariusz, była nieobecna lub nie potrafiła odnaleźć ostatecznej wersji i tylko logika budowy kwestionariusza ratowała sytuację.

5.1. Rozkład częstości jednej zmiennej

Podstawową statystyką jest rozkład częstości jednej zmiennej. Jeśli chcemy się dowiedzieć, jaki odsetek firm korzystał z funduszu, powinniśmy użyć tej statystyki. Może być ona przedstawiona za pomocą liczb, może być również procentowana. W przypadku wyboru wartości procentowych warto zaznaczyć, co jest podstawą procentowania (informację tę zamieszcza się zazwyczaj na dole tabeli). Zazwyczaj będzie to cała badana populacja, zdarzają się jednak przypadki, gdy podstawą nie są wszyscy badani, a tylko jakiś ich wycinek. Może to być związane z filtrowaniem (pytanie nie było zadawane jakiejś grupie respondentów, więc trudno, by stanowili oni część podstawy oprocentowania), bądź z występowaniem dużej liczby odpowiedzi odmownych[3]. W takim przypadku należy się zastanowić nad tym, czy pewnych przypadków nie wyłączyć z podstawy oprocentowania, by pokazać różnice. Obrazuje to przykład 5.

Przykład 5

	X1	X2
A1. Czy w przyszłym roku będą Państwo aplikować o środki z funduszu X?		
1. zdecydowanie nie będziemy aplikować	8%	18%
2. …	12%	27%
3. …	10%	22%
4. …	9%	20%
5. zdecydowanie będziemy aplikować	6%	13%
6. *nie wiem/trudno powiedzieć*	55%	–
Podstawa oprocentowania (N =)	600	270

W przykładzie 5 w kolumnie X2 wyłączyłem zbyt liczną grupę, która nie potrafiła odpowiedzieć, czy będzie aplikować o środki z danego funduszu. W rezultacie, jak widać, zmniejszyłem podstawę oprocentowania (z 600 do 270).

Oczywiście, wyniki z kolumny X1 są równie prawdziwe, jak te z kolumny X2; różnią się jedynie interpretacją. Podczas gdy dane z kolumny

[3] Każdy badacz staje w pewnym momencie przed dylematem, nazywanym przeze mnie „dylematem dochodu". Dylemat dochodu dotyczy sytuacji, gdy mamy do czynienia z wyjątkowo dużą liczbą odmów odpowiedzi, bądź braków danych. Wówczas procentowanie niewiele nam mówi, gdyż mamy cztery kategorie, z czego jedna, dla przykładu, stanowi 45% populacji. Pozostałe kategorie mają do zagospodarowania jedynie pozostałe 55%, w związku z tym jakiekolwiek różnice stają się niewidoczne. Nazwałem tę sytuację dylematem dochodu, gdyż badani zazwyczaj bardzo nie lubią mówić ankieterom, ile naprawdę zarabiają.

X1 interpretujemy jako: „spośród wszystkich badanych, 8% firm odpowiedziało, że zdecydowanie nie będzie aplikować", dane z kolumny X2 mówią nam o tym, że „spośród tych firm, które potrafiły ocenić swoje uczestnictwo w przyszłych edycjach, 18% odpowiedziało, że zdecydowanie nie będzie aplikować". Ten drugi sposób wydaje się bardziej zawikłany, jednak umożliwia nam opisywanie tylko tej grupy firm, która jest zdecydowana podjąć jakieś działanie – co może być (choć oczywiście nie musi) być bardziej interesujące z punktu widzenia celów badania. Umożliwia także bezpośrednie porównanie liczby firm, które są zdecydowane nie aplikować, z tymi, które są zdecydowane na to, by pozyskać środki z funduszu X. W populacji wszystkich zbadanych firm różnica ta wynosi 2% (8% – 6%), wydaje się więc niewielka. Jednak w populacji firm zdecydowanych podjąć jakieś działanie różnica ta wynosi aż 5% (18%–13%).

Z prostego rozkładu jednej zmiennej możemy uzyskać także coś więcej niż częstości. Tak jak wspomniałem, niektóre skale pytań nadają się do tego, by policzyć z nich średnie. W przypadku wartości średnich nie mamy możliwości manewru: wszystkie dane, które nie są liczbowymi odpowiedziami na pytanie, muszą być wyrzucone z analizy liczenia średniej. W programach statystycznych zazwyczaj używa się komendy, która czyni te odpowiedzi brakami danych. W przypadku danych z przykładu 5 uzyskalibyśmy średnią **2,84** liczoną oczywiście z 270 przypadków. Jeżeli zauważymy, że środkiem skali jest wartość 3, taka a nie inna wartość średniej pozwoli nam wysnuć wniosek, że badane firmy są raczej niechętne aplikowaniu o środki z funduszu X.

Warto przy okazji zauważyć, że gdybyśmy nie wyrzucili z liczenia średniej wartości „9", otrzymalibyśmy średnią na poziomie **6,23**, co przekracza wartości, które przyjmuje skala i jest zupełnie nieinterpretowane – stanowi jeden z najczęściej występujących błędów w przypadku liczenia średnich.

Oprócz wartości średniej w przypadku rozkładów jednej zmiennej stosuje się wskaźniki, które, jak do tej pory, nie doczekały się polskich odpowiedników, określa się je angielskimi terminami jako ***top box*** bądź ***low box***. Określenie *top box* zazwyczaj stosuje się do najwyższego pozytywnego stwierdzenia z danego pytania, określenie *low box* – do najniższego. W przytoczonym przykładzie 5, *top box* wynosiłby 6% ogółu wszystkich firm, zaś *low box* – 8%.

Prócz tych dość prostych wskaźników można wykorzystywać **top2boxy** i **low2boxy**, a przy szerszych skalach również **top3boxy** i **low3boxy**[4]. *Top2box* z naszego przykładu wyniósłby **15%** (6% + 9%), zaś low2box – **20%** (8% + 12%).

---

[4] *Top3boxy* i *low3boxy* możliwe są do zastosowania w przypadku skal przynajmniej siedmiostopniowych, gdzie mamy do czynienia z trzema stwierdzeniami pozytywnymi i trzema negatywnymi.

Zebrawszy to wszystko, otrzymujemy zestaw statystyk dla jednej zmiennej przedstawiony w tabeli 1.

Tabela 1.

Top box	**6%**	Top2box	**15%**	Średnia	**2,84**
Low box	**8%**	Low2box	**20%**		
Różnica (top – low)	*–2%*		*–5%*		
N =	*600*		*600*		*270*

Mając taką ilość danych, możemy stwierdzić, że liczba firm, które chcą aplikować o środki z funduszu X jest nieco niższa niż tych, które nie będą aplikować, co zdają się potwierdzać wszystkie przytoczone statystyki.

Należy pamiętać, że w powyższym przykładzie różne są podstawy poszczególnych statystyk i o ile *top* i *low boxy* zostały policzone na próbie wszystkich firm, o tyle średnia uwzględnia tylko te, które potrafiły udzielić konkretnej odpowiedzi na zadane pytanie. Gdybyśmy chcieli, z jakiś względów, skupić się tylko na tych firmach, które udzieliły odpowiedzi 1–5, tabela wyglądałaby nieco inaczej, choć ogólny wniosek byłby rzecz jasna ten sam (zob. tabela 2).

Tabela 2.

Top box:	**13%**	Top2box:	**33%**	Średnia	**2,84**
Low box:	**18%**	Low2box:	**45%**		
Różnica (top – low)	*–5%*		*–12%*		
N=	*270*		*270*		*270*

5.2. Tabele krzyżowe

Dzięki tabelom krzyżowym możemy dowiedzieć się, jaka jest wzajemna zależność zmiennych. Zazwyczaj w tabelach krzyżowych mamy uwzględnione liczebności oraz jakiś rodzaj procentowania. W przypadku prostych tabel krzyżowych mamy do czynienia z procentowaniem „do ogółu", z procentowaniem w wierszach i procentowaniem w kolumnach. Przykład 6a obrazuje rozkład liczebności, w przykładach 6b–6d przedstawiono każdy z rodzajów procentowania.

Przykład 6a przedstawia rozkład częstości dwóch zmiennych. Uwzględnia ona zarówno chęć aplikowania o środki, jak i wielkość badanej firmy. W takiej formie tabela niesie ze sobą, co prawda, podstawowe

Przykład 6a

	A2. Wielkość firmy			*suma*
	mała	średnia	duża	
A1. Chęć aplikowania o środki				
1. zdecydowanie nie będziemy aplikować	10	8	12	30
2. raczej nie będziemy aplikować	20	12	8	40
3. raczej będziemy aplikować	12	10	8	30
4. zdecydowanie będziemy aplikować	7	8	5	20
suma	49	38	33	120

Przykład 6b

	A2. Wielkość firmy			*suma*
	mała	średnia	duża	
A1. Chęć aplikowania o środki				
1. zdecydowanie nie będziemy aplikować	8%	7%	10%	25%
2. raczej nie będziemy aplikować	17%	10%	7%	33%
3. raczej będziemy aplikować	10%	8%	7%	25%
4. zdecydowanie będziemy aplikować	6%	7%	4%	17%
suma	41%	32%	28%	100%

informacje, jednak dużo więcej będzie można wywnioskować, jeśli użyjemy któregoś ze sposobów procentowania.

Powyższe procentowanie jest procentowaniem do ogółu. Nie wyróżnia ono żadnej z dwu zmiennych użytych do zbudowania tabel krzyżowych. Z przykładu 6b możemy wywnioskować chociażby, że:

- wśród wszystkich zbadanych firm jest 25% takich, które zdecydowanie nie będą aplikować o środki (pole szare),
- wśród wszystkich zbadanych firm najmniej chętne do aplikowania o środki są duże firmy (pole białe),
- wśród wszystkich badanych małe firmy, które nie będą aplikowały o środki, stanowią 8% (pole czarne).

Wszystkie przytoczone wnioski są wnioskami na poziomie ogólnym. Innym sposobem spojrzenia na dane, jest wybranie procentowania w wierszach bądź kolumnach. W każdym z tych przypadków wybieramy którąś ze zmiennych użytych w tabeli krzyżowej jako podstawę oprocentowania.

Przykład 6c

	A2. Wielkość firmy			*suma*
	mała	średnia	duża	
A1. Chęć aplikowania o środki				
1. zdecydowanie nie będziemy aplikować	20%	21%	36%	25%
2. raczej nie będziemy aplikować	41%	32%	24%	33%
3. raczej będziemy aplikować	24%	26%	24%	25%
4. zdecydowanie będziemy aplikować	14%	21%	15%	17%
suma	100%	100%	100%	100%

W przypadku przedstawionym w przykładzie 6c możemy stwierdzić, że:

- wśród zbadanych małych firm, było 20% takich, które stwierdziły, że zdecydowanie nie będą aplikować o środki (pole czarne),
- wśród zbadanych firm, to średnie firmy były najbardziej przychylnie nastawione do zdobywania funduszy (pole szare),
- jedynie 15% dużych firm było zdecydowanych aplikować o środki (pole białe).

Proszę zwrócić uwagę, że w tym przypadku jako podstawę oprocentowania wybraliśmy wielkość firmy (czyli, w tym przypadku, wybraliśmy procentowanie w kolumnach). Wszystkie wnioski dotyczą więc wielkości firm i porównań między nimi. Odmienny punkt widzenia prezentuje procentowanie w wierszach (przykład 6d).

Przykład 6d

	A2. Wielkość firmy			*suma*
	mała	średnia	duża	
A1. Chęć aplikowania o środki				
1. zdecydowanie nie będziemy aplikować	33%	27%	40%	100%
2. raczej nie będziemy aplikować	50%	30%	20%	100%
3. raczej będziemy aplikować	40%	33%	27%	100%
4. zdecydowanie będziemy aplikować	35%	40%	25%	100%
suma	41%	32%	28%	100%

Z przykładu 6d, przedstawiającego procentowanie w wierszach, możemy wyciągnąć następujące wnioski:

- 40% firm, które nie będą aplikować o środki to duże firmy (pole białe)
- połowa firm, które raczej nie będą aplikować o środki, to firmy małe (pole czarne)
- cztery na dziesięć firm, które zdecydowanie będą aplikować o środki, to firmy średnie (pole szare)

Jak widać z przytoczonych przykładów, należy uważnie dobierać rodzaj procentowania, od niego bowiem w dużej mierze zależy zasadność wysuwanych wniosków. Każdy z powyżej zastosowanych sposobów przedstawienia danych jest poprawny; od pytania badawczego, na które zamierzamy odpowiedzieć, zależy wybór odpowiedniego narzędzia. Inaczej będzie wyglądać odpowiedź na pytanie „Firmy jakiej wielkości najczęściej będą aplikować o środki?", inaczej będziemy musieli podejść do pytania „Ogólnie rzecz biorąc, które firmy są najbardziej niechętne aplikowaniu o środki?".

Rozkłady częstości jednej zmiennej oraz tabele krzyżowe stanowią w badaniach ewaluacyjnych ponad 90% wszystkich analiz i statystyk[5]. Pozostałe 10% to statystyki dedykowane do poszczególnych badań, trudne do opisania bez odwoływania się do konkretnych projektów. W przypadku analizowania rozkładów częstości i tabel krzyżowych powinniśmy móc odpowiedzieć na pytania:

- Jak dokładnie brzmiało pytanie, na które odpowiada tabela bądź wykres?
- Jaka jest podstawa oprocentowania? Czy jest to cała próba, czy też nie? Jeśli nie, to czy wykluczono braki danych, czy też był założony filtr na poziomie kwestionariusza?
- Jeżeli mamy do czynienia z procentowaniem, to co jest podstawą oprocentowania? Czy zastosowane procentowanie pozwala nam odpowiedzieć na postawione pytanie?

6. PODSUMOWANIE

Cały proces badawczy w przypadku badania ilościowego składa się z dwóch etapów. Pierwszy etap polega na przygotowaniu badania, drugi na opracowaniu danych. Na etapie projektowania badania należy zebrać informacje na temat tego, co właściwie ma być przedmiotem badania i jakie są główne pytania, na które badanie ma udzielić odpowiedzi. Pytania badawcze zebrane są w dokumencie nazywanym briefem i stanowią punkt wyjścia do konstrukcji kwestionariusza. Opracowanie pytań

[5] Wiąże się to, między innymi, z niewielkimi próbami stosowanymi zazwyczaj w badaniach ewaluacyjnych. Opinia, jakoby statystycy potrafili liczyć, ale dopiero od tysiąca wzwyż jest oczywiście nieprawdziwa, lecz tkwi w niej ziarno prawdy – niektóre statystyki wymagają prób wielokrotnie wyższych od tych stosowanych w badaniach ewaluacyjnych.

badawczych nie jest jednoznaczne z opracowaniem pytań kwestionariuszowych. W rzeczywistości określenia, którymi posługuje się zamawiający w briefie, bardzo rzadko nadają się do zamieszczenia w kwestionariuszu. Wśród socjologów krążyła swego czasu anegdota, że zamawiający w jednym z badań zamierzał zadać pytanie „Które z wymiarów Ja odnajduje Pani w tym opakowaniu?". Na szczęście badaczowi udało się przekuć to dość hermetyczne sformułowanie na pytanie przyjazne dla odbiorcy.

W przypadku konstruowania kwestionariusza należy zwrócić szczególną uwagę na jego logikę i zastanowić się, czy układ pytań jest wewnętrznie spójny. Warto jest przyjrzeć się też skalom i filtrom, po to, by uniknąć przykrych niespodzianek na etapie analizy. Należy jeszcze raz rozważyć zasadność pytań otwartych, stanowiących zawsze najsłabsze ogniwo późniejszych analiz (odpowiedzi na pytania otwarte są niekwantyfikowalne i bezużyteczne z punktu widzenia statystyki). Należy również zastanowić się nad tym, czy pytania zawarte w kwestionariuszu umożliwią nam późniejszą odpowiedź na pytania zawarte w briefie.

Od rodzaju pytań zawartych w briefie zależy nie tylko ostateczny kształt kwestionariusza, ale również decyzja o sposobie przeprowadzenia badania i doborze próby. Warto pamiętać, że choć sposoby przeprowadzenia ankiet są do siebie podobne, to jednak do niektórych rodzajów badań lepiej jest wykorzystać telefon, do innych zdecydowanie korzystniej będzie wykorzystać przewagę, jaką daje fizyczna obecność ankietera[6]. Od badacza ostatecznie zależy, którą z dostępnych metod przeprowadzenia wywiadu wybierze.

Osobną kwestią jest konstrukcja próby, od której również w ogromnym stopniu zależy jakość otrzymanych ostatecznie wyników. Ważny jest zarówno zakładany błąd oszacowania, którym obarczona jest każda próba, jak i rodzaj zastosowanej próby (losowa, celowa). Konstrukcja próby, określająca osoby, które zamierzamy przebadać, jest również dokumentem umożliwiającym następnie konstrukcję wag analitycznych, a co za tym idzie „wyprostowanie" zrealizowanej próby do założeń próby zakładanej.

Na konstrukcji próby i kwestionariusza kończy się przygotowawczy etap badania. Drugim etapem jest etap analityczny, czyli przygotowanie raportów i prezentacji z przeprowadzonego badania. Raport powinien zawierać odpowiedzi na pytania badawcze zawarte w briefie. W trakcie pisania raportu należy zwrócić uwagę na kilka rzeczy. Po pierwsze,

---

[6] Dla przykładu rozważmy dwie sytuacje: kiedy mamy do przeprowadzenia 7000 wywiadów z respondentami rozrzuconymi po całej Polsce, raczej skorzystamy z metody CATI, dzięki czemu unikniemy angażowania ogromnej liczby ankieterów, którzy będą musieli dotrzeć do każdego respondenta. Z kolei badania przeprowadzane np. na osobach piastujących kierownicze stanowiska zdecydowanie lepiej wypadają w metodach face-to-face (CAPI, PAPI), gdyż osoby te bywają niedostępne, gdy próbujemy się do nich dodzwonić.

każdy wykres czy tabela zawarte w raporcie powinny, prócz informacji o rozkładzie odpowiedzi, zawierać pełną treść pytania wraz z numeracją z kwestionariusza. Po drugie, tabele i wykresy powinny zawierać informację o podstawie oprocentowania (jeśli zastosowano w nich procentowanie). Brak tych elementów jest najczęstszym zaniedbaniem piszących raporty i konstruujących prezentacje, co skutkuje tym, że nie wiadomo, jakie pytanie zadano respondentowi ani ile osób odpowiedziało na pytanie. Skutkiem tego nieraz napotkać można „szokujące" wyniki badań, z których wynika, że 98% badanych zna i używa markę X, których „szokującość" znikłaby, gdybyśmy wiedzieli, że pytanie zadano respondentom, którzy używają akurat tej marki.

Należy być bardzo ostrożnym z ekstrapolowaniem wyników badań na całą populację, gdyż istnieje wiele przeszkód uniemożliwiających bądź ograniczających taką ekstrapolację. Jak pisałem w podrozdziale poświęconym próbom, tylko niektóre z nich tworzą warunki sprzyjające przenoszeniu wyników na całą populację; nie bez znaczenia jest również poziom błędu estymacji, o czym również była mowa w tym rozdziale.

Udane przeprowadzenie badania ilościowego to proces twórczy. Od jakości próby i kwestionariusza zależą późniejsze wyniki przedstawione w raporcie; sam proces pisania raportu jest czynnością wymagającą uwagi i pomysłowości. Proces odpowiadania na pytania marketingowe czy społeczne ma swoje poważne ograniczenia, z których dobrze jest zdawać sobie sprawę i wystrzegać się stwierdzeń nieuprawnionych bądź niekompletnych.

BIBLIOGRAFIA

Babbie E. (2003), *Badania społeczne w praktyce*, Warszawa, PWN.

Barnett V. (1982), *Elementy teorii próby*, Warszawa, PWE.

Fowler F.J. (1995), *Improving Survey Questions. Design and Evaluation*, London, Sage.

Frankfort-Nachmias Ch., Nachmias D. (2001), *Metody badawcze w naukach społecznych*, Poznań, Zysk i S-ka.

Górniak J, Wachnicki J. (2000), *Pierwsze kroki w analizie danych*, Kraków, SPSS Polska.

Karpiński J. (2006), *Wprowadzenie do metodologii nauk społecznych*, Warszawa, Wydawnictwo WSPiZ.

Lutyński J. (1994), *Metody badań społecznych*, Łódź, ŁTN.

Mayntz R., Holm K., Hübner P. (1985), *Wprowadzenie do metod socjologii empirycznej*, Warszawa, PWN.

Noelle E. (1976), *Reprezentatywne badania ankietowe. Wprowadzenie do metodologii. Wybór tekstów*, Warszawa, OBOP.

Nowak S. (red.) (1965), *Metody badań socjologicznych. Wybór tekstów*, Warszawa, PWN.

Sawiński Z. (red.) (1992), *Podręcznik socjologicznych badań ankietowych*, Warszawa, IFIS PAN.

Sztabiński P.B. (1999), Wywiad telefoniczny ze wspomaganiem komputerowym: nowa technika czy wywiad kwestionariuszowy przez telefon?, *Przegląd Socjologiczny* **48/1**.

Sztabiński P.B. (2001), Wywiad osobisty ze wspomaganiem komputerowym (CAPI): czy rzeczywiście idealna technika?. *ASK. Społeczeństwo. Badania. Metody* **10/2001**, Warszawa, IFiS PAN.

Sztabiński P.B. (2002), Punkt „środkowy" w skalach bipolarnych: zamieszczać czy nie zamieszczać?, *ASK. Społeczeństwo. Badania. Metody* **11/2002**, Warszawa, IFiS PAN.

Sztabiński P.B., Sztabiński F., Sawiński Z. (red.) (2004), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN.

KOMENDY SPSS

Tabela częstości (frekwencje)

Składnia w SPSS (zmienna x1):

```
FREQUENCIES
VARIABLES=x1
/ORDER= ANALYSIS .
```

lub

```
fre x1.
```

Tabela krzyżowa (bez procentowania, same liczebności)

Składnia w SPSS (zmienne: x1, x2):

```
CROSSTABS
/TABLES= x1 BY x2
/FORMAT= AVALUE TABLES
/CELLS= COUNT
/COUNT ROUND CELL .
```

lub

```
cross x1 by x2.
```

Tabela krzyżowa (procentowanie w wierszach)

Składnia w SPSS (zmienne: x1, x2):

```
CROSSTABS
/TABLES= x1 BY x2
/FORMAT= AVALUE TABLES
/CELLS= ROW
/COUNT ROUND CELL .
```

lub

```
cross x1 by x2
/cell row.
```

Tabela krzyżowa (procentowanie w kolumnach)

Składnia w SPSS (zmienne: x1, x2):

```
CROSSTABS
/TABLES= x1 BY x2
/FORMAT= AVALUE TABLES
/CELLS= COLUMNS
/COUNT ROUND CELL .
```

lub

```
cross x1 by x2
/cell col.
```

SŁOWNICZEK

ankieta – w niniejszym tekscie rozumiana jako ankieta samowypełnialna; w odróżnieniu od kwestionariusza nie zakłada udziału ankietera

badanie *ad hoc* – badanie jednorazowe

badanie panelowe – rodzaj badania trackingowego; przeprowadzane za każdym razem z tymi samymi respondentami, na niezmienionym kwestionariuszu; służy wychwyceniu zmian

badanie trackingowe – badanie powtarzalne

baza danych – zbiór zakodowanych informacji zebranych w trakcie badania

booster (nadreprezentacja) – element próby, dobierany celowo ze względu na niską szansę wylosowania odpowiedniej ilości respondentów z danej grupy.

brief – dokument zawierający pytania badawcze

CAPI – Computer Assisted Personel Interview (badanie *face-to-face*, za pomocą skryptu)

case – przypadek; w statystycznym żargonie oznacza wiersz w bazie danych odpowiadający respondentowi

CATI – Computer Assisted Telephone Interview (badanie telefoniczne)

CAWI – Computer Assisted Web Interwiew (badanie internetowe)

filtr – ma dwa znaczenia: w kwestionariuszu określa warunki zadania danego pytania; filtr w bazie danych wyklucza określone przypadki z badania

frekwencja – częstość; komenda statystyczna pozwalająca na zobaczenie rozkładu zmiennej

kafeteria – część pytania w kwestionariuszu, wyszczególniająca możliwe odpowiedzi na pytanie

koder – w przypadku ankiet PAPI; osoba wprowadzająca dane do bazy danych

kwestionariusz – narzędzie badawcze, zawierające pytania, w odróżnieniu od ankiety zakłada współdziałanie respondenta i ankietera

kwestionariusz rekrutacyjny – część kwestionariusza, odpowiedzialna za właściwy dobór respondenta do dalszej części badania

metryczka – część kwestionariusza, zawierająca podstawowe informacje o badanym, niezwiązane bezpośrednio z celem badania (takie jak płeć, wiek, wykształcenie, czy liczba pracowników, obrót)

PAPI – badanie przeprowadzane za pomocą kwestionariusza drukowanego

prekodowanie – odpowiedzi prekodowane stosowane są jako uzupełnienie pytań otwartych; stosowane, gdy badacz może przewidzieć przynajmniej część odpowiedzi na dane pytanie

próba – dobrany zestaw respondentów, wybrana część badanej populacji

próba celowa – próba, którą dobiera się ze względu na cel badania

próba losowa – próba, którą dobiera się przez losowanie

próba losowo-warstwowa – próba, w której dokonuje się losowania respondentów kilkakrotnie w poszczególnych warstwach (np. dochodowych, czy ze względu na wielkość przedsiębiorstwa)

próba reprezentatywna – próba, do której każdy element ma szansę wejść z takim samym prawdopodobieństwem

respondent – osoba badana

skala – rodzaj kafeterii odpowiedzi w pytaniu kwestionariuszowym.

skala niewerbalna – skala, w której kolejne punkty opisane są wyłącznie liczbowo

skala quasi-niewerbalna – skala, w której słownie opisano jedynie krańce, wszystkie pozostałe wartości są opisane wyłącznie liczbowo

skala symetryczna – skala, która posiada tyle samo odpowiedzi pozytywnych, co negatywnych

skala werbalna – skala, której kolejne punkty opisane są słownie

skrypt – elektroniczna wersja kwestionariusza

tabela częstości (frekwencje) – najczęściej tabela zawierająca informacje o jednej zmiennej

tabela krzyżowa (*crosstab*) – najczęściej tabela zawierająca informacje o dwóch zmiennych

top box – najwyższa (pozytywna) wartość skali

top2box – zsumowane dwie najwyższe (pozytywne) wartości skali

low box – najniższa (negatywna) wartość skali

low2box – zsumowane dwie najniższe (negatywne) wartości skali

waga analityczna – narzędzie statystyczne używane w przypadku skrzywienia próby, bądź dopasowania próby do ekstrapolacji na populację

wiązka próby – określa minimalną liczbę wywiadów do zrealizowania w danym miejscu (np. miejscowości czy dzielnicy); minimalizuje koszty dojazdu do respondentów

ROZDZIAŁ 2.5.

Krzysztof Piech

EWALUACJA ODDZIAŁYWANIA FUNDUSZY UNIJNYCH – TEORIA I PRAKTYKA MODELOWANIA MAKROEKONOMICZNEGO

Celem niniejszego rozdziału[1] będzie zaprezentowanie jednego z rodzajów metod stosowanych w ewaluacji – modeli makroekonomicznych. Są one jedną z najtrudniejszych metod ewaluacyjnych, a jednocześnie jedną z najbardziej złożonych w ramach całych nauk ekonomicznych – od dziesiątek lat budowanie modeli pozostaje wyzwaniem, z którym zmaga się elita ekonomistów świata. Niemniej jednak autor starał się zaprezentować tę problematykę w sposób możliwie zrozumiały i przystępny nie tylko dla zaawansowanych ekonomistów. Z pewnością jednak niektóre fragmenty mogą sprawiać trudność w zrozumieniu dla czytelnika bez wykształcenia ekonomicznego.

Modele pomagają w projektowaniu i realizacji różnych polityk, szczególnie polityki spójności. Ewaluacja powinna być przeprowadzana w stosunku do każdych większych wydatkowanych pieniędzy publicznych. Jednak do tej pory w naszym kraju koncentrowano się prawie wyłącznie na funduszach unijnych. Stąd również niniejszy rozdział będzie głównie

[1] Dziękuję Janowi in't Veld z Komisji Europejskiej za uwagi do fragmentów wcześniejszej wersji niniejszego rozdziału, a także Dorocie Pelle z Instytutu Badań Strukturalnych, Michałowi Greszcie z Narodowego Banku Polskiego, Tomaszowi Kaczorowi z Instytutu Badań nad Gospodarką Rynkową, oraz innym osobom, których życzliwe uwagi zostały wykorzystane.

nim poświęcony, tj. ocenie (głównie *ex ante*) oddziaływania funduszy unijnych[2] na polską gospodarkę.

Rozdział stara się łączyć kwestie teoretyczne z prezentacją przykładów różnych wyników badań, by całość opracowania była możliwie interesująca. Pierwsza część rozdziału (podrozdziały 1–4) poświęcona jest przeglądowi rodzajów modeli, zarówno ich typów, jak i konkretnych przykładów, w tym również polskich. Następnie zaprezentowane i omówione zostały wybrane wyniki, uzyskane za pomocą modeli: zarówno europejskich (podrozdział 5), jak i polskich (podrozdział 6), również w ujęciu regionalnym (podrozdział 7). W zakończeniu (podrozdział 8) zawarte są treści, niepowtarzające zawartości opracowania, a osadzające je w szerszym kontekście.

1. RODZAJE MODELI MAKROEKONOMICZNYCH

„Interwencja wspólnoty może mieć istotne oddziaływanie na gospodarkę, jako całość, szczególnie w krajach i regionach, gdzie skala transferów stanowi istotną część ich PKB czy inwestycji krajowych. W takich przypadkach, tylko modelowanie makroekonomiczne może symulować złożone współzależności pomiędzy zmiennymi ekonomicznymi na poziomie makroekonomicznym dla ewaluacji *ex ante* wpływu makroekonomicznego głównych Planów Celów Konwergencji czy Ram" (EC 2006: 21). Modele makroekonomiczne są „praktycznie jedynym narzędziem, które może być użyte do formalnego potwierdzenia, czy polityka europejska osiągnęła założone cele" (*Evaluating…* 2003).

Mimo wielu zalet, nie są one pozbawione wad. Dla uzyskania właściwych rezultatów modelowania potrzebne są oczywiście wysokiej jakości dane. Ale nie tylko to. Dużą rolę odgrywa też sam model. W przeszłości większość modeli makroekonomicznych była pozbawiona podstaw mikroekonomicznych, takich jak maksymalizacja zysków i minimalizacja kosztów producentów (przełom w tym zakresie wniosły modele typu CGE).

Modele makroekonometryczne mogą wykorzystywać modele ekonomiczne oraz oczywiście – narzędzia ekonometryczne. Wśród podstawowych modeli makroekonomicznych wyróżnić można np. model IS-LM, model IS-LM-BP, modele krótkookresowej równowagi makroekonomicznej, modele cyklu koniunkturalnego, modele wzrostu gospodarczego i in. (Garbicz i Golachowski 2005).

Modele ekonomiczne stosowane w ewaluacji oddziaływania polityki strukturalnej można podzielić na dwie, główne grupy (w ramach drugiej z nich pewne klasy modeli) (BICEPS 2008: 7–10):

---

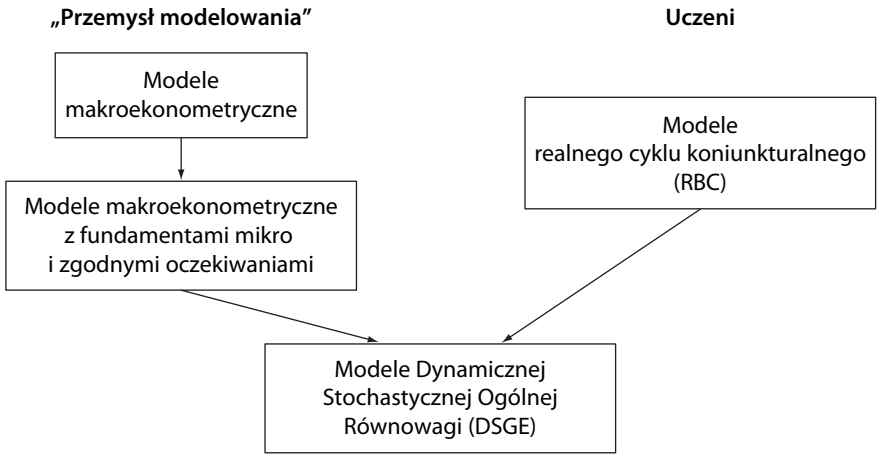
[2] Tam, gdzie jest mowa w niniejszym opracowaniu o oddziaływaniu środków unijnych, uwzględnia się współfinansowanie ze środków krajowych.

- ekonometryczne modele regresji, oparte na pojedynczych równaniach, bazujące na idei konwergencji,
- modele symulacyjne, przybliżające zachowania gospodarki:
  - modele estymowane/kalibrowane (np. HERMIN, QUEST II, ECOMOD),
  - modele CGE (stosowane w wielu krajach, np. w Grecji, Portugalii, w Polsce),
  - modele regionalne,
  - modele częściowej równowagi.

Ze względu na techniki stosowane w modelach, można je podzielić na następujące klasy:

- modele realnego cyklu koniunkturalnego,
- modele typu CGE,
- modele typu DCGE.

Próbę zaprezentowania przemian w modelowaniu makroekonomicznym, zachodzących w ostatnich latach, przedstawia rysunek 1.



Rysunek 1. Dynamika modeli makroekonomicznych

Źródło: Varga i Veld 2008.

Modele realnego cyklu koniunkturalnego (*real business cycle* – RBC) wykorzystują nie tylko estymację parametrów modelu za pomocą technik ekonometrycznych, lecz również kalibrację. Polega to na tym, że w przypadku, gdy model daje wyniki inne niż wynikałoby to z danych rzeczywistych, zmienia się parametry tak, by wyniki modelu i rzeczywistości się pokrywały (bez ingerowania w strukturę modelu, tj. poprawiania go). W związku z tym przydatność tych modeli do celów prognostycznych jest ograniczona. Mimo że te modele próbują łączyć krótkoterminowe wahania produkcji z mechanizmami długookresowego wzrostu, to jest on wyznaczany przez neoklasyczny model wzrostu Solowa, a zatem nie są wskazane źródła podstępu technologicznego (Tokarski 2005: 31).

Obok modeli RBC, najbardziej popularnymi modelami makroekonometrycznymi są **modele CGE** (*computable general equilibrium* – obliczeniowy model równowagi ogólnej). Służą one do symulacji zmian polityki, technologii i in. Wychodzą od modelu *input-output* Leontiewa oraz keynesowskich równań równowagi. Były one szczególnie popularne na świecie w latach 80. Modele te są szczególnie użyteczne w otrzymywaniu długoterminowych prognoz. Nie za dobrze radzą sobie radzą jednak ze sferą pieniężną (finansową).

Częściowo w odpowiedzi na krytykę Lucasa[3] powstały **modele DSGE** (*dynamic stochastic general equilibrium* – dynamiczny, stochastyczny model równowagi ogólnej). Starają się one bazować na fundamentach mikroekonomicznych. W przeciwieństwie do wielu modeli, są one odporne na krytykę Lucasa. Modele te z reguły mają mniej zmiennych i są trudniejsze do rozwiązania (też przy użyciu komputerów). Dość dobrze radzą sobie z krótkim okresem. Często oparte są na danych kwartalnych, a nie jedynie na rocznych. Stąd lepsze są dla krótkookresowego prognozowania, w tym dla badania cykli koniunkturalnych oraz efektów polityki gospodarczej.

2. NAJWAŻNIEJSZE MODELE MAKROEKONOMICZNE W EUROPIE

Różne z wymienionych rodzajów modeli są stosowane w praktyce przez różne instytucje: banki centralne, ministerstwa, uczelnie czy instytuty badawcze. W Europie rozwiniętych zostało kilka głównych modeli makroekonomicznych. Wymienia się najczęściej: QUEST, E3ME i HERMIN. Są to przykłady modeli ekonometrycznych, szeroko używanych w symulowaniu wpływu polityki pieniężnej, konwergencji i funduszy strukturalnych w UE[4].

Poniżej dokonano krótkiej charakterystyki głównych modeli makroekonomicznych stosowanych w Europie (por. też *Evaluating*… 2003):

- **QUEST** to grupa modeli (oznaczonych numerami I, II, III) rozwijanych przez Komisję Europejską. Modele te mają charakter globalny, tj. nie ograniczają się jedynie do Unii Europejskiej, ale modelując gospodarkę światową, badają relacje UE z resztą świata (np. z USA).

[3] Robert Lucas, laureat Nagrody Nobla, stwierdził po raz pierwszy w swoim artykule pt. *Econometric Policy Evaluation: A Critique* (1976), że niewłaściwe jest stosowanie wielkich modeli makroekonomicznych do oceny konsekwencji alternatywnych scenariuszy polityki, gdyż symulacje takie odbywają się na podstawie niezmienionych parametrów modelu, podczas gdy wprowadzana polityka je zmieni. A zmiana ta następuje wskutek dostosowywania się jednostek do oczekiwanych efektów polityki. „W rezultacie nie jest właściwe posługiwanie się modelami ekonometrycznymi dla celów predykcji przyszłej polityki" (Landreth i Colander 1998: 737).

[4] Model REMI Policy Insight jest stosowany przede wszystkim w Stanach Zjednoczonych, przy czym po jego modyfikacjach stosowano go też w Europie (południowe Włochy, Andaluzja, Toskania) (*Evaluating*… 2003).

Ich celem jest analiza cykli koniunkturalnych oraz procesów długookresowego wzrostu. Traktują gospodarkę tak, jakby był w niej jedynie jeden sektor (model jednosektorowy).

Druga wersja modelu (**QUEST II**) to model oparty na funkcji Cobba-Douglassa, co – zdaniem Bradley'a i Untieda (2007: 20) – może powodować wyższe (niż w modelu HERMIN) efekty wypierania. Zawiera zmienne zasobów i przepływów, w tym kapitał fizyczny, aktywa zagraniczne netto, pieniądz, dług rządowy, które są endogeniczne. Model umożliwia analizę przepływu oszczędności, decyzji produkcyjnych i inwestycyjnych gospodarstw domowych, firm i rządu. Strona podażowa modelu determinowana jest przez neoklasyczną, zagregowaną funkcję produkcji; długookresowe stopy wzrostu determinowane są przez egzogeniczny postęp technologiczny i stopę wzrostu liczby ludności. Wyniki symulacji mogą być prezentowane jako odchylenia od podstawowego scenariusza. Realne stopy procentowe i kurs walutowy są determinowane endogenicznie, co umożliwia badanie „efektu wypierania" inwestycji sektora prywatnego przez fundusze strukturalne. Model QUEST II był stosowany m.in. do szacowania wpływu kryteriów z Maastricht na wzrost i zatrudnienie, wpływu reform strukturalnych, polityki pieniężnej i fiskalnej itd. Ostatnio był użyty również do szacowania wpływu polityki spójności w nowych krajach członkowskich UE.

Od niedawna rozwijana jest trzecia wersja modelu – **QUEST III**. Należy on do klasy modeli DSGE. Zawiera m.in. sektor B+R (co odpowiada na wyzwania stawiane w Strategii Lizbońskiej), a także, co ciekawe, trzy rodzaje zasobów siły roboczej: w podziale ze względu na umiejętności niskie, średnie i wysokie. Docelowo zawierał on będzie submodele dla każdego kraju członkowskiego UE (w tym Polski): najpierw parametry będą kalibrowane – do końca lata 2008 roku, a następnie estymowane – do 2010 roku (Veld 2008).

- **ECOMOD** to model rozwijany przez EcoMod i Wolny Uniwersytet w Brukseli. Jest on dość zdezagregowany (zawiera sześć sektorów oraz 60 rodzajów działalności). Był używany do ewaluacji wpływu funduszy w okresie programowania 2007–2013 dla wszystkich krajów członkowskich.
- Model **HERMIN** ma trzy główne, zintegrowane komponenty (strona podażowa, absorpcji i dystrybucji dochodu). Mechanizm zagregowanego popytu modelowany jest w sposób keynesowski, natomiast podaż kształtowana jest przez ceny i konkurencyjność kosztową (a zatem w sposób neoklasyczny). Brakuje w nim sektora nauki, a także sektora pieniężnego gospodarki. Stąd postęp techniczny, stopy procentowe i kurs walutowy są egzogeniczne (nie są wyjaśniane przez model). Model ten zastosowany był np. w Grecji, Hiszpanii, Irlandii, Portugalii oraz w krajach transformacji (np. Węgry, Polska, Słowacja, Łotwa, Estonia).

Inne, ważniejsze modele makroekonomiczne stosowane w Europie to:

- **AWM** (*area-wide model*) Europejskiego Banku Centralnego, odnoszący się do całej strefy euro. Zawiera 89 równań (w tym 15 behawioralnych – estymowanych). Stosuje funkcję Cobba-Douglasa o stałych korzyściach skali. Czasami przeprowadza się przy jego użyciu symulacje polityki gospodarczej (Fagan 2001).
- Modele **NIGEM** (National Institute's Global Econometric Model)[5] i **NIDEM** (National Institute's Domestic Econometric Model) Narodowego Instytutu Badań Gospodarczych i Społecznych (NIESR) w Wielkiej Brytanii.
- **E3ME** (Energy-Environment-Economy Model of Europe) to model wielosektorowy (zawiera aż 42 sektory gospodarki oraz 41 kategorii wydatków konsumentów), dynamiczny, ekonometryczny, oparty na danych rocznych. Zawiera rozbudowane komponenty: energetyczny i środowiskowy. Uwzględnia działalność B+R i innowacje jako źródło wzrostu. Model pokrywa 29 krajów europejskich. Korzysta z doświadczeń modeli CGE w zakresie kalibracji ostatnich danych. Rozwijany był przez Komisję Europejską z zespołem instytucji partnerskich prowadzonych przez Cambridge Econometrics.
- Model **Interlink** OECD, jak nazwa wskazuje, obejmuje wiele krajów OECD (i nie tylko) wraz z połączeniami między nimi. Wersja modelu z końca lat 80. zawierała od 200 do 250 równań dla większych krajów oraz 130–150 równań dla mniejszych. Przeciętnie dany kraj OECD opisywały 44 równania „łącznikowe" (Richardson 1988).
- Model **GEM-E3** autorstwa Centrum Studiów Gospodarczych Katolickiego Uniwersytetu Leuven (i partnerów) to model typu CGE obejmujący gospodarkę światową, w tym europejską. Stosowany był głównie w latach 2000–2004, głównie w zakresie polityki energetycznej i środowiska naturalnego.

Można też wspomnieć o innych, ważnych modelach, spoza Unii Europejskiej. Najważniejsze z nich to modele GEM Międzynarodowego Funduszu Walutowego, które są jednymi z najważniejszych na świecie. Oba pochodzą z najnowszej klasy modeli DSGE. Model **GEM** (Global Economy Model) jest jednym z pierwszych tego typu (DSGE) modeli na świecie. Bazuje na paradygmacie makroekonomii nowej, otwartej gospodarki (New Open Economy Macroeconomics). Ze względu na oparcie się na danych kwartalnych, był skonstruowany dla badania krótkookresowej dynamiki. Jednakże utrudniało to badanie wielu zagadnień polityki

---

[5] Model NIGEM stosowany jest do kwartalnych prognoz gospodarki brytyjskiej, uwzględnia wpływ gospodarki światowej na nią. Stosowany był do ewaluacji np. wpływu zmian oszczędności gospodarstw domowych, zmian cen domów czy migracji na brytyjską gospodarkę. Stosowany jest też przez MFW, EBC oraz różne banki centralne; był zakupiony również (wraz z komponentem polskim) przez NBP i Ministerstwo Finansów (Zaleski et al. 2004: 9).

podatkowej. Stąd powstał drugi model – **GFM** (Global Fiscal Model), który oparty jest na danych rocznych oraz skoncentrowany na średnio- i długookresowych zjawiskach, takich jak efekty polityki fiskalnej (Botman et al. 2007).

3. MODELE W POLSCE I OŚRODKI MODELOWANIA

Komisja Europejska zobowiązała każdy kraj członkowski do przeprowadzania ewaluacji *ex ante* swoich programów operacyjnych (i nie tylko nich). Jednym z narzędzi ewaluacji są modele makroekonomiczne. Zgodnie z wytycznymi Komisji (EC 2006: 22), modele makroekonomiczne stosowane w ewaluacji powinny:

- zawierać stronę popytową i podażową gospodarki, przy czym ostatnia powinna na różne sposoby odzwierciedlać takie obszary interwencji jak infrastruktura, kapitał ludzki, badania i rozwój[6] oraz produktywne inwestycje, które mogą istotnie wzmacniać efekty zewnętrzne,
- posiadać następujące, główne zmienne wyjaśniane: PKB, inwestycje, zatrudnienie, płace, ceny, deficyt budżetowy, import i eksport,
- umożliwiać symulowanie przynajmniej następujących scenariuszy: same fundusze narodowe, fundusze UE oraz oba źródła razem,
- rozważać koszty alternatywne wydatków publicznych, tj. efekty innych alternatywnych sposobów wydatkowania funduszy krajowych i unijnych,
- pokazywać wrażliwość rezultatów na zmiany wydatków funduszy strukturalnych i spójności oraz na zmiany polityki gospodarczej.

W Polsce jest kilkanaście ośrodków prognostycznych oraz kilkadziesiąt osób profesjonalnie zajmujących się prognozowaniem; wśród nich m.in.:

- instytucje rządowe (ministerstwa),
- uczelnie wyższe,
- niezależne instytuty badawcze,
- instytucje finansowe, w tym banki (a w nich biura głównych ekonomistów oraz departamenty analityczne i dealingowe),
- publicyści ekonomiczni.

Większość z nich nie ma własnych modeli makroekonomicznych dla celów prognostycznych. Do najbardziej znanych ośrodków dysponujących nimi można zaliczyć:

- Narodowy Bank Polski, który ma kilka modeli ekonometrycznych. Podstawowym celem modeli ECMOD, MSMI i NSA jest przeprowadzenie

[6] Modelowanie sektora nauki (badawczo-rozwojowego) jest dość nową i przez to trudną kwestią do przeprowadzenia – na tyle, że w Polsce zawiera go prawdopodobnie zaledwie jeden model (model łódzki prof. W. Welfe).

projekcji makroekonomicznych (prognoz), a także – w dalszej kolejności – symulacje polityki pieniężnej. Wśród modeli NBP wymienić można:

- model NECMOD, który od czerwca 2008 roku jest głównym modelem prognostycznym NBP. Wyniki uzyskiwane za jego pomocą publikowane są w *Raportach o inflacji*. Model pełniej odzwierciedla ewolucję gospodarki w ostatnich latach, niż model ECMOD, a także przewidywany jej rozwój w latach kolejnych. Model zachowuje hybrydowy charakter i długookresową stabilność, ma rozbudowaną stronę podażową gospodarki (np. produkt potencjalny, stopa bezrobocia równowagi, kurs równowagi), modeluje oczekiwania, ma też „potencjalne możliwości symulacyjne, pozwalające uzyskać „odpowiedzi" na wiele aktualnych pytań (efekt zmian fiskalnych, **funduszy strukturalnych**, emigracji; Budnik et al. 2008). Zastąpił on model ECMOD, który obecnie nie jest już używany.
- model ECMOD – model hybrydowy, obejmujący elementy neoklasyczne (w długim okresie) i keynesowskie (w krótkim okresie); model ma charakter kwartalny, ma ponad 100 równań (w tym 15 behawioralnych); stosowana jest funkcja produkcji Cobba-Douglasa z kapitałem i pracą jako czynnikami produkcji; niektóre z parametrów są kalibrowane (Fic et al. 2005);
- model MSMI (mały strukturalny model inflacji) – mały model (20 równań), zbudowany w latach 1999–2000 do analizy procesów inflacyjnych, a także do analizy scenariuszy np. wejścia Polski do UE (Kłos 2000);
- model NSA (nowy schemat analityczny), stosowany w latach 2000–2003 co miesiąc, a następnie co kwartał dla przygotowywania projekcji inflacji dla Rady Polityki Pieniężnej (Kłos et al. 2004).

- Ministerstwo Finansów[7],
- Ministerstwo Gospodarki, dysponujące dwoma modelami:
  - typu CGE, służący do symulacji polityki strukturalnej (oraz makroekonomicznej); obejmuje 39 sektorów produkcyjnych; wykorzystywane są cztery rodzaje podstawowych czynników produkcji (kapitał oraz siła robocza o trzech poziomach umiejętności: wysokim, średnim i niskim), gospodarstwa domowe podzielone są na bogate i biedne, a także według jednego z pięciu typów zatrudnienia; model zawiera też sektor bankowy; funkcja produkcji oparta jest na równaniu CES;
  - typu DSGE, stosowany do symulowania polityki makroekonomicznej; wzrost gospodarczy jest w tym modelu wynikiem działalności badawczo-rozwojowej, a także akumulacji kapitału ludzkiego oraz infrastruktury.

[7] Autor nie posiada informacji na temat modeli stosowanych przez MF.

- Instytut Badań nad Gospodarką Rynkową (model MaMoR2 – typu CGE, por. dalej).
- Niezależny Ośrodek Badań Ekonomicznych, kierowany przez prof. W. Orłowskiego (model typu CGE, endogeniczny, użyty w 2002 roku do symulacji korzyści wejścia Polski do Unii Europejskiej; por. Orłowski 2000).
- Instytut Badań Rynku, Konsumpcji i Koniunktur (dawny Instytut Koniunktur i Cen Handlu Zagranicznego) (por. dalej).
- Uniwersytet Łódzki – model W8D-2002 – model makroekonometryczny, wykorzystujący osiągnięcia endogenicznej teorii wzrostu, w tym uwzględniający innowacje, prace badawczo-rozwojowe i kapitał ludzki; model zawiera ponad 200 równań (z których prawie 40 poświęconych jest postępowi technicznemu), ponad 300 zmiennych (w tym ponad 200 endogenicznych); za pomocą tego modelu prowadzono symulacje kilku scenariuszy rozwoju polskiej gospodarki do 2025 roku (np. Welfe 2005).
- Wrocławska Agencja Rozwoju Regionalnego (wykorzystująca irlandzki model HERMIN – por. dalej).
- Instytut Badań Strukturalnych, dysponujący trzema modelami typu DSGE. Wszystkie z nich są modelami najmłodszymi spośród wymienionych. Tej klasy modeli jest zaledwie kilka w Europie (stosuje je m.in. bank centralny Włoch, Komisja Europejska – QUEST III)
  - Model EU-ImpactMod, który został zastosowany do oszacowania funduszy unijnych na polską gospodarkę, jest najmłodszym modelem makroekonomicznym polskiej gospodarki – powstał w 2007 roku.
  - Model OLG-DSGE skonstruowany został dla analizy podaży pracy w Polsce (i oddziaływania szoków makroekonomicznych, w tym zmiany technologicznej na nią), stąd zawiera moduł demograficzny, bazujący na klasycznym schemacie nakładających się pokoleń (OLG). Uwzględnia on też reprezentację systemu emerytalnego w Polsce (Bukowski i Zawistowski 2008b).
  - „Model DSGE dużej skali" skonstruowany został dla symulowania wprowadzenia euro w Polsce. Zamiast funkcji CES stosuje funkcję o stałej, relatywnej stopie substytucji, dzięki czemu autorzy kalibrują zmienne przy pomocy rachunków sektorowych, statystyk rynku pracy i rynku pieniężnego (Bukowski 2008).

Oczywiście każdy z tych modeli dostarcza innych wyników. W przypadku prognozowania (w tym ewaluacji *ex ante*) trudno jest z góry określić, którym wynikom należy bardziej ufać, a którym mniej. Niektórzy wręcz twierdzą, że „nie ma «naukowego» sposobu wyboru «najlepszego modelu»" (BICEPS 2008: 13). Jednak ekonomiści, od zarania ekonomii, jako nauki, porównują modele i wśród naukowców zajmujących się tą problematyką istnieje pewnego rodzaju konsensus co do tego, który

model jest właściwszy, lepszy do zastosowania. Wymieniłbym tu następujące kryteria oceny modeli:

- Aktualność i akceptowalność przez środowisko ekonomistów teorii stojącej za modelem. Obecnie każdy zgadza się, że np. modele neoklasyczne są lepsze od opartych na teorii klasycznej; ponadto uwzględnienie teorii wzrostu endogenicznego również uznawane jest za poprawiające ocenę modelu. W tym zakresie uznaje się także, iż im model makroekonomiczny ma solidniejsze fundamenty mikroekonomiczne, tym jest lepszy. Do oceny stopnia zdolności teoretycznej i empirycznej modeli służy klasyfikacja Pagana (2003), zgodnie z którą wyróżnia się modele czysto teoretyczne, statystyczne i hybrydowe (będące ich połączeniem).
- Klasa modelu – obecnie za najlepsze modele uznaje się te typu DSGE (choć jeszcze nie jest pewne, czy modele te na stałe przyjmą się w praktyce modelowania).
- Liczba zmiennych, w tym zmiennych endogenicznych, np. odnoszących się do działalności badawczo-rozwojowej. Uznaje się, że im więcej ich jest, tym model jest bardziej zaawansowany.
- Sposób wyznaczania parametrów modelu: estymacja ekonometryczna – od strony naukowej (choć niekoniecznie praktycznej) – uznawana jest za właściwszy sposób, niż kalibracja.
- Trafność prognoz modelu, stopień zbliżenia uzyskiwanych z modelu prognoz do rzeczywistości: nawet jeśli model jest dobry z punktu widzenia powyższych kryteriów, a dawałby nierzeczywiste, niemożliwe do osiągnięcia wyniki (lub gdy *ex post* okazałoby się, że różnice między danymi teoretycznymi a rzeczywistymi są bardzo duże), trudno byłoby uznać taki model za dobry, zwłaszcza dla celów decyzyjnych. Dużą trudność sprawia ocenienie jakości prognoz *ex ante*, gdy są one dość możliwe do osiągnięcia. Wtedy przyjmuje się często podejście oparte na konsensusie prognoz – przyjmuje się np. średnią z wielu prognoz i ocenia prognozy o skrajnych wartościach jako najmniej wiarygodne (choć *ex post* wcale tak nie musi być).

4. MODELE MAKROEKONOMICZNE STOSOWANE W EWALUACJI W POLSCE

Jak dotąd, w procesie ewaluacji oddziaływania funduszy unijnych w Polsce wykorzystane były cztery modele:

- model HERMIN dr. Johna Bradleya, wdrażany w Polsce przez Wrocławską Agencję Rozwoju Regionalnego,
- model d. Instytutu Koniunktur i Cen Handlu Zagranicznego (obecnie: Instytut Badań Rynku, Konsumpcji i Koniunktur),
- model MaMoR2 Instytutu Badań nad Gospodarką Rynkową,
- model EU-ImpactMod Instytutu Badań Strukturalnych.

Prezentację modeli rozpoczniemy w kolejności chronologicznej – według pierwszeństwa ich stosowania do potrzeb ewaluacji.

Model HERMIN

Model ten jest najstarszym z modeli makroekonomicznych stosowanych w procesie ewaluacji w Polsce. Jest on jednocześnie jednym z najstarszych modeli makroekonomicznych stosowanych do dzisiaj w Europie. Został w całości opublikowany (tj. jego 4-sektorowa wersja), a ponadto od wielu lat dostępne są publikacje na jego temat – stąd można najpełniej go opisać, co jest dużą zaletą tego modelu. Autor modelu ponadto wielokrotnie (i obiektywnie) wychodzi poza swoją tematykę, dokonując porównań ze sobą różnych modeli.

Model ten powstał w 1982 roku w ESRI (Economic and Social Research Institute) – słynnym, irlandzkim instytucie badawczym. Wzorowany był na większym modelu Komisji Europejskiej o nazwie HERMES, stworzonym na początku lat 80. Autorem modelu jest b. pracownik ESRI – dr John Bradley. Model ten został skonstruowany z myślą o małych, otwartych gospodarkach, będących nowymi członkami UE. Umożliwiał bowiem stosowanie krótkich szeregów czasowych i radził sobie z brakiem danych sektorowych[8]. Dla zastosowania metod ekonometrycznych w modelowaniu wymagane są natomiast na ogół dłuższe szeregi czasowe. Stąd zdarza się, że w modelu HERMIN równania są kalibrowane.

Model ten w 4-sektorowej wersji (HPO4) składa się z następujących sektorów: przemysłowego, usług rynkowych, rolnictwa, usług publicznych (według terminologii stosowanej w modelu – usługi nierynkowe). W wersji 5-sektorowej wyodrębniony został sektor budownictwa (ze względu na duże zainteresowanie nim w ostatnich latach)[9].

Model tworzą trzy bloki strukturalne: podażowy, absorpcji i dystrybucji przychodów.

Powiązania pomiędzy głównymi częściami modelu przedstawia rysunek 2.

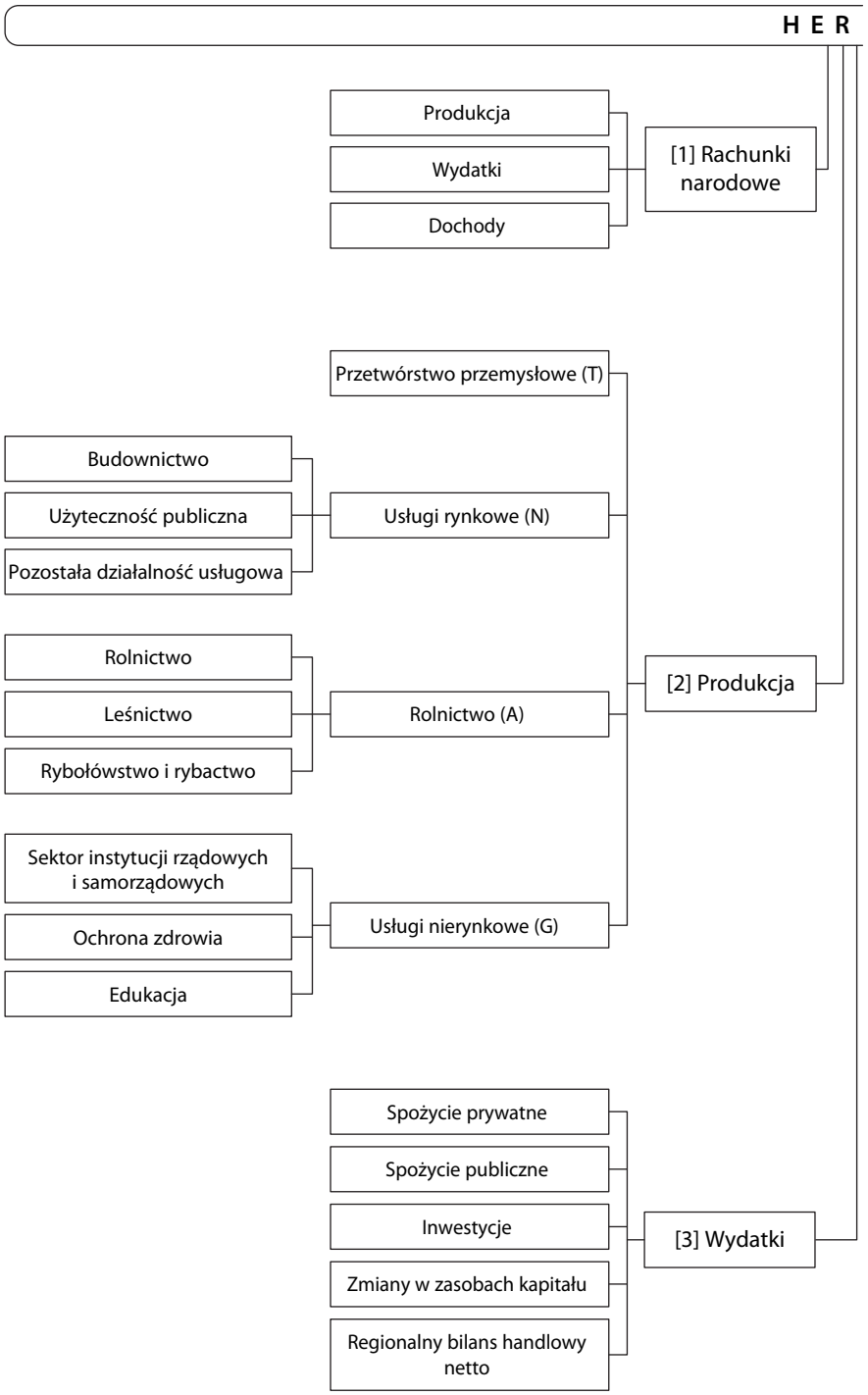
Model zawiera (w 4-sektorowej wersji) 262 równania (w tym 162 nieliniowe), 406 zmiennych (w tym 257 zmiennych endogenicznych i 149 zmiennych egzogenicznych).

Podstawą modeli typu HERMIN są konwencjonalne mechanizmy keynesowskie. Stąd relacje pomiędzy wydatkami a dochodami powiązane są przez mechanizmy mnożnikowe. Strona podażowa modelowana jest w części keynesowsko (przez popyt), ale też w sposób neoklasyczny

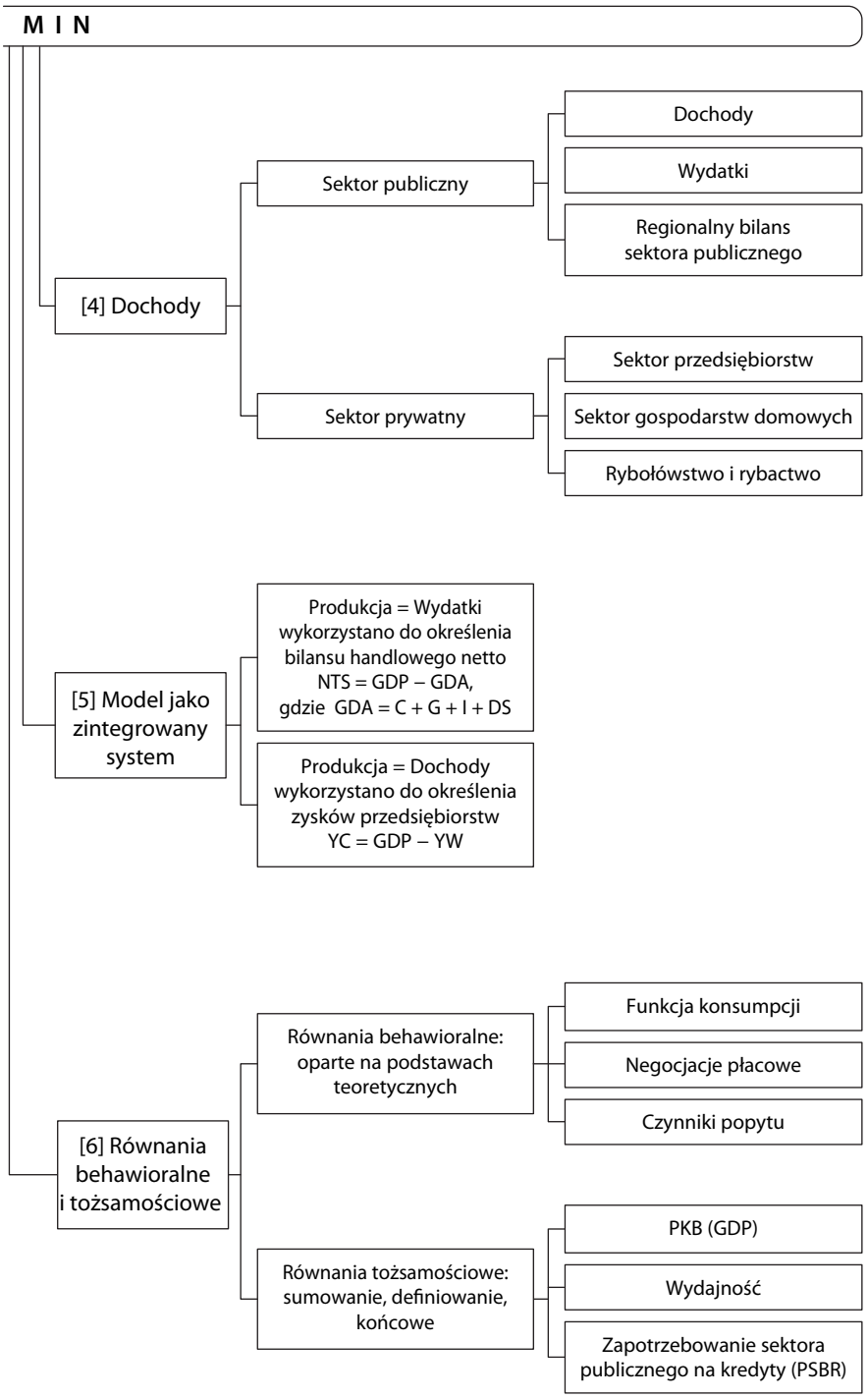
---

[8] Szczególnie istotna była kwestia krótkich szeregów: w nowych krajach członkowskich bywały na ogół kłopoty z posiadaniem odpowiednio rozbudowanych baz danych statystycznych, a także problemy z jakością (i porównywalnością) posiadanych danych.

[9] Dla Ministerstwa Rozwoju Regionalnego zbudowano model HERMIN w wersji 10-sektorowej, oznaczony jako HPON (zawierał on m.in. sektor budownictwa). Model ten nie jest obecnie wykorzystywany, również ze względu na duże nakłady prac konieczne z jego utrzymaniem, przetwarzaniem baz danych itp.

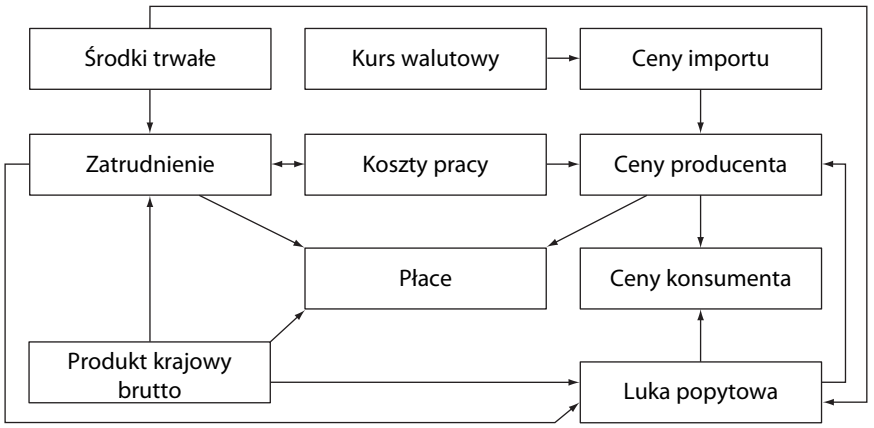


Rysunek 2. Model HERMIN – schemat modelowania

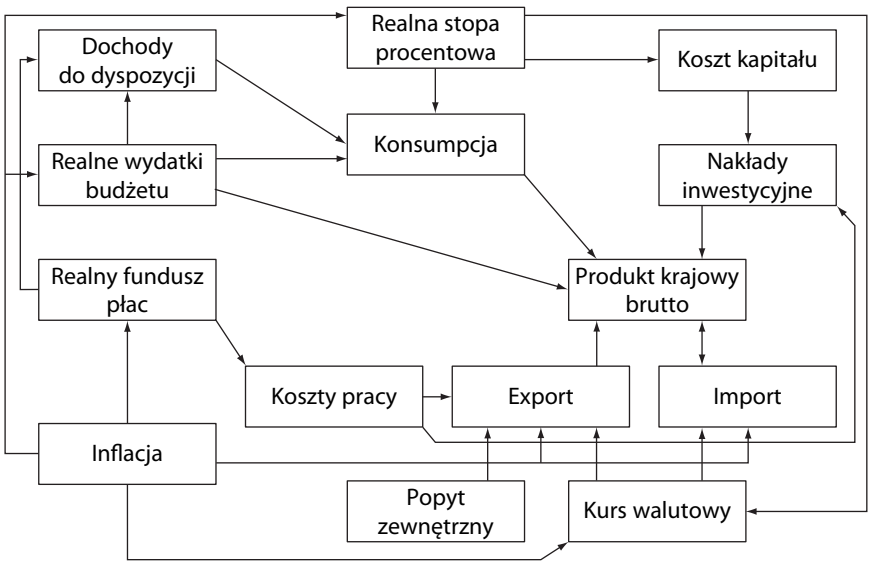


Źródło: Na podstawie Zaleski et al. 2005: 10.

a) strona podażowa (podstawowe zależności)



b) strona popytowa (podstawowe zależności)



Rysunek 3. Schemat modelu polskiej gospodarki służącego do weryfikacji diagnozy i prognozy makroekonomicznej

Źródło: Karpińska-Mizielińska et al. 2006: 161.

(podaż determinowana też cenami i konkurencyjnością kosztową). Popyt na czynniki produkcji w przemyśle i usługach rynkowych określa funkcja produkcji CES. Model nie uwzględnia endogenicznych mechanizmów wzrostu gospodarczego, ani sposobów podnoszenia kapitału ludzkiego[10].

[10] Ta ułomność modelu i potrzeba jej usunięcia podnoszona jest też przez innych autorów – por. Szafrański 2006.

W model HERMIN wbudowany jest model Philipsa, który zakłada wymienność między stopą bezrobocia a stopą inflacji[11].

Model d. Instytutu Koniunktur i Cen Handlu Zagranicznego

Model ten został określony przez autorów jako model hybrydowy II generacji. Przygotowany na zlecenie Ministerstwa Gospodarki, wykorzystywany był dla przygotowania średniookresowych prognoz makroekonomicznych, w tym szczególnie w zakresie związanym z handlem zagranicznym.

Model ten łączy „neoklasyczną teorię zapewniającą długookresową równowagę gospodarce, z teorią nowokeynesowską, bazującą na zmianach popytu. Takie połączenie powoduje nakładanie się na długookresową ścieżkę wzrostu, średniookresowych wahań koniunkturalnych" (Karpińska-Mizielińska et al. 2006: 112). W modelu stosuje się funkcję produkcji Cobba-Douglasa oraz CES. Model wydaje się dość dobrze wyjaśniać czynniki wzrostu gospodarczego: uwzględnia np. inwestycje w rozwój zasobów ludzkich (w sposób analogiczny do tego, jak traktuje się nakłady na badania i rozwój w endogenicznych modelach wzrostu), które mają efekt długofalowy, natomiast bezpośrednie wsparcie sektora produkcyjnego skutkuje przejściowym wzrostem dynamiki PKB. Takie podejście do modelowania sprawia, że prognozowana przez d. IKiCHZ ścieżka wzrostu gospodarczego była inna niż w przypadku modelu HERMIN: zawierała się w przedziale 4,0–4,4% (bez uwzględnienia oddziaływania funduszy unijnych), podczas gdy „w modelu HERMIN tempo to rośnie wykładniczo (do nieskończoności w odpowiednio długim okresie)" (Karpińska-Mizielińska et al. 2006: 114).

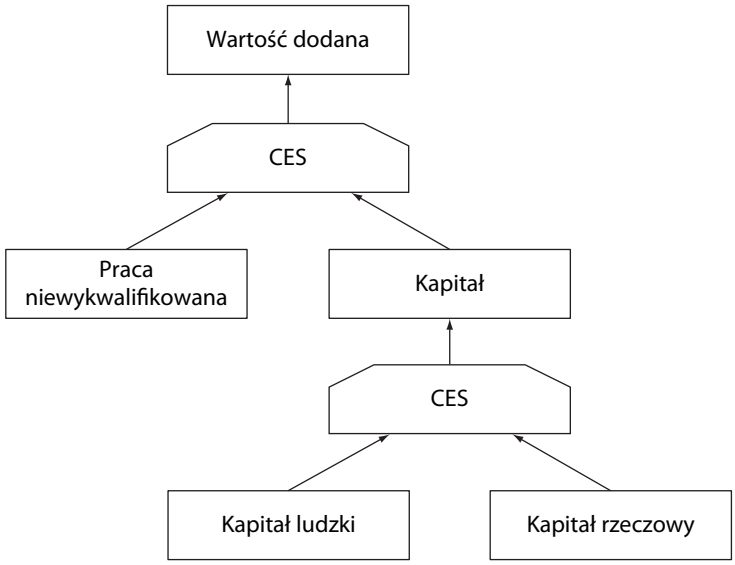
Na rysunku 3 przedstawione zostały główne mechanizmy strony popytowej i podażowej modelu.

Model MaMoR2

Jest to model Instytutu Badań nad Gospodarką Rynkową (autorstwa Tomasza Kaczora); model typu CGE. Uwzględnia trzy czynniki produkcji: praca (osób niewykwalifikowanych), kapitał ludzki oraz rzeczowy, połączone w ramach funkcji produkcji CES.

Model utrzymuje założenie małej, otwartej gospodarki (ceny krajowe są „importowane" ze świata), kurs walutowy jest egzogeniczny. „Najważniejsze parametry modelu zostały bądź skalibrowane, bądź przyjęte na podstawie literatury" (Kaczor 2006: 6). Inaczej niż HERMIN, model ten w większej mierze wydaje się przewidywać międzyokresową optymalizację. Podobnie jak HERMIN, model MaMoR2 zakłada stałość kursu

[11] Współcześnie uznaje się, że krzywa Philipsa ma ograniczone zastosowanie – jedynie do krótkiego okresu.



Rysunek 4. Schemat procesu produkcyjnego w modelu MaMoR2

Źródło: Kaczor 2006: 7.

walutowego. *Novum* tego modelu polega na tym, że model ten „tworzy" sektor krajowy (i prognozy dla gospodarki całego kraju) z gospodarek regionów, które są w dużej mierze autonomiczne względem siebie (Kaczor i Mitroczuk 2007).

Model EU-ImpactMod

Model Instytutu Badań Strukturalnych należy do najnowszej klasy modeli makroekonometrycznych (Bukowski et al. 2008). Jest to model małej gospodarki, w której zmienne dotyczące zagranicy są egzogeniczne. Obejmuje sferę realną (bez pieniężnej), w tym rozbudowany segment fiskalny. Środki unijne w modelu podzielone są na aż siedem kategorii wydatków: zadania związane z podnoszeniem umiejętności ludzi dorosłych, wydatki niezbędne do obsługi funduszy strukturalnych, środki współfinansujące tworzenie infrastruktury edukacyjnej, inwestycje w kapitał publiczny (głównie infrastrukturę ochrony środowiska, komunalną i ochrony zdrowia), inwestycje w infrastrukturę transportową oraz łącznościowo-informacyjną, wydatki na kulturę, na dofinansowanie sektora edukacyjnego i jego funkcji naukowej (w tym zwłaszcza na wynagrodzenia dla instruktorów), a także środki Wspólnej Polityki Rolnej (Bukowski et al. 2008a).

Podkreślić można, że model wydaje się odzwierciedlać sporą część najnowszego dorobku nauk ekonomicznych. Jego autorzy wykazują dużą znajomość literatury przedmiotu.

5. PORÓWNANIE WYNIKÓW EWALUACJI POLITYKI SPÓJNOŚCI W EUROPIE I W POLSCE

Modele makroekonomiczne stosowane są na przykład po to, by sprawdzić, jakie będą efekty interwencji. Taką interwencją może być też polityka spójności czy napływ funduszy strukturalnych do danego kraju.

Kwestią bardzo interesującą jest uzyskanie odpowiedzi na pytanie: który z modeli stosowanych w ewaluacji jest lepszy? Niestety, odpowiedź na to pytanie nie jest ani łatwa, ani w pełni możliwa. Każdy z modeli jest inny, bazuje na innych teoriach, przyjmuje inne założenia. Jednakże naukowcy (i praktycy – twórcy modeli) starają się jakoś ten problem rozwiązać i w tym celu stosują – oprócz analizy założeń, konstrukcji modeli (co w dużym skrócie zaprezentowano powyżej) – również porównanie wyników uzyskiwanych za ich pomocą. Ponownie, nie zawsze dają one jednak jasne konkluzje.

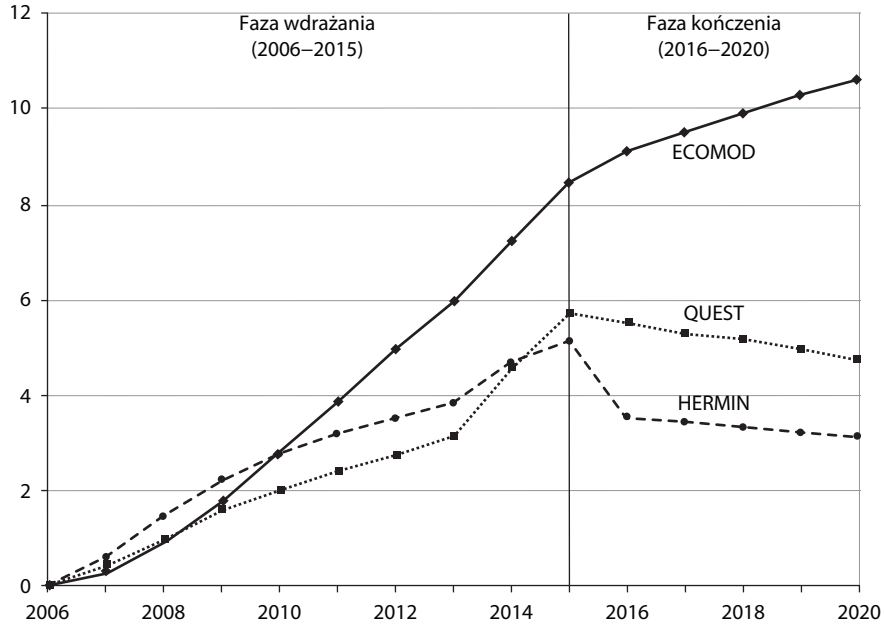
„Symulacje wpływu funduszy unijnych od 2000 do 2006 roku, przeprowadzone dla Hiszpanii, wskazują uzyskiwanie większych wyników przy wykorzystaniu modelu HERMIN (wzrost PKB o 2,1%), niż w pracy Beutela (2002) – 1,7%, przy wykorzystaniu modelu MOISEES – 1,0% czy QUEST II – 0,6%" (EC 2005: 29). Szacunki otrzymane przy wykorzystaniu modelu HERMIN były najwyższe, spośród czterech wyżej wymienionych modeli.

Gdy porównamy wyniki obliczeń dla Polski (dotyczących oddziaływania funduszy unijnych w ubiegłych i w kolejnych latach) widzimy, że najsłabszy pod względem trafności wyników wydaje się model ECOMOD, a wyniki modelu HERMIN są wyższe od wyników QUEST II w trakcie napływu funduszy UE do kraju, a niższe – po zakończeniu perspektywy finansowej. Dotyczy to m.in. wyników dotyczących wzrostu gospodarczego. W tabeli 1 oraz na rysunku 5 zostały zaprezentowane bardziej

Tabela 1. Porównanie wyników symulacji oddziaływania funduszy UE na lata 2007–2013 na PKB w przypadku Polski

Rok	ECOMOD	HERMIN	QUEST II
2006	0,0	0,0	0,0
2010	2,6	2,6	1,9
2015	8,2	4,6	5,0
2020	10,4	3,5	4,2

Źródło: BICEPS 2008: 11.



Rysunek 5. Wpływ funduszy unijnych na poziom PKB w Polsce według modeli ECOMOD, QUEST II i HERMIN (procentowe odchylenie od scenariusza bazowego)[12]

Źródło: Bradley i Untied 2007: 12.

szczegółowe dane. Dotyczą one różnicy pomiędzy scenariuszem bazowym (bez napływu funduszy UE) a końcowym (wraz z ich napływem i oddziaływaniem).

Spostrzeżenie dotyczące wyników modelu ECOMOD znajduje potwierdzenie również w danych dotyczących zatrudnienia.

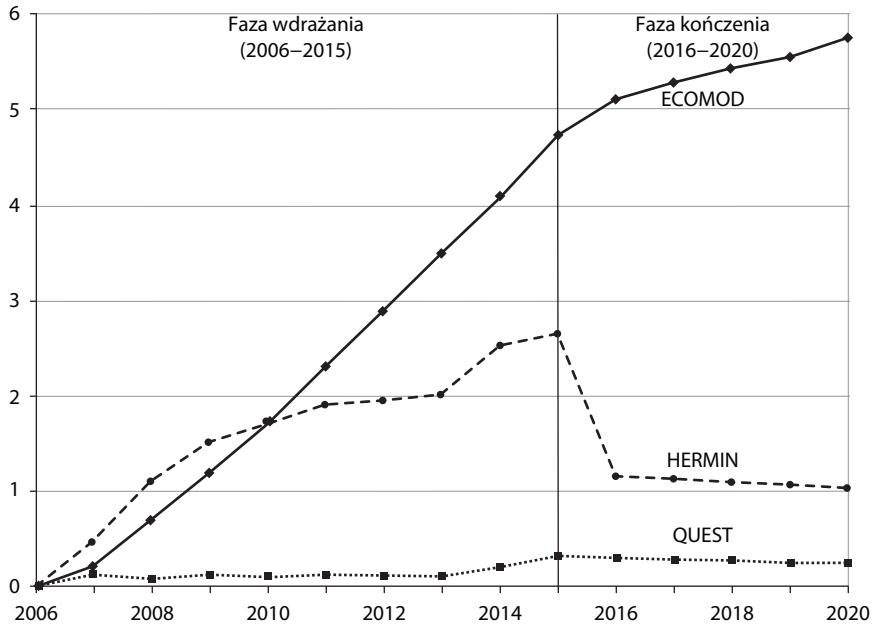
Z powyższego porównania widzimy, że model ECOMOD dostarczał zdecydowanie wyższych wyników niż pozostałe, przy czym wyniki modeli HERMIN i QUEST II są do siebie dość zbliżone[13].

---

[12] Dane dla QUEST od 2016 roku pochodzą spoza modelu. Faza wdrażania 2006–2015, faza kończenia 2016–2020 (ostatnim rokiem finansowania z perspektywy 2007–2013 jest 2015).

[13] Bradley i Untied 2007 upatrują różnice między wynikami modelu QUEST II i HERMIN (brak informacji nt. modelu ECOMOD uniemożliwia jego głębszą analizę i porównania z innymi modelami) w:

- przyjęciu różnych funkcji produkcji,
- różnej liczbie sektorów gospodarki w obu modelach,
- różnych założeniach relacji gospodarki polskiej z zagranicą,
- różnym przyjęciu efektów wydajności pracy: w modelu Komisji wzrost wydajności w pełni przekłada się na wzrost płac, a w modelu J. Bradleya część zysków przedsiębiorcy z tytułu wzrostu wydajności pozostaje w firmie (dzięki czemu może np. obniżyć cenę oferowanych produktów),
- innej perspektywie oczekiwań.



Rysunek 6. Wpływ funduszy unijnych na poziom zatrudnienia w Polsce według modeli ECOMOD, QUEST II i HERMIN (procentowe odchylenie od scenariusza bazowego)[14]

Źródło: Bradley i Untied 2007: 13.

Tabela 2. Wpływ funduszy UE na PKB w Łotwie i w Estonii[15]

Rok	2000	2002	2004	2006	2008	2010	2015
Łotwa	2,98	6,10	6,25	6,12	5,99	5,84	
Estonia			7,29	8,52	13,28	13,52	6,58

Źródło: BICEPS 2008: 12.

Nie oznacza to, że nie występują trudności w przypadku modelowania przy wykorzystaniu HERMIN-a dla innych krajów transformacji systemowej, co zauważono w: BICEPS 2008 (por. tabela 2).

Wyniki dla 2004 roku pokazały, że względem scenariusza bazowego wzrost gospodarczy na Łotwie byłby wyższy o 6,3 p.p., a w Estonii o 7,3 p.p., podczas gdy rzeczywisty wzrost gospodarczy w 2004 roku wynosił na Łotwie 8,7%, zaś w Estonii – 7,3%. Szczególnie wynik dla Estonii wydaje się nierealny do osiągnięcia (wynikałoby z niego, że bez funduszy unijnych gospodarka estońska w ogóle by się w 2004 roku nie rozwijała).

[14] Uwagi jak do rysunku 5.

[15] Dane pokazują różnice między scenariuszem bez oraz z użyciem funduszy UE.

Na podsumowanie powyższych, bardzo skrótowych rozważań na temat jakości modeli, przytaczam poniższe: „Własny model Komisji QUEST, który uwzględnia silne mechanizmy wypierania (…) ma skłonność do dawania niższego wpływu funduszy strukturalnych. Ostatnie badanie polityki spójności w holenderskim CPB [chodzi o Holenderskie Biuro Analizy Polityki Gospodarczej – K.P.] (…) sugeruje znacznie niższy wpływ polityki", niż jest to w wynikach dostarczanych przez model HERMIN. „Zatem metodologia bazująca na modelach HERMIN jest jedynie jedną z wielu możliwych alternatyw" (Bradley et al. 2003: 15)[16]. Współautor modelu QUEST II odnotowuje jednak, że przeszłe ewaluacje funduszy strukturalnych oparte były na „modelach makroekonomicznych [tj. na modelu HERMIN – K.P.], w których strona popytowa ma w zasadzie keynesowską naturę i nie występują w nich efekty wypierania" (Veld 2007: 2).

6. PORÓWNANIE PROGNOZ ROZWOJU POLSKIEJ GOSPODARKI I ODDZIAŁYWANIA FUNDUSZY UNIJNYCH W POLSCE DO 2020 ROKU

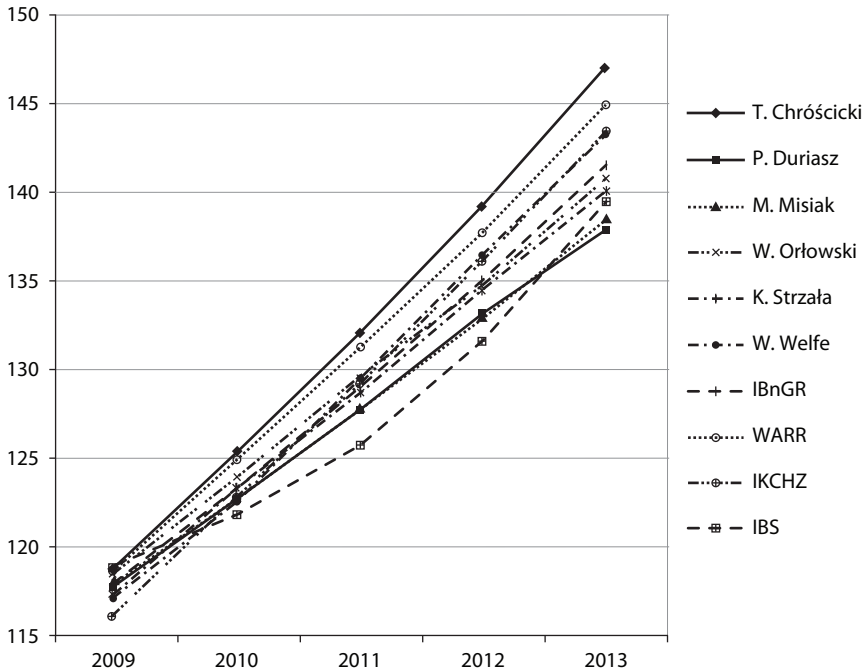
Niekiedy krótkookresowe, eksperckie prognozy makroekonomiczne mogą być trafniejsze niż wyniki otrzymywane na podstawie modeli ekonometrycznych. Częściowo spowodowane jest to tym, że autorzy modeli przykładają większą wagę do odzwierciedlenia w nich struktury gospodarki i jej mechanizmów, niż do trafności prognoz. Stąd przykładowo model HERMIN, MaMoR2 czy EUImpactMod nie są modelami typowo prognostycznymi.

W praktyce jednak to trafność uzyskiwanych prognoz z danego modelu w dużej mierze świadczy o tym, na ile wyniki uzyskane przy zastosowaniu danego modelu są wiarygodne. O ile nie jesteśmy w stanie sprawdzić, jakie rzeczywiście było oddziaływanie funduszy UE np. na PKB, czy na bezrobocie, o tyle możemy wskazać, na ile uzyskiwane prognozy różniły się później od rzeczywistości.

Rysunek 7 zawiera prognozy na następne pięć lat, przeliczone tak, by rokiem bazowym był 2006. Pokazane jest nie roczne tempo wzrostu

[16] W modelu QUEST zakłada się, że oczekiwania mają charakter antycypacyjny (prognostyczny) (*forward-looking*), podczas gdy w modelu HERMIN – adaptacyjny (*backward-looking*). Innymi słowy, w modelu Komisji Europejskiej zakłada się, że ludzie formułują swoje oczekiwania na podstawie przypuszczenia na temat tego, co może się dziać w przyszłości, podczas gdy w drugim modelu kierują się przeszłymi wynikami gospodarczymi. Powoduje to, że efekty wypierania (wydatków prywatnych przez wydatki publiczne) w modelu QUEST są większe, niż w modelu HERMIN. Por. też Bradley et al. 2003: 5.

O tym, jak ważne jest właściwe szacowanie efektu wypierania, świadczy wysokość tego efektu: dla hiszpańskiej Andaluzji oszacowano ją na aż 32%, tzn. wzrost wydatków publicznych o 100 EUR zmniejsza wartość wydatków prywatnych o 32 EUR (Fuente 2003: 17).



Rysunek 7. Wybrane wyniki prognoz wzrostu gospodarczego w latach 2009–2013 (2006 rok = 100)[17]

Źródło: obliczenia autora na podstawie prognoz IBnGR (Kaczor i Socha 2008: 39), WARR (Zaleski et al. 2008: 23–24), IKCHZ (Karpińska-Mizielińska et al. 2006), IBS (Bukowski et al. 2008: 26, 28) oraz prognoz zebranych przez PTE 2007: 25.

gospodarczego, ale skumulowane – prezentujące prognozowaną zmianę poziomu PKB względem roku 2006. Zebrane zostały prognozy pojedynczych ekspertów oraz pochodzące z modeli makroekonomicznych uwzględniających napływ (i oddziaływanie) funduszy unijnych.

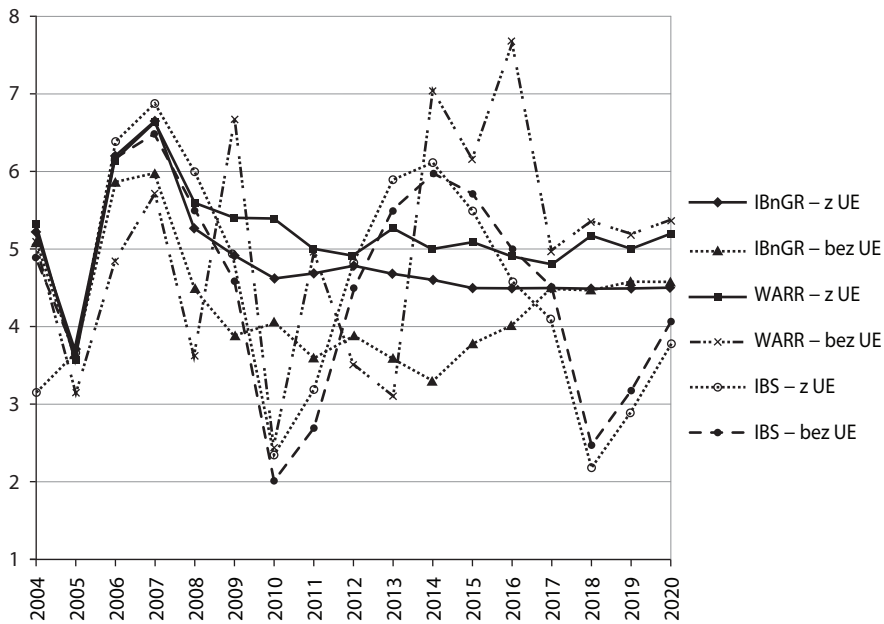
Jak widzimy, mimo pewnych rozbieżności, wszyscy progności wskazywali na bardzo dynamiczny wzrost gospodarczy w kolejnych latach. Prognozy uzyskiwane za pomocą modeli były na ogół nieco bardziej optymistyczne niż prognozy eksperckie, przy czym najbardziej optymistyczne spośród nich były szacunki dostarczone przez WARR, zaś najmniej – IBS.

Poniżej zostaną zaprezentowane dokładniejsze wyniki prognoz wzrostu gospodarczego Polski oraz ewaluacji oddziaływania funduszy unijnych, opublikowane w lipcu 2008 roku. Każdy z zastosowanych w tym celu modeli należy do innej klasy, co znacząco wzbogaca naszą wiedzę na temat scenariuszy rozwoju i oddziaływania. I tak:

[17] Prognozy na wcześniejsze lata pominięto (ze względów graficznych).

- model HERMIN bazuje na technikach modelowania z lat 70. i 80. XX wieku, opiera się w dużej mierze na mechanizmach keynesowskich (mnożnikowych),
- model MaMoR2 należy do klasy CGE, popularnej w latach 80. i 90. XX wieku,
- model EU-ImpactMod to model z rodziny DSGE, rozwijanej w ciągu ostatnich kilku lat.

Widzimy (rysunek 8), że w przypadku modelu bazującego na najnowszych technikach (tj. DSGE), oddziaływanie funduszy unijnych na wzrost gospodarczy jest niewielkie. Nieco większe są one w trakcie okresu programowania w przypadku modelu klasy CGE, po czym wzrost gospodarczy w obu scenariuszach jest prawie taki sam. Największe rozbieżności widoczne są w przypadku najstarszego modelu – widać dużą podatność wzrostu gospodarczego na wahania wykorzystania środków unijnych. Z innych, interesujących kwestii, zauważyć można bardzo duże wahania wzrostu gospodarczego w scenariuszu bez napływu funduszy



Rysunek 8. Wzrost gospodarczy Polski: bez i wraz z uwzględnieniem napływu środków unijnych do końca 2015 roku (w %)[18]

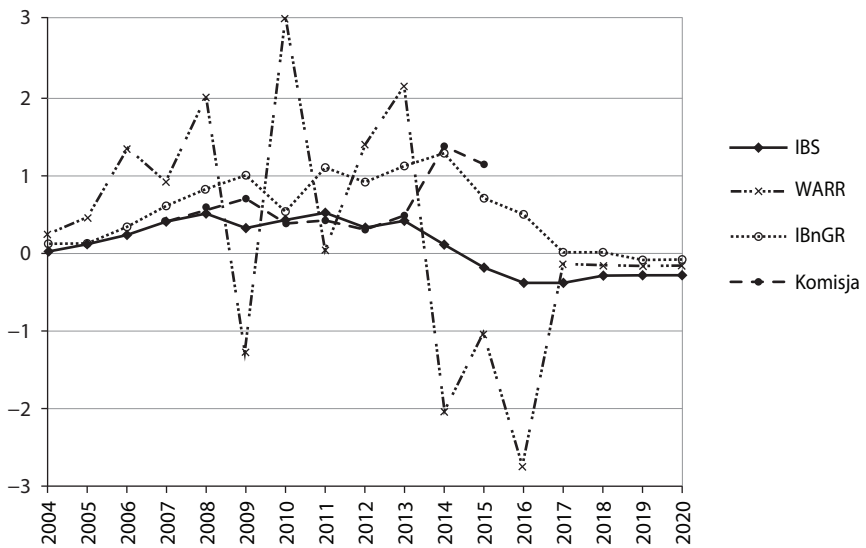
Źródło: Bukowski et al. 2008: 26, 28; Kaczor i Socha 2008: 39; Zaleski et al. 2008: 23–24 oraz obliczenia własne na ich podstawie.

[18] Wzrost gospodarczy – wzrost realnego PKB w danym roku. Bez UE – scenariusz bez napływu i oddziaływania funduszy strukturalnych (w ramach NPR 2004–2006 oraz NSRO 2007–2013); z UE – wraz z nimi.

strukturalnych w przypadku modelu HERMIN, a także ciekawy przebieg danych IBS, które wyraźnie naśladują cykl koniunkturalny[19].

Porównajmy poniżej same zakresy oddziaływania funduszy unijnych na polską gospodarkę według tych trzech modeli. Dla porównania skali dodano też wyniki uzyskane za pomocą modelu QUEST II Komisji Europejskiej, choć obejmują one wyłącznie fundusze z perspektywy 2007–2013 (obliczenia były przeprowadzone ok. roku przed pozostałymi). Co ciekawe, w początkowych latach oddziaływanie funduszy unijnych na gospodarkę może być niższe niż wielkość samego napływu funduszy, co Veld (2007) tłumaczy efektem wypierania[20].

Widzimy, że o ile IBS i IBnGR podają dość stabilne w czasie i dość zbliżone do siebie wyniki, o tyle wyniki szacunków WARR cechują się bardzo dużą zmiennością i dużą rozbieżnością wyników: od +3 p.p. do prawie –3 p.p. Skala oddziaływania funduszy według WARR w 2010 roku (2,98 p.p.) byłaby 5–7-krotnie wyższa, niż według IBS i IBnGR (0,4 i 0,5 p.p.). Na rysunku 9 dodano też szacunki Komisji Europejskiej i choć nie są one w pełni porównywalne, to pokazują przynajmniej skalę



Rysunek 9. Oddziaływanie funduszy unijnych na wzrost gospodarczy Polski (w p.p.)[21]

Źródło: Bukowski et al. 2008: 28; Kaczor i Socha 2008: 39; Zaleski et al. 2008: 24 oraz obliczenia własne na podstawie Veld 2007: 24.

---

[19] IBS włączył w prognozy bazowe czynnik cykliczny, przy czym cykl koniunkturalny w Polsce ma – według IBS – długość 8 lat (zdaniem autora jest ona bliższa 10 latom). IBS uzyskał tę wartość przez ekstrapolację cyklu z lat 1992–2003.

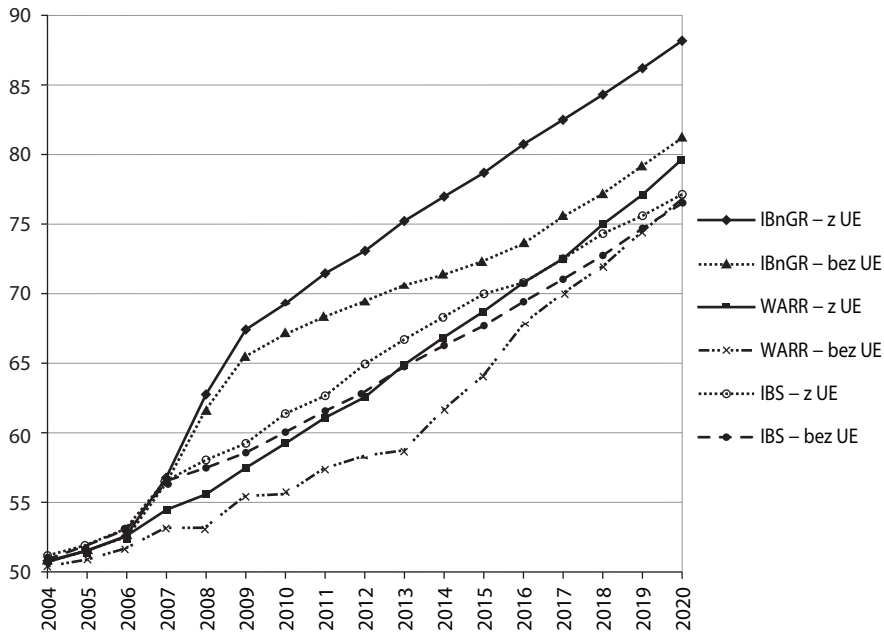
[20] Bazując na tych wyjaśnieniach i danych, efekt wypierania w Polsce można by oszacować na 0,2–0,8 p.p. PKB w latach 2007–2010. Są to więc już dość duże wielkości.

[21] Szacunki Komisji Europejskiej (lata 2007–15) nie są w pełni porównywalne z pozostałymi.

oddziaływania – dyspersja wyników nie jest tak duża jak w przypadku danych Wrocławskiej Agencji Rozwoju Regionalnego.

W kilku miejscach wyniki uzyskiwane za pomocą modelu HERMIN są dość zastanawiające – np. w 2009 roku fundusze unijne miałyby mieć ujemny wpływ na polską gospodarkę. Również napływ funduszy unijnych i ich oddziaływanie (*externalities*) w latach 2014–2020 miałyby być wyraźnie ujemne. To, że prognozy na lata kolejnej perspektywy finansowej pokazują ujemne oddziaływanie funduszy, nie jest zaskakujące. Jest to sztuczny efekt, wynikający z założenia braku napływu do Polski funduszy unijnych w kolejnej perspektywie finansowej. Zastanawiająca jest jednak tak duża skala ujemnego oddziaływania w latach 2014–2016, prognozowana przez model HERMIN (należałoby wyjaśnić przyczyny tego zjawiska i sprawdzić, jak je minimalizować).

Na rysunku 10 zaprezentowane zostaną ponadto ścieżki konwergencji Polski z UE-27, również z uwzględnieniem oddziaływania funduszy unijnych (z perspektywy finansowej 2007–2013) oraz bez niego.



Rysunek 10. Wzrost gospodarczy Polski względem UE27: bez i wraz z uwzględnieniem napływu środków unijnych do końca 2020 roku (UE-27 = 100)[22]

Źródło: Bukowski et al. 2008: 26, 28; Kaczor i Socha 2008: 42; Zaleski et al. 2008: 27.

[22] Według WARR wzrost PKB *per capita* PPS dla UE w latach 2010–2020 założono na poziomie 2% (dla lat wcześniejszych – prognozy Eurostatu). W pozostałych przypadkach – nie podano. Dane historyczne w przypadku IBS różniły się od danych podawanych przez Eurostat.

Porównania rysunków 8, 9 i 10 nasuwają wątpliwości co do scenariusza założonego dla UE w przypadku obliczeń IBnGR. Widzimy znaczne przyspieszenie konwergencji w latach 2008–2009, co nie ma odzwierciedlenia w danych dotyczących wzrostu gospodarczego Polski. Według autora modelu, mogło być to spowodowane założeniem aprecjacji złotego.

Wyniki pozostałych dwóch instytucji są ze sobą zbieżne w długim okresie, szczególnie w zakresie scenariusza bez oddziaływania funduszy UE – zarówno IBS, jak i WARR na 2020 rok prognozowały bardzo podobny poziom rozwoju Polski względem UE-27.

Na zakończenie popatrzmy jeszcze na oddziaływanie funduszy unijnych na rynek pracy, a dokładniej – na tworzenie nowych miejsc pracy, zgodnie z aktualnymi wynikami dostarczanymi przez trzy modele.

Różnice pomiędzy wynikami modeli są dość duże (pamiętać jednak należy o nieporównywalności wyników): zarówno jeśli chodzi o wyniki dla poszczególnych lat, jak i wartości skumulowane. WARR wskazuje

Rysunek 11. Oddziaływanie funduszy unijnych na tworzenie nowych miejsc pracy w poszczególnych latach (w tys.)[23]

Źródło: Bukowski et al. 2008: 28; Kaczor i Socha 2008: 47; obliczenia własne na podstawie Zaleski et al. 2008: 37.

[23] Użyte na rysunku dane WARR to niepodawane przez Agencję „nowo utworzone miejsca pracy ogółem", obejmujące właściwie „trwałe, nowo utworzone miejsca pracy" (Zaleski et al. 2008: 24), ale „nowo utworzone miejsca pracy ogółem" (Zaleski et al. 2008: 37), będące sumą nowych i odtworzonych miejsc pracy (wartości te ponadto zostały wyliczone na podstawie danych skumulowanych). Prezentowane dane – ze względu na inną definicję obrazowanego pojęcia oraz sposób obliczeń – są zbliżone tematycznie (jedyne, tego typu, nieskumulowane dane pochodzące z trzech modeli), aczkolwiek nieporównywalne.

na bardzo wysokie wzrosty liczby nowych miejsc pracy od początku poprzedniego okresu programowania (od 50 tys. w 2004 roku do 170 tys. w 2006 roku), zaś wyniki IBnGR i IBS są pod tym względem ostrożniejsze i zbliżone do siebie – do 2007 roku, od którego oba modele wskazują inne dane i ich trendy. Wyniki podane przez IBnGR pokazują możliwość ubytku nowych miejsc pracy (w 2010 roku oraz po 2015 roku), dane WARR od 2014 roku przyjmują stałą wartość, zaś IBS wskazuje na coraz bardziej przyrastającą liczbę nowych miejsc pracy do 2013 roku, po czym przyrosty te zmniejszają się, by w 2020 roku „spotkać" się z poziomem podanym przez WARR.

7. MODELE EKONOMICZNE STOSOWANE NA POZIOMIE REGIONALNYM

Modelowanie rozwoju regionalnego staje się coraz bardziej popularne w ostatnich latach. Przyczynia się też do tego polityka spójności i konieczność jej ewaluacji. Spośród kilku modeli, które były stosowane w tym zakresie, wymienić można:

- REMI – model ekonometryczny, który jest stosowany dla potrzeb regionalnych. Struktura modelu zakłada „przed-szacowanie" wskaźników, dzięki czemu można na jego podstawie uzyskać więcej informacji, niż stosując jedynie dostępne dane krajowe, co jest ważne w przypadku zastosowań regionalnych (gdzie występują problemy z dostępnością danych) (Treyz i Treyz 2003). Według REMI (www.remi.com), Massachusetts Institute of Technology podał m.in. następujące charakterystyki modelu: jest skalibrowany do warunków lokalnych, używając stosunkowo dużej ilości danych lokalnych; ma wyjątkowo silne podstawy teoretyczne; uwzględnia cykle koniunkturalne.
- HELM – model użyty do ewaluacji *on-going* dla prowincji Hainaut w Belgii. Rozważa on efekty wzrostu endogenicznego powiązane z długoterminową produktywnością. Wpływ interwencji funduszy strukturalnych na produktywność (pracy) ograniczony w nim został do sektora produkcyjnego.
- W Hiszpanii dla oszacowania efektów funduszy strukturalnych w regionie Galicja w latach 1994–1999 oraz 2000–2006 zastosowano model podobny do HERMIN.
- W Polsce, spośród czterech modeli makroekonomicznych stosowanych dla potrzeb ewaluacji, dwa były wykorzystane na poziomie regionalnym. Był to model HERMIN oraz MaMoR2[24]. Submodele regionalne HERMIN były konstruowane jako uproszczone kopie modeli krajowych, natomiast w modelu MaMoR2 stanowią one integralną część modelu całej gospodarki. Nie oznacza to jednak, że są to rzeczywiście

[24] Ośrodek łódzki (prof. W. Welfe i zespół) również pracuje nad modelem dla województwa łódzkiego. Prace mają zakończyć się do końca 2008 roku.

modele poszczególnych regionów: postać równań jest bowiem taka sama, a różne są jedynie dane wejściowe (a także większość parametrów). Stąd te regionalne modele nie odzwierciedlały specyfiki danego województwa (np. innych mechanizmów rozwojowych, innych czynników wpływających na emigrację itd.). Niemniej jednak stanowią cenne (i właściwie jedyne) źródło prognoz rozwoju regionalnego naszego kraju, uwzględniające (i wydzielające) oddziaływanie funduszy unijnych.

W tabeli 4 zaprezentowano wybrane wyniki obu modeli opublikowane w 2006 roku, odpowiednio przeliczone przez autora (por. Piech

Tabela 4. Prognozy rozwoju regionalnego, Polska = 100

Województwo	2005	2010		2015		2020	
		WARR	IBnGR	WARR	IBnGR	WARR	IBnGR
dolnośląskie	103,3	105,7	102,9	109,0	102,8	111,9	103,1
kujawsko-pomorskie	87,2	84,6	89,3	83,4	91,4	80,4	91,2
lubelskie	68,3	67,0	70,2	66,4	71,9	66,2	72,3
lubuskie	90,2	88,6	87,1	86,2	85,3	80,7	86,4
łódzkie	91,8	91,9	92,2	94,6	92,7	96,8	92,9
małopolskie	85,3	82,3	86,6	81,9	87,7	83,5	87,4
mazowieckie	158,4	166,8	152,3	173,6	149,6	182,0	149,2
opolskie	82,8	81,8	80,5	83,6	79,9	87,7	79,6
podkarpackie	69,0	66,0	71,9	64,4	74,5	63,6	74,4
podlaskie	74,0	72,9	75,6	71,0	76,7	68,0	76,0
pomorskie	98,2	96,3	98,5	92,4	97,8	87,9	97,9
śląskie	107,9	108,6	111,3	107,1	109,7	104,8	109,7
świętokrzyskie	74,8	74,6	79,5	73,2	81,3	70,4	81,4
warmińsko-mazurskie	76,5	75,0	77,6	73,1	79,8	70,2	79,4
wielkopolskie	106,9	101,4	103,9	95,8	104,3	89,9	104,7
zachodniopomorskie	92,8	92,1	95,3	90,3	94,9	89,5	95,2

Źródło: GUS 2008 i obliczenia własne na podstawie GUS 2008; Bradley et al. 2006 oraz Kaczor 2006a, 2006b; za: Piech 2008: 218.

Tabela 5. Wpływ funduszy unijnych na poziom PKB na mieszkańca w relacji do średniej UE-27 (w cenach bieżących, p.p.)

Województwo	MaMoR2	HERMIN	MaMoR2	HERMIN	MaMoR2	HERMIN
	2007		2013		2015	
dolnośląskie	0,8	0,8	5,8	5,0	7,8	4,6
kujawsko-pomorskie	0,5	0,6	3,9	3,0	5,4	2,2
lubelskie	0,5	0,5	3,7	1,8	5,0	0,7
lubuskie	1,1	0,5	6,4	3,2	8,7	2,3
łódzkie	0,9	0,8	6,0	3,8	7,9	4,1
małopolskie	0,5	0,4	4,0	1,5	5,6	0,5
mazowieckie	0,2	1,0	3,5	4,4	4,9	2,5
opolskie	0,1	0,5	1,5	1,5	2,2	0,9
podkarpackie	0,5	0,4	4,2	2,3	5,8	1,8
podlaskie	0,7	0,4	4,6	1,9	6,3	0,7
pomorskie	0,9	1,1	5,9	4,3	7,5	3,7
śląskie	0,6	0,9	4,9	4,1	6,4	3,3
świętokrzyskie	0,6	0,5	4,2	2,1	5,6	0,9
warmińsko-mazurskie	0,8	0,6	5,7	2,8	7,7	1,8
wielkopolskie	1,0	1,0	4,2	4,1	5,5	3,7
zachodniopomorskie	0,8	1,0	4,4	1,9	5,8	0,6

Źródło: MRR 2008: 131.

2008)[25]. Wyniki prezentują scenariusz uwzględniający napływ funduszy UE. Ważne jest, by interpretując wyniki, pamiętać, że przedstawione są one w porównaniu do średniej dla Polski (=100). Scenariusz dla kraju

[25] Wyniki opublikowane w 2008 roku uniemożliwiają porównanie obu modeli w sposób obrany przez autora (tj. przy założeniu, że Polska = 100). Nie został opublikowany komplet potrzebnych w tym celu danych (dostępne są jedynie dla UE-27=100, a także dane dotyczące wpływu na rozwój regionów względem średniej dla Polski, bez pokazania wartości poziomów rozwoju regionów względem Polska=100). Stąd autor oparł się na własnych szacunkach, bazujących na starszych wynikach WARR i IBnGR.

powstał na podstawie wyników obu modeli na poziomie krajowym, odpowiednio przeliczonym przez autora (nie są to zatem oryginalne wyniki obu instytucji), by suma gospodarek województw dawała wynik dla całego kraju. Tendencje rozwoju Polski do 2020 roku (używając nowszych szacunków) zostały już zaprezentowane wcześniej.

Widzimy, że rozbieżności między wynikami dostarczanymi przez oba modele są częstokroć wysokie. Jest to oczywiście spowodowane różnicami w ich budowie – oba modele na poziomie regionalnym różnią się od siebie jeszcze bardziej niż modele dla gospodarki krajowej. Niemniej jednak, niekiedy generalne wyniki, tendencje zarysowane przez oba modele się pokrywają. I tak, przykładowo, najbogatszym regionem w Polsce nadal pozostanie mazowieckie, zaś województwa najbiedniejsze (lubelskie, podkarpackie, podlaskie, warmińsko-mazurskie) nadal będą należały do grona najuboższych (choć bezwzględny poziom ich rozwoju oczywiście się podniesie). Według kalkulacji bazujących na danych WARR, dystans tych regionów do średniego poziomu rozwoju w Polsce powinien się zwiększyć w 2020 roku, natomiast bazując na danych IBnGR, uzyskujemy przeciwny wniosek. Uśredniając wyniki, należałoby stwierdzić, że pozycja najbiedniejszych regionów względem średniej krajowej nie zmieni się.

W lipcu 2008 roku opublikowane zostały dane pokazujące oddziaływanie funduszy unijnych na polskie regiony, porównując je do średniej unijnej (UE-27=100).

W przypadku tych danych również niekiedy widzimy rozbieżności pomiędzy wynikami modeli dla poszczególnych województw, „jednak niezależnie od zastosowanego narzędzia, pozytywne efekty są obserwowane we wszystkich regionach" (MRR 2008: 131). Zgodnie z obydwoma modelami w 2013 roku, najsłabsze będzie oddziaływanie funduszy w województwie opolskim (po 1,5 p.p.), zaś największe według IBnGR w woj. lubuskim i łódzkim, a według WARR – w województwie dolnośląskim.

8. ZAKOŃCZENIE

Stosowanie zaawansowanych narzędzi makroekonomicznych w procesach ewaluacji nie jest łatwe. Jest to jednak niezmiernie potrzebne. Można oczywiście próbować opierać się na opiniach eksperckich, które częstokroć na poziomie całej gospodarki są trafniejsze, jeśli chodzi o formułowane prognozy, niż wyniki modelowania. Jednakże eksperci bardzo często nie są w stanie odpowiedzieć na szczegółowe pytania, np. o skalę wpływu (oddziaływania) danej interwencji na podstawowe nawet parametry gospodarki, jak PKB, zatrudnienie czy inflacja. Ponadto pozostaje kwestia obiektywności ekspertów, oceny ich wiedzy i doświadczenia itp.,

które są trudne do jednoznacznego ocenienia. Stąd niezbędne jest stosowanie zaawansowanych narzędzi statystycznych w ekonomii, w tym modeli gospodarek (krajowych i regionalnych).

W krajach o dłuższych tradycjach wspierania podejmowania decyzji przez wykorzystanie narzędzi naukowych (np. Holandia), stosowanie modeli makroekonomicznych (ale też mikroekonomicznych) jest kwestią dość oczywistą. W Polsce procesy transformacji sprawiały bardzo duże kłopoty w modelowaniu zjawisk gospodarczych – ich dynamika była często zbyt wysoka, a szeregi czasowe możliwe do zastosowania zbyt krótkie, by decydenci otrzymywali odpowiednie wsparcie naukowe w tym zakresie. Stąd zainteresowanie rozwojem modeli nie było duże: tematyką tą zajmowali się naukowcy (np. ośrodek łódzki), władze pieniężne i gospodarcze kraju.

Wymagania Komisji Europejskiej sprawiły, że konieczne stało się przeprowadzenie ewaluacji *ex ante* Narodowych Strategicznych Ram Odniesienia 2007–2013 i powiązanych z nimi programów operacyjnych. Wykorzystano w tym celu jeden z najstarszych modeli używanych w Europie dla takich zastosowań – model HERMIN (który został też zaadaptowany na poziomie regionalnym). Oprócz niego zamówiono wyniki uzyskane w sumie za pomocą trzech modeli, każdy różnej klasy (z których jeden dodatkowo modelował regiony).

Dobrze skonstruowany model gospodarki jest w praktyce (wielu, rozwiniętych krajów świata) jedynym wiarygodnym źródłem informacji o możliwych, przyszłych scenariuszach rozwoju. Intuicja czy doświadczenie ekspertów nie może go zastąpić w jakichkolwiek, bardziej zaawansowanych symulacjach oddziaływania interwencji publicznych. Modele są więc bardzo potrzebne dla zwiększenia efektywności wykorzystania posiadanych, ograniczonych środków.

Wydaje się, że Ministerstwo Rozwoju Regionalnego rozumie tę potrzebę i intensywnie podchodzi do kwestii ewaluacji oddziaływania *ex-post* i *ex ante* środków unijnych, czego wyrazem jest m.in. zlecenie przeprowadzenia trzech takich symulacji, za pomocą bardzo różnych modeli i przez różne ośrodki. Zwiększa to naszą wiedzę z zakresu efektywności wykorzystania funduszy unijnych.

Polska uzyskała niepowtarzalną szansę: ogromnej wartości środki pomocowe (na miarę Planu Marshalla) do wykorzystania w ciągu najbliższych kilku lat. Druga taka szansa nie powtórzy się zapewne na przestrzeni następnych, kilkudziesięciu lat. Nie chodzi jednak o to, żeby środki te wydać w 100%, bo to potrafi każdy – ważne jest, by to zrobić jak najefektywniej. Polityka gospodarcza, kształtowanie interwencji państwa, to sztuka dokonywania wyborów, to sztuka podejmowania najwłaściwszych decyzji. Do tego właśnie potrzebne jest naukowe wsparcie procesów ewaluacji, do tego też potrzebne jest modelowanie makroekonomiczne.

BIBLIOGRAFIA

BICEPS (2008), *EU funds macroeconomic impact assessment. First stage report*, Riga, Baltic International Centre for Economic Policy Studies, January.

Botman D., Karam Ph., Laxton D., Rose D. (2007), *DSGE Modelling at the Fund: Applications and Further Developments*, "IMF Working Paper" No. 200, August.

Bradley J., Morgenroth E., Untiedt G. (2003), *Macro-regional evaluation of the Structural Funds using the HERMIN modelling Framework*, "ESRI Working Paper" No. 152, September.

Bradley J., Untiedt G. (2007), *Do economic models tell us anything useful about Cohesion Policy impacts? A comparison of HERMIN, QUEST and ECOMOD*, "GEFRA Working Paper", No. 3, July.

Bradley J., Zaleski J., Tomaszewski P., Zembaty M., Wojtasiak A. (2006), *Ocena wpływu Narodowych Strategicznych Ram Odniesienia i wybranych Programów Operacyjnych na lata 2007–2013 na gospodarki polskich województw przy pomocy modeli regionalnych HERMIN. Raport zbiorczy*, Wrocław, Wrocławska Agencja Rozwoju Regionalnego, grudzień.

Budnik K., Greszta M., Hulej M., Kolasa M., Murawski K., Rot M., Rybaczyk B., Tarnicka M. (2008), *NECMOD: Prezentacja nowego modelu prognostycznego*, Warszawa, Narodowy Bank Polski, 16.06.

Bukowski M. (red.) (2008), *Wprowadzenie euro a bezrobocie i zatrudnienie w Polsce*, Warszawa, Instytut Badań Strukturalnych.

Bukowski M., Pelle D., Saj W. (2008), *Wpływ funduszy unijnych na gospodarkę Polski w latach 2004–2020*, Warszawa, Instytut Badań Strukturalnych, maj.

Bukowski M., Dyrda S., Kowal P. (2008a), *EU-ImpactMod – opis modelu*, Instytut Badań Strukturalnych, lipiec, http://www.funduszestrukturalne.gov.pl/NR/rdonlyres/734EB359–54E4-4242-8C5F-D849AEFE6B51 /47963/31EUImpactMod_opis.pdf

Bukowski M., Zawistowski J. (red.) (2008b), *Zmiana technologiczna na polskim rynku pracy*, Warszawa, Instytut Badań Strukturalnych, czerwiec.

EC (2005), *Country Study: Spain in EMU: a virtuous long-lasting cycle?*, European Commission – Directorate-General for Economic and Financial Affairs, „Occasional Papers" No. 14, February.

EC (2006), *The New Programming Period 2007–2013. Indicative guidelines on evaluation methods: Ex ante evaluation*, European Commission – Directorate General Regional Policy, „Working Document" No. 1, August.

Evaluating Socio Economic Development, SOURCEBOOK 2: Methods & Techniques Econometric models (2003), Regional Policy – Inforegio (Evalsed), http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/downloads/sb2_econometric_models.doc

Fagan B., Henry J., Mestre R. (2001), *An area-wide model (AWM) for the euro area*, European Central Bank – "Working Paper" No. 42, January.

Fic T., Kolasa M., Kot A., Murawski K., Rubaszek M., Tarnicka M. (2005), Model gospodarki polskiej ECMOD, *Narodowy Bank Polski – Materiały i Studia* **194**.

Fuente A. (2003), *The effect of Structural Fund spending on the Spanish regions: an assessment of the 1994–99 Objective 1 CSF*, Fundatión de Estucios de Economia Aplicada "Working Papers" No. 11, April.

Garbicz M., Golachowski E. (2005), *Elementarne modele makroekonomiczne*, Warszawa, Oficyna Wydawnicza SGH.

Główny Urząd Statystyczny (2008), http://www.stat.gov.pl

Kaczor T. (2006), *Model MaMoR2. Informacje o konstrukcji i założeniach*, Warszawa, Instytut Badań nad Gospodarką Rynkową, listopad.

Kaczor T. (2006a), *Prognoza oddziaływania makroekonomicznego realizacji Narodowych Strategicznych Ram Odniesienia 2007–2013*, Warszawa, Instytut Badań nad Gospodarką Rynkową listopad.

Kaczor T., Mitroczuk I. (2007), Goniąc Europę – fundusze publiczne a rozwój Polski i regionów, *Transformacja i rozwój* **113**, Warszawa, Instytut Badań nad Gospodarką Rynkową.

Kaczor T., Socha R. (2008), *Badanie wpływu Narodowego Planu Rozwoju 2004–2006 oraz Narodowych Strategicznych Ram Odniesienia 2007–2013 na wybrane wskaźniki dokumentów strategicznych*, Warszawa, Instytut Badań nad Gospodarką Rynkową.

Karpińska-Mizielińska W., Smuga T., Burzyński W., Ważniewski P., Barteczko K., Duchnowska E., Przystupa J., Marzec A., Marczewski K. (2006), *Ocena szacunkowa Narodowych strategicznych ram odniesienia 2007–2013. Raport końcowy*, Warszawa, Instytut Koniunktur i Cen Handlu Zagranicznego.

Kłos B. (2000), Empiryczny model inflacji, *Bank i Kredyt* **9**, 9–31.

Kłos B., Kokoszczyński R., Łyziak T., Przystupa J., Wróbel E. (2004), Modele strukturalne w prognozowaniu inflacji w Narodowym Banku Polskim, *Narodowy Bank Polski – Materiały i Studia* **180**.

Landreth H., Colander D. (1998), *Historia myśli ekonomicznej*, Warszawa, WN PWN.

MRR (2008), *Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004–2007*, Ministerstwo Rozwoju Regionalnego, Warszawa, lipiec.

Orłowski W. (2000), *Koszty i korzyści z członkostwa w Unii Europejskiej. Metody – modele – szacunki*, Warszawa, CASE.

Pagan A. (2003), *Report on modelling and forecasting at the Bank of England*, London, Bank of England.

Piech K. (2008), *Konwergencja regionów w Polsce – wpływ funduszy unijnych*, w: Pangsy-Kania S., Piech K. (red.), *Innowacyjność w Polsce w ujęciu regionalnym: nowe teorie, rola instytucji, funduszy unijnych i klastrów*, Warszawa, Instytut Wiedzy i Innowacji.

PTE (2007), Polskie Towarzystwo Ekonomiczne, http://www.pte.pl/pliki/2/10/dlapte19XII.pdf

Richardson P. (1998), The structure and simulation properties of OECD's Interlink model, *OECD Economic Studies* **10**, Spring.

Szafrański G. (2006), *Prognoza ekonomicznych efektów realizacji Regionalnego Programu Operacyjnego Województwa Łódzkiego (RPO WŁ) na lata 2007–2013*, ekspertyza wykonana na zlecenie Zarządu Województwa Łódzkiego, Łódź.

Tokarski T. (2005), *Wybrane modele podażowych czynników wzrostu gospodarczego*, Kraków, Wydawnictwo Uniwersytetu Jagiellońskiego.

Treyz F., Treyz G. (2003), *Evaluating the Regional Economic Effects of Structural Funds Programs Using the REMI Policy Insight Model*, referat na V Europejską Konferencję o Ewaluacji Funduszy Strukturalnych pt. *Challenges for evaluation in an enlarged europe*, Budapest, 26–27.06.

Varga J., J. in't Veld (2008), *Macro-Economic Impact Assessment of EU Cohesion Policy with the QUEST Model*, BICEPS conference (Stockholm School of Economics), Riga, 29.05.

Veld J. in't (2007), The Potential Impact of the Fiscal Transfers under the EU Cohesion Policy Programme, *European Economy Economic Papers* **283**.

Veld J. in't (2008), informacje uzyskane przez autora w rozmowie w trakcie konferencji BICEPS (Stockholm School of Economics), Ryga, 30.05.

Welfe W. (2005), *Scenariusze średnio- i długookresowego rozwoju polskiej gospodarki*, referat na konferencję Uniwersytetu Gdańskiego (Katedra Ekonometrii) pt. *Ekonometryczne modelowanie i prognozowanie wzrostu gospodarczego*, Leźno, 20.05.

Zaleski J., Wojtasiak-Terech A., Tomaszewski P., Zembaty M., Bradley J. (2008), *Wpływ realizacji inwestycji finansowanych z funduszy unijnych na kształtowanie się głównych wskaźników dokumentów strategicznych – Narodowego Planu Rozwoju i Narodowej Strategii Spójności oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modelu krajowego i modeli regionalnych HERMIN*, Wrocław, Wrocławska Agencja Rozwoju Regionalnego, 5.05.

Zaleski J., Tomaszewski P., Wojtasiak A., Bradley J. (2004), *Raport 3. Metodologia wykonania średniookresowej prognozy dla polskiej gospodarki*, Wrocław, Wrocławska Agencja Rozwoju Regionalnego, 30.09.

Zaleski J., Tomaszewski P., Zembaty M., Wojtasiak A., Bradley J. (2005), *Raport. Baza danych 16 modeli regionalnych dla polskich województw*, Wrocław, Wrocławska Agencja Rozwoju Regionalnego, 30.11.

ROZDZIAŁ 2.6.

Agnieszka Haber

ZARZĄDZANIE BADANIEM EWALUACYJNYM – PERSPEKTYWA ZLECAJĄCEGO

Zarządzanie projektem ewaluacyjnym to pole działania dwóch głównych podmiotów: zlecającego projekt i jego wykonawcy. I choć w niniejszym rozdziale zajmiemy się perspektywą pierwszego, to warto podkreślić, że niezależnie od perspektywy zaangażowanego podmiotu, **zarządzanie ewaluacją** jest sztuką osiągania celów projektu, wymagającą zdobycia zarówno wiedzy praktycznej, umiejętności interpersonalnych, elastyczności w działaniu, jak również zdolności do przekładania wiedzy teoretycznej na rozwiązania praktyczne, dostosowane do indywidualnej, niepowtarzalnej specyfiki danego projektu, ukierunkowanego na stworzenie unikalnego produktu bądź dostarczenia konkretnej wiedzy.

W niniejszym rozdziale przedstawiony zostanie proces zarządzania ewaluacją z perspektywy zamawiającego. Kolejno omówione będą poszczególne fazy przebiegu projektu:

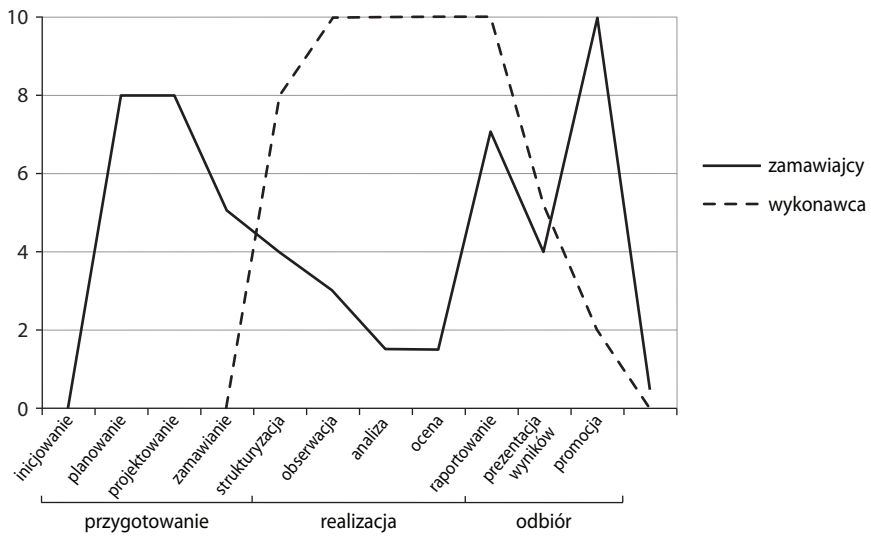
- inicjowanie ewaluacji,
- planowanie – konceptualizacja,
- projektowanie – operacjonalizacja,
- wybór wykonawcy ewaluacji,
- realizacja badania ewaluacyjnego,
- raportowanie (analiza i ocena),
- prezentacja wyników i promocja.

Poruszone zostaną kwestie:

- złożoności otoczenia i relacji występujących w projektach ewaluacyjnych,

- warunków, jakie muszą być spełnione przez zamawiającego w celu otrzymania produktu wysokiej jakości,
- zagadnień problemowych, z jakimi zamawiający może spotkać się, realizując projekty ewaluacyjne.

Zaangażowanie zamawiającego i wykonawcy na poszczególnych etapach jest różne (rysunek 1). Dla zapewnienia wysokiej jakości produktu końcowego udział obydwu stron powinien być w stosunku do siebie komplementarny i uzupełniać się wzajemnie z zachowaniem bezpiecznego dystansu.



Rysunek 1. Zaangażowanie zamawiającego i wykonawcy w trakcie trwania procesu ewaluacji

Źródło: opracowanie własne.

1. INICJOWANIE EWALUACJI

Początkiem cyklu życia projektu ewaluacyjnego jest inicjowanie ewaluacji – **uświadomienie określonej potrzeby realizacji** ewaluacji. Może ona wynikać z narzuconego z góry obowiązku bądź też z rzeczywistych potrzeb informacyjnych, mających wspomóc proces zarządzania interwencją. O ile z pierwszą potrzebą nikt nie dyskutuje, o tyle druga wymaga wielu trudnych zabiegów inicjujących.

Na etapie identyfikacji i definiowania potrzeb informacyjnych oraz określania celów ewaluacji konieczne jest **zaangażowanie zarówno decydentów** odpowiedzialnych za proces zarządzania interwencją, jak i wyższego kierownictwa organizacji zlecającej ewaluację. Jest to istotnie nie tylko z punktu widzenia „uzasadnienia biznesowego", ale

również ostatecznego „sukcesu ewaluacji", czyli wykorzystania jej wyników. Ewaluacja będzie tym, czym uczynią ją szefowie. Ich poglądy i podejście do niej będą kształtowały opinie, postawy i zaangażowanie pozostałych członków organizacji.

Etap inicjowania ewaluacji poprzedza projekt ewaluacyjny. Jego celem jest **określenie korzyści**, jakie przyniesie jego realizacja oraz **zaangażowanie głównych odbiorców wyników** (interesariuszy). Dobrze zaplanowany i koordynowany może stać się siłą napędową projektu ewaluacyjnego. Warunkiem uzyskania akceptacji i zaangażowania kluczowych podmiotów jest doprecyzowanie **celu, kontekstu**, w jakim będzie realizowane badanie oraz **sposobu wykorzystania wyników ewaluacji**, tak by były one zrozumiałe dla wszystkich zainteresowanych. Obok określenia funkcji ewaluacji konieczne jest także uwzględnienie „absorpcyjnej zdolności organizacji" (w zakresie konkretnych ewaluacji). Jak wskazuje Robert Picciotto (2006), zbyt mała porcja ewaluacji może mieć niszczący wpływ na organizację, podobnie jak nadmierne wysiłki ewaluacyjne mogą przyczynić się do niepotrzebnych kosztów transakcyjnych i przeładowania informacjami kadry zarządzającej, do której wyniki są kierowane.

Bez wątpienia zarządzanie tym odcinkiem prac jest łatwiejsze, jeżeli organizacja ma nie tylko doświadczenie oraz kompetencje niezbędne do zarządzania tego typu projektami, ale również wykształciła kapitał społeczny i kulturę organizacyjną zorientowaną na „nowe publiczne zarządzanie" (zob. Olejniczak 2008) oraz, co równie istotne, ukierunkowanie na zarządzanie wiedzą, współpracę, komunikację z otoczeniem (zewnętrznym/ wewnętrznym) oraz odbiorcami (klientami) swoich działań. (zob. Haber, Szałaj 2008).

2. PLANOWANIE – KONCEPTUALIZACJA

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 (Haber 2007).

Planowanie poprzedza realizację badania i obejmuje wszelkie działania przygotowujące i rozpoczynające realizację projektu. Ma nie tylko zapewnić optymalizację zakresu ewaluacji i podejmowanych działań, ale również utwierdzić projektującego, że planowana ewaluacja jest warta ponoszonych kosztów i da się ją zrealizować. Etap ten pomaga zapewnić, by wykonywana praca i wytworzone produkty były odpowiedzią na główne cele ewaluacji oraz spełniły potrzeby informacyjne klienta.

Punktem wyjścia planowania jest decyzja o tym, jakich informacji organizacja lub poszczególni interesariusze potrzebują i dlaczego ich

potrzebują[1]. Zdefiniowanie problemu stanowi uzasadnienie wszystkich następnych działań w zakresie gromadzenia informacji oraz określenia, do kiedy powinny być one dostarczone. Jednymi z najczęściej popełnianych błędów, obok 1) nieumiejętnego przetwarzania dostępnych danych, 2) błędnego lub 3) nieprecyzyjnego określenia potrzeb informacyjnych, są 4) nazbyt szeroki zakres badania, w wyniku którego gromadzone są dane, które tylko w niewielkim stopniu lub w ogóle nie zostaną wykorzystane w analizie problemu badawczego, lub 5) nazbyt wąski sposób patrzenia na dany problem badawczy – skupianie się, w ramach prowadzonego pomiaru wyników, na zjawiskach mierzalnych kosztem elementów trudno lub w ogóle niemierzalnych[2].

Ustalenie kluczowych zagadnień badawczych jest trudne nie tylko ze względu konieczność dokonania selekcji zagadnień, ale również dlatego, że wymaga przebicia się przez retorykę problemu często naznaczoną politycznie, instytucjonalnie, społecznie bądź ideologicznie. Ważne jest, by przy **definiowaniu problemu,** jak wskazuje E. Bardach, bezkrytycznie nie przyjmować owej retoryki, lecz używać jej jako wyjściowego materiału do opracowania wstępnej definicji problemu, która może okazać się analitycznie użyteczna. „(...) Zazwyczaj należy określić wstępną główną orientację analizy, aby nie dopuścić do sytuacji, w której analizowany problem, poprzez swoją wysoką złożoność, trudno będzie uczynić przedmiotem analizy (w przypadku gdy problemy nie są zbyt złożone, można spróbować określić ich więcej niż jeden)" (por. Bardach 2005: 31) lub nie popaść w szereg innych pułapek perspektywy badawczej, np. „pod-optymalizację", którą można zdefiniować jako stawianie przez menedżerów wąskich celów lokalnych, kosztem ogólnych celów organizacji, „obsesję pomiaru" – czyli położenie nacisku na jeden pomiar sukcesu, a nie na leżące u jego podstaw cele (por. Smith 1995: 377–410).

Następnie konieczne jest **określenie przedmiotu ewaluacji i jej wstępnego zakresu**; **możliwości jej realizacji i ewentualnych obszarów ryzyka oraz możliwości wykorzystania wyników** i **organizacyjnych, społecznych**, **politycznych implikacji**[3], (czyli w jakim środowisku i otoczeniu – ze względu na interwencję lub organizację – będzie realizowany projekt). Identyfikacja potrzeby, wartości oraz znaczenia ewaluacji danej interwencji, priorytetów i konkretyzacja koncepcji umożliwiają skuteczną koncentrację zasobów (finansowych, informacyjnych, czasowych, ludzkich) organizacji, niezbędnych do realizacji projektu.

---

[1] Podstawowe pytania, jakie instytucja musi sobie postawić na tym etapie to: „Jakiej/ich informacji poszukujemy? Do czego potrzebne są nam te informacje?" i w końcu „Kim są kluczowi odbiorcy wyników ewaluacji?" oraz „Czy będziemy w stanie wykorzystać wnioski i rekomendacje płynące z ewaluacji?".

[2] Więcej zob. Leeuw 2008.

[3] Więcej zob. Robson 1997.

Dobry plan, wyznaczający ramy badania potrzebne do realizacji założonych celów ewaluacji, powinien być przede wszystkim: **celowy, wykonalny, operatywny, racjonalny, oparty na rzetelnej wiedzy, kompletny**, tzn. musi **obejmować całość zadania i wszystkie istotne czynniki mające na niego wpływ, a także nie powinien generować sprzeczności i dopuszczać zmiany w trakcie realizacji.**

Jak wskazują Nioche, Ghertman i Laroche (1996), planowanie jest angażowaniem się w przyszłość, gdzie z góry trzeba ocenić zbiór problemów i sposobów ich rozwiązywania tak, aby pożądane wyniki osiągnąć w sposób najbardziej efektywny i pewny ujęty w sekwencję kolejnych zadań. I choć nie można wskazać jednolitego sposobu planowania ze względu na szereg odmienności instytucjonalnych, doświadczeń, indywidualnych celów/funkcji ewaluacji oraz sytuacji niepewności i ryzyk projektowych, to możemy bez wątpienia powiedzieć, że sukces projektu na tym etapie zapewnia jasno określony cel przedstawiony interesariuszom możliwie prostym językiem oraz wyznaczenie kompetentnego zespół ludzi po stronie instytucji, który dostarczy informacji z różnych dziedzin.

3. PROJEKTOWANIE – OPERACJONALIZACJA

Planowanie dostarcza podstawy dla projektu (faza operacjonalizacji), na których będzie można zestawić lub przygotować dokument rozpoczynający ewaluację – Specyfikację Istotnych Warunków Zamówienia[4] – przedstawiającą oczekiwania i wizję realizacji zamawiającego ewaluację.

---

[4] Standardowy SIWZ – Specyfikacja Istotnych Warunków Zamówienia (ang. Terms of Reference – TOR) obejmuje:

- podstawę prowadzenia postępowania oraz tryb udzielenia zamówienia;
- zasady postępowania konkursowego;
- cel badania (jasny, ściśle określony oraz użyteczny);
- opis przedmiotu ewaluacji;
- głównych odbiorców wyników ewaluacji i sposoby ich wykorzystania;
- zakres dostępnych danych;
- jasno określone zadania oraz porządek działań;
- hipotezy do zweryfikowania;
- pytania ewaluacyjne;
- zakres badań (przedmiotowy, podmiotowy, czasowy, przestrzenny);
- wymagania dotyczące metodologii lub wskazaną metodologię wraz z metodą doboru próby jej wielkością i charakterystyką;
- szacunkowy kosztorys;
- harmonogram;
- kompetencje poszczególnych stron (wykonawcy/zlecającego).

Etap ten obejmuje dokonanie w kontekście nakreślonego *indicatum*[5], problemów i celów badania, szeregu rozstrzygnięć dotyczących m.in.:

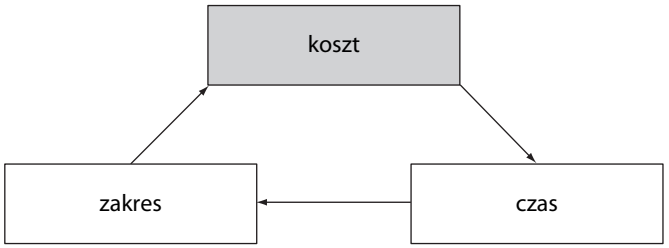
- pojęć i definicji, za pomocą których opisujemy rzeczywistość, a które pośredniczą między rzeczywistym zjawiskiem a badaczem, eliminując niejasności, wieloznaczności;
- wskaźników, mierników będących konsekwencją wstępnego zdefiniowania i konceptualizacji problemu badawczego;
- diagnozy warunków interwencji, problemu;
- badanej zbiorowości (*study population*);
- metod i technik badawczych;
- decyzji co do zasady analizy uzyskanego materiału empirycznego itd. (zob. więcej Babbie 2003; Bardach 2005; Robson 1997; Haber 2007). Należy przy tym uważać, by źle zaprojektowana ewaluacja nie wypaczała obrazu rzeczywistości. Robert Piccotto (2006) wskazuje, że tam, gdzie wynik jest trudny do zmierzenia, istnieje pokusa do używania najłatwiejszych wskaźników, niekoniecznie najbardziej adekwatnych, często mylących, niemających bezpośredniego wpływu na opisywane zjawisko, prowadząc tym samym do niezgodnego z prawdą osądu. Jak pisze, ocenianie programu według z góry ustalonego klucza może przynieść złe skutki: cele mogą być niezwiązane z tematem, są określone nazbyt ambitnie lub przeciwnie skromne. Takie ryzyko istnieje w tych metodach ewaluacyjnych, które są nastawione na osiągnięcie określonego celu. Bez odwołania się do specyfiki programu ewaluacja nie przyczyni się wzrostu odpowiedzialności, a jej wpływ na jakość organizacji będzie minimalny. Wiarygodność i bezstronność ewaluacji zasadzają się na rygorystycznym oszacowaniu celów programu. Nie jest to warunek łatwy do wdrożenia. Istnieje więc pokusa do używania gotowych, zorientowanych na kontrolę zjawiska wskaźników, ze szkodliwymi konsekwencjami dla pobudzenia całej struktury[6].

Uwzględniając powyższe, trzeba jednak podkreślić, że ostateczny zakres projektu warunkowany jest zawsze tzw. magicznym trójkątem, w którego skład wchodzą następujące parametry: **zakres, koszt, czas** (zob. rysunek 2).

---

[5] Zjawisko, którego pomiar nas interesuje.

[6] Doskonałym przykładem tego typu praktyk jest stosowanie w badaniach programów publicznych skierowanych do sektora przedsiębiorstw pomiaru wskaźnika zatrudnienia. Nagminnie dokumenty programowe wskazują na konieczność pomiaru tego wskaźnika w oderwaniu od skomplikowanej rzeczywistości funkcjonowania przedsiębiorstw i rzeczywistego wpływu programu na beneficjenta pomocy.



Rysunek 2. „Magiczny trójkąt" w zarządzaniu projektem

Jednocześnie należy pamiętać, że wszystkie wskazane składowe są od siebie silnie zależne, i tak np. ustalenie zakresu projektu ewaluacyjnego i czasu jego wykonania ustala jego koszt zgodnie z zasadą, że im większy zakres badania, zróżnicowane metody, tym większy budżet, i odwrotnie, zmniejszenie czasu wykonania podnosi koszty lub zawęża zakres badania.

Pozostałe kluczowe kwestie, jakie należy uwzględnić, planując projekt ewaluacji, to niezbędne **zasoby ludzkie** zarówno po stronie zamawiającego badanie (kierownik projektu, skład grupy ds. ewaluacji, eksperci pomocniczy, komitet monitorujący ewaluacją itd.), jak i po stronie wykonawcy (głównie kluczowych ekspertów) oraz **zasoby techniczne** (własne oraz wykonawcy).

Ponadto przy projektowaniu zakresu oraz sposobu realizacji badania należy zwrócić szczególną uwagę na identyfikację **zasobów danych**, ich źródła oraz dostępność, a także **potencjalne kluczowe ryzyka** zidentyfikowane przez zamawiającego, które mogą się pojawić w tracie realizacji ewaluacji. Jedną z procedur poprawnego przygotowania ewaluacji interwencji jest przeprowadzenie zarówno audytu dostępnych danych, jak również wypracowanie pełnego systemu wskaźników, niezbędnego do przeprowadzenia poprawnej oceny wraz z identyfikacją źródeł ich pozyskiwania[7].

Operacjonalizacja problemu badawczego jest najobszerniejszym etapem przygotowania badań. Podjęte na tym etapie decyzje będą rzutować nie tylko na wyniki, ale również na produkt końcowy badania. Dlatego też samo projektowanie budzi wiele kontrowersji. Dla jednych jest ono niezbędne, dla innych stanowi jedynie swoistego rodzaju dekorację lub nawet, w przypadku projektów zbyt szczegółowych, nawet przeszkodę. Wielu teoretyków i praktyków badań ewaluacyjnych uważa, że zlecający projekt powinien jedynie ustalić cel badania i zapewnić środki

[7] Z tego typu projektami można zapoznać się m.in. na stronach Polskiej Agencji Rozwoju Przedsiębiorczości – www.parp.gov.pl/index/index/110 (przygotowanie systemu wskaźników do ewaluacji działań wdrażanych przez PARP w ramach PO KL, PO IG) oraz Ministerstwa Rozwoju Regionalnego – www.mrr.gov.pl

prowadzące do celu, a rolą wykonawcy jest znaleźć optymalny sposób działania oraz optymalne metody badania i oceny danej interwencji. Bez wątpienia takie założenie jest słuszne, jednak doświadczenie wielu instytucji publicznych pokazuje, iż takie rozwiązanie często generuje zdanie się, w warunkach Prawa Zamówień Publicznych, na oferty niskiej jakości, o ograniczonym zakresie. Nazbyt szczegółowe wymagania wobec zakresu i metodologii, niedopuszczające do uzupełnienia bądź modyfikacji zakresu, uniemożliwiają jednak wniesienie wkładu merytorycznego przez wykonawcę, co może powodować zawężenie badania i brak kreatywnych, innowacyjnych rozwiązań metodologicznych, nadmierną automatyzację działań w procesie badawczym, a w najgorszym przypadku nawet „patologie badawcze", prowadzące do nieprawidłowości bądź konstruowania artefaktów. Ze względu na powyższe uwarunkowania, jeżeli zamawiający chce zapewnić sobie wysokiej jakości produkt końcowy, ale nie dysponuje odpowiednimi kwalifikacjami, zasobami ludzkim bądź czasowymi, powinien ten etap zlecić niezależnemu ekspertowi zewnętrznemu bądź bezpośrednio wykonawcy.

4. WYBÓR WYKONAWCY EWALUACJI

Decyzja o podmiocie dokonującym oceny oraz w przypadku zamówień publicznych, trybie wyboru wykonawcy, ma kluczowe znaczenie nie tylko dla jakości produktu końcowego, ale również dla przyszłej współpracy z wykonawcą. Dlatego też w celu zapewnienia sobie:

- „Najlepszego wykonawcy" – niezależnie od trybu wyboru, obok określenia: odpowiednich zasobów ludzkich i technicznych, gwarantujących sprawną, wysokiej jakości realizację projektu; wymogu bezstronności i niezależności wykonawcy w stosunku do zamawiającego oraz przedmiotu oceny; referencji i doświadczenia w realizacji projektów 1) w danym obszarze tematycznym oraz 2) o podobnym charakterze (ze względu na wielkość projektu, zasięg itp.), należy możliwie precyzyjnie określić kluczowy skład zespołu ekspertów i ich kwalifikacje. To właśnie zatrudnieni eksperci będą stanowili o kreatywności, rzetelności, jakości oraz ostatecznie trafności i użyteczności rekomendacji.
- „Najlepszej propozycji technicznej" – należy 1) upublicznić postępowanie w celu zapewnienia możliwie szerokiej konkurencji lub zaprosić kluczowe podmioty specjalizujące się w danym obszarze badań ewaluacyjnych, w celu negocjacji, 2) niezależnie od trybu, zapewnić możliwie dużo czasu na przygotowanie oferty oraz zgromadzenie najlepszego zespołu ekspertów, 3) precyzyjnie określić cel, zakres oraz warunki brzegowe zamówienia, 4) zapewnić wystarczające środki, które pozwalają na realizację wskazanego celu/zakresu ewaluacji,

a także, choć ten pogląd nie jest zbyt popularny wśród wykonawców, na co wskazywałam wcześniej, 5) określić tzw. minimum z możliwością wniesienia do projektu przez wykonawcę istotnych modyfikacji/propozycji merytorycznych.

- „Najkorzystniejszej ceny oraz harmonogramu" – należy przez upublicznienie[8] lub zaproszenie do udziału w postępowaniu[9] zapewnić konkurencyjność ofert. Ze względu na jakość i rzetelność realizacji projektu, należy przy tym pamiętać, by kryterium ceny i harmonogramu nie było decydujące (tzn. w przypadku ceny nie wynosiło 100%[10]). Zamawiający zawsze powinien być zorientowany w aktualnych przedziałach cenowych oraz czasochłonności realizacji poszczególnych etapów prac w projekcie, co zapewni mu nie tylko realistyczne zaprojektowanie własnego budżetu i orientacyjnego terminu zakończenia projektu, ale również ewentualne wychwycenie zaniżania cen lub przedstawiania nierealistycznego wykonania po stronie wykonawcy. Jak pokazuje praktyka, tego typu „konkurencyjne praktyki", zawsze prędzej czy później wpływają na jakość projektu oraz uniemożliwiają ewentualną elastyczność w projekcie ze względu na np. nieprzewidziane wcześniej zadania, bądź udział dodatkowych ekspertów.
- Sprawnej realizacji i przyszłej współpracy – należy jasno i precyzyjnie określić m.in. 1) warunki kontraktu, 2) warunki zabezpieczenia bezpieczeństwa poufności i anonimowości danych przekazanych wykonawcy oraz zgromadzonych przez niego w trakcie procesu badawczego, 3) prawa, obowiązki, kompetencje poszczególnych stron (wykonawcy/zlecającego), 4) zasady przyszłej współpracy oraz, o ile to

---

[8] Przetarg nieograniczony.

[9] Bez zastosowania ustawy, negocjacje z ogłoszeniem, negocjacje bez ogłoszenia.

[10] Trudno rekomendować konkretne ustawienie kryterium ceny, przede wszystkim ze względu na:

- **Indywidualną politykę i wytyczne instytucji (zamawiającego)** w zakresie postępowań na realizację zamówienia publicznego.
- **Szczegółowość zakresu projektu: zakresu prac oraz metodologii badania**. Im projekt bardziej dokładny, tzn. m.in. zawiera wielkość/ci prób/y, techniki badawcze, zakresy analiz itd. i nie zostawia wykonawcy pola do wprowadzenia nowych elementów do projektu, a jedyne elementy, które są „ruchome", dotyczą zwiększenia próby o wartość *x* lub *y*, bądź wprowadzenia kolejnego wywiadu, tym kryterium ceny może być wyższe i *vice versa*.
- **Możliwości finansowe zamawiającego.** Z jednej strony, zamawiający powinien chcieć uzyskać jak najlepszą cenę, z drugiej zaś zamawiający chce otrzymać produkt wysokiej jakości; nie możemy liczyć, iż zostanie on zrealizowany po niżej pewnych kosztów. Jak pokazuje praktyka stosowania wysokiego kryterium ceny, konkurencja na rynku podmiotów ubiegających się o zamówienia publiczne w obszarze realizacji projektów ewaluacyjnych dopuszcza sytuację, w której wygrywają oferty słabsze, mniej kreatywne, z zespołem ekspertów o mniejszym doświadczeniu, oraz realizowane przez firmy dysponujące gorszym zapleczem technicznym.

możliwe, 5) obszary oraz zasady, na jakich mogą odbyć się ewentualne modyfikacje projektu ewaluacyjnego bądź innych warunków kontraktu, takich jak cena lub udział poszczególnych ekspertów w projekcie.

5. REALIZACJA BADANIA EWALUACYJNEGO

Ponieważ proces od zainicjowania projektu do wyboru wykonawcy jest procesem długotrwałym oraz rozwiązywanie problemów bądź poszukiwanie kluczowych informacji mogą mieć charakter niestały, przed rozpoczęciem badania terenowego badacz powinien, wspólnie z zamawiającym, powtórzyć każdy z wcześniejszych etapów i ostatecznie zweryfikować projekt zamawiającego. W procesie strukturyzacji zamawiający, na podstawie analizy i propozycji wykonawcy zawartej w dokumencie zwanym „raportem metodologicznym" powinien, poza dostarczeniem danych i dokumentacji programu, zaakceptować ostateczny kształt badania ewaluacyjnego i tym samym „dać zielone światło" dla dalszych działań wykonawcy. Od tego momentu w procesie realizacji, aż do etapu odbioru pierwszej wersji „raportu końcowego" rola wykonawcy jest drugoplanowa i sprowadza się głównie do kontroli, koordynacji wewnątrz instytucji oraz wsparcia organizacyjnego.

Na tym etapie zamawiający powinien pilnie obserwować i eliminować ewentualne błędy pojawiające się w procesie realizacji – począwszy od kwestii 1) zakresu ewaluacji, by wykonawca nazbyt nie zawężał lub niepotrzebnie rozszerzał jej zakresu, przez 2) poprawność metodologiczną i 3) kontrolę poprawności narzędzi badawczych[11] po 4) gromadzenie danych zarówno w wymiarze poprawności realizacji, jak i zakresu gromadzenia danych.

Etap gromadzenia danych jest jednym z bardziej newralgicznych, narażonych na najwięcej błędów, zarówno po stronie wykonawców jak i zamawiających. Pomijając błędy metodologiczne, szeroko opisywane w literaturze fachowej, do najczęściej spotykanych należy złe rozplanowanie procesu badawczego, który np. konsumuje czas przeznaczony na kolejne etapy oraz tzw. zalewanie danymi (zbyt rozbudowane narzędzia badawcze, gromadzenie danych, które tylko w niewielkim stopniu zostaną uwzględnione w analizach, czy danych kontekstowych, mało przydatnych z punktu widzenia badanego przedmiotu; por. Haber 2007: 53). Zdaniem E. Bardacha, dzieje się tak, ponieważ zajmowanie

[11] Doświadczeni badacze wskazują że:
- „O ile istnieje taka możliwość, czytaj kwestionariusz zarówno jako zamawiający, jak i obierający wyniki badania".
- „Jeżeli masz trudności z wypełnieniem, interpretacją pytań, istnieje duże prawdopodobieństwo, że respondenci również będą je mieli".
- „Źle przygotowane narzędzie zawsze przyniesie błędne lub niepełne wyniki".

Tabela 1. Podstawowe pytania kontrolne przy analizie realizacji badań terenowych

Pomiar
Za pomocą jakich pojęć określono przedmiot badań ewaluacyjnego?
Czy ewaluatorzy nakreślili różne wymiary badanej zmiennej?
Jakie wskaźniki jakościowe i ilościowe zostały wybrane jako miernik tych wymiarów i pojęć? Czy wybrane wskaźniki rzeczywiście mierzą dane zjawisko i są rzetelne?
Jaki jest poziom pomiaru każdej zmiennej: nominalny, porządkowy, interwałowy (przedziałowy), ilorazowy? Czy jest to właściwy poziom pomiaru?
Czy były stosowane jakieś miary złożone (indeksy, skale, typologie)? Jeśli tak, czy były one właściwie ze względu na przedmiot i cel badań oraz czy były poprawne?
Dobór próby
Czy ograniczenie badań do próby jest uzasadnione?
Czy próba dobrana została losowo, celowo, kwotowo czy też była dobierana za pomocą metody „kuli śnieżnej"? Czy było to uzasadnione i poprawnie przeprowadzone?
Na temat jakiej populacji badacz chce wyciągnąć wnioski?
Jeśli zastosowano próbę losową, co było operatem losowania? Czy operat ten odpowiednio reprezentował populację będącą przedmiotem zainteresowania ewaluatora?
Jaki schemat losowania zastosowano? losowanie proste, systematyczne, grupowe? Czy przed losowaniem zastosowano warstwowanie? Czy dokonano go na podstawie właściwych cech, tzn. dla cech istotnych dla zmiennych analizowanych w badaniu?
Jak duża była próba? Jak duży był jej procent realizacji?

W jaki sposób ewaluator sprawdzał reprezentatywność próby (np. czy porównywał rozkład płci, wieku, wykształcenia itd.)?
Czy badane jednostki są reprezentatywne dla szerszej populacji – czy wnioski z próby można odnieść do znaczącej populacji?
Jeśli ze względu na charakter badań, nie miała zastosowania statystycznie reprezentatywna próba losowa, to czy badani/obserwacje zostali dobrani w sposób zapewniający szeroki ogląd analizowanego zjawiska? Czy zwrócono uwagę/wyłączono przypadki nietypowe?
Badania sondażowe (kwestionariuszowe)
Jakie pytania zadawano respondentom?
Jak pytania były dokładnie sformułowane? (błędnie sformułowane pytania generują błędne wyniki badań)
Jak w pytaniach otwartych kategoryzowano odpowiedzi? (błędy kodowania)
Czy pytania są zrozumiałe i jednoznaczne/czy nie zachodzi ryzyko ze respondenci błędnie je odczytają?
Czy respondenci byli w stanie/mogli wiarygodnie odpowiedzieć na stawiane im pytania?
Czy pojawiły się pytania „2w1" (pytania z „i", „lub", „albo") – pytania odnoszące się do dwóch poglądów/sytuacji itd.
Czy pojawiły się pytania zawierające przeczenia – pytania, które mogły być błędnie zinterpretowane i dostarczyć tym samym błędnych odpowiedzi?
Czy ewaluatorzy postawili pytania odnoszące się do norm/wartości społecznych, pytania obciążone tzw. „poprawnością prawno/polityczno/społeczną"?
Czy był przeprowadzony pilotaż narzędzi badawczych? Jakie wyniki przyniósł? Czy na jego podstawie dokonano modyfikacji narzędzi itd.?
Czy i na jakich etapach narzędzie było modyfikowane?

Tabela 1. cd.

Badania terenowe
Czy badania podjęto w celu testowania hipotezy, czy zbudowania teorii na bazie obserwacji?
Jakie są najważniejsze zmienne w badaniach? Jak są definiowane i mierzone? Czy dostrzegasz problemy z trafnością?
Czy badania są rzetelnie skonstruowane/zrealizowane (analizowane)?
Jaką metodę badania wybrano? Czy była ona właściwa? (np. pytanie wymagające zastanowienia lub zawierające dużą liczbę pytań, trwające ok. godziny, realizowane metodą szybkiego badania telefonicznego itd.)
Jaką metodę dotarcia do respondenta wybrano? (badanie CATI – telefoniczne/internetowe/pocztowe/z udziałem ankietera)
Jak dobrano respondentów/próbę?
Czy ewaluator/kluczowi eksperci w jakimkolwiek stopniu uczestniczyli w projektowaniu/realizacji/byli beneficjantami interwencji?
Czy badanym ujawniono/wyjaśniono, w jakim celu przeprowadzana jest ewaluacja? Na czyje zlecenie? Czy to mogło mieć wpływ na udzielane odpowiedzi/zachowania?
Czy można dostrzec u badacza jakieś osobiste uczucia/interesy własne wobec badanych/zleceniodawcy? Jeśli tak, to jaki wpływ mogły one mieć na poczynione obserwacje i wnioski?
Analiza danych zastanych
Kto/kiedy/w jakim celu/w jaki sposób zbierał dane?
Czy dane są kompletne i nie zawierają błędów?
Czy źródło danych jest wiarygodne/rzetelne?

Co było jednostką analizy dla tych danych?
Czy zmienne analizowane/porównywane w obecnych badaniach pokrywają się w definicji/zakresie z danymi zastanymi (np. definicje MSP: wielkość zatrudnienia, przychód netto/brutto, innowacyjność, sposób i okres pomiaru wielkości sprzedaży, wartości zakupów importowych, wartości sprzedaży produktów/usług itp.)
Analiza danych
Czy cel/temat/zakres badania wymagał analizy jakościowej, ilościowej czy mieszanej?
Jaka metoda/rodzaj analizy został wybrany? (jedno/wielo-wymiarowa/analiza SWOT; kosztów/korzyści; kosztów/efektywności; w oparciu o modele ekonometryczne; benchmarking itd.)
Jak zostały zakodowane dane niestandardowe?
Czy ewaluator przedstawił wszystkie istotne dla badań analizy?
Czy wszystkie istotne zmienne zostały określone i zbadane?
Czy konkretne wnioski z analizy są rzeczywiście istotne ze względu na poszukiwane informacje?
Czy ewaluator oparł swoje wnioski i rekomendacje na faktach i wynikach badań? Czy nie ma nadinterpretacji?
Czy analiza i interpretacja danych poprowadzone zostały poprawnie od strony logicznej?
Czy użyte techniki analizy statystycznej były odpowiednie dla poziomu pomiaru analizowanych zmiennych?

Źródło: opracowanie własne na podstawie Babbie 2003.

się gromadzeniem danych wygląda na czynność produktywną, podczas gdy innowacyjne myślenie jest trudne i frustrujące. Ponadto, zleceniodawcy, widząc, że analityk pilnie gromadzi dane, są uspokojeni i przekonani, że mają coś, co jest warte wyłożonych przez nich pieniędzy (Bardach 2005: 39).

6. RAPORTOWANIE (ANALIZA I OCENA)

Na etapie analizy, oceny i pisania raportu, rola zamawiającego powinna być zminimalizowana ze względu na zapewnienie niezależności i bezstronności ewaluatorów. Zaburzenie ról (celowe – w wyniku nacisków ze strony zamawiającego lub zupełnie niezamierzone – np. w wyniku sympatii ewaluatora dla zespołu po stronie zamawiającego ewaluację) w tym stadium, może powodować wiele obciążeń, zniekształceń oceny końcowej oraz konflikty.

Produkt końcowy, jakim jest raport, powinien przede wszystkim dostarczać spójnej diagnozy problemów postawionych przed badaczami, oraz ich rozwiązanie bądź wskazanie kierunków działań, które powinny być podjęte w celu osiągnięcia optymalizacji działań, celów określonej polityki lub interwencji. Zarówno diagnoza stanu, jak i propozycje rozwiązań powinny być zawsze przedstawione na podstawie wyników badania oraz niezależnej analizy i oceny ewaluatora.

W celu wypracowania raportu wysokiej jakości, odpowiadającego wyobrażeniom obu stron, warto, przed rozpoczęciem nad nim prac 1) spotkać się z zespołem ewaluatorów i ustalić jego szczegółowy zakres, strukturę, zawartość oraz styl komunikacji z czytelnikiem oraz 2) przekazać ewaluatorom listę kryteriów, wedug której zamawiający będzie dokonywał oceny raportu.

Raport wysokiej jakości to taki, który z jednej strony: 1) odpowiada na potrzeby informacyjne zamawiającego i potencjalnych grup odbiorców – udziela wyczerpujących odpowiedzi zgodnie z zamówieniem; 2) zawiera interpretację, a nie tylko opis; 3) jasno, logicznie, wyczerpująco, czytelnie, prezentuje wnioski i rekomendacje oraz wskazuje na źródła i rodzaje informacji, jakie posłużyły do wyciągnięcia wniosków wraz z oceną ich znaczenia dowodowego, 4) zawiera odpowiedni dobór przykładów, cytatów (zwłaszcza przy badaniach jakościowych) z drugiej zaś jest 5) dostosowany do czytelnika pod względem języka, treści i zapewnia „komfort użytkowania" przez przejrzystość, estetykę, dobre streszczenia.

W celu wypracowania raportu wysokiej jakości konieczne jest zapewnienie odpowiedniej komunikacji między stronami oraz aktywne włącznie poszczególnych odbiorców ewaluacji w proces odbioru raportu oraz tworzenia rekomendacji. Po wstępnej prezentacji wyników i rekomendacji powinna zawsze odbyć się dyskusja z interesariuszami nad ich trafnością, użytecznością, możliwością wdrożenia, a także

Tabela 2. Przykładowy formularz oceny raportu ewaluacyjnego

Biorąc pod uwagę poniższe kryteria, raport ewaluacyjny jest:	nie do przyjęcia	słaby	dobry	bardzo dobry
Spełnione oczekiwania: Czy ewaluacja rozstrzyga wszystkie zagadnienia ewaluacyjne i czy odpowiada wymaganiom sprecyzowanym w SIWZ?				
Właściwy zakres: Czy trafność programu, jego produkty, rezultaty, oddziaływania, interakcje z innymi politykami i nieoczekiwane efekty zostały objęte analizą?				
Odpowiedni projekt: Czy wybrane podejście badawcze i użyte narzędzia były adekwatne w stosunku do postawionych pytań (uwzględniając istniejące ograniczenia metodologiczne)?				
Poprawność doboru danych: Czy użyte dane pierwotne i wtórne były odpowiednie do przedmiotu analizy? Czy są wystarczająco wiarygodne?				
Solidna analiza: Czy informacje jakościowe i ilościowe były poprawnie analizowane, zgodnie z bieżącą metodologiczną wiedzą i standardami? Czy były kompletne i wystarczające, by odpowiedzieć na zadane pytania?				
Wiarygodne wyniki: Czy wyniki badania są logiczne, wypływają z przeprowadzonej analizy i mają oparcie w zebranych danych oraz ich interpretacji opartej na sformułowanych hipotezach wyjaśniających?				
Ważne konkluzje: Czy raport dostarcza jasnych konkluzji? Czy są one poparte wiarygodnymi dowodami?				
Bezstronne i użyteczne rekomendacje: Czy rekomendacje są uczciwe, niezniekształcone przez osobiste uprzedzenia zlecającego? Czy są wystarczająco szczegółowe, by móc je wdrożyć?				
Przejrzysty raport: Czy raport opisuje kontekst, cele, organizację i rezultaty analizowanego programu w taki sposób, że przedstawiane informacje są przystępne?				
Biorąc pod uwagę ograniczenia kontekstowe badania ewaluacyjnego, ogólnie raport można uznać za:				

Źródło: European Commission 1999, vol. 1, p. 180.

przypisania im konkretnych adresatów odpowiedzialnych za ich wdrożenie. Przypisanie podmiotu odpowiedzialnego za wdrożenie rekomendacji, na tym etapie, znacznie wzmacnia przyszłe wdrożenie, ułatwia określenie horyzontu czasowego, w jakim będzie to możliwe, bądź jest konieczne oraz, co najważniejsze, wzmacnia zaangażowanie interesariuszy w końcowy proces ewaluacji. Przez operacjonalizację formułowanych rekomendacji, budowanie zrozumienia i akceptacji dla zasadności, kształtowane jest nie tylko angażowanie, lecz także świadomość adresatów, która może okazać się kluczowa dla przyszłych zmian w procesie zarządzania interwencją. Zaangażowanie i czas poświęcony na wypracowanie mechanizmów konstruowania i testowania rekomendacji oraz angażowania odbiorców wyników i motywowania adresatów do wdrażania rekomendacji na tym odcinku pracy, zawsze przynosi korzyści zarówno wykonawcy, jak i zlecającemu. Zarządzanie końcowym procesem wymaga nie tylko profesjonalizmu wykonawcy, lecz także wiedzy zamawiającego w zakresie obszarów związanych ze złożoną rzeczywistością interwencji (tj. towarzyszące jej polityki, otoczenie instytucjonalne, prawne oraz społeczne) oraz pewnej dozy wyobraźni strategicznej i defensywnej, koniecznej do spojrzenia okiem oponentów proponowanych zmian czy też polityki.

7. PREZENTACJA I PROMOCJA WYNIKÓW

Ostatnia faza, do niedawna pomijana w wielu opracowaniach i praktykach zarządzania ewaluacją, obejmuje **upublicznienie i promocję wyników**. Niestety zadziwiająco często zdarza się, że etap ten jest nie dość dobrze przemyślany i dopracowany zarówno przez wykonawcę, jak i zamawiającego. Jego realizacja wymaga posiadania odpowiednich umiejętności oraz znajomości odbiorców i najlepszych kanałów komunikacji. A przecież od formy, zakresu, sposobu, dobrania środków komunikacji i dystrybucji wyników oraz znajomości specyfiki odbiorców zależy wykorzystanie wyników ewaluacji.

Raport końcowy jest produktem, który, tak jak inne, wymaga oprawy i promocji. Bez wątpienia napisanie dobrego raportu jest sztuką, ale jeszcze większą jest napisanie go i zaprezentowanie w przystępnej formie oraz dotarcie z wynikami do jak największej grupy odbiorców. Dziś już nie wystarczy organizacja jednego spotkania, na którym wykonawca zaprezentuje najważniejsze wnioski i rekomendacje; umieszczenie raportu na stronach internetowych instytucji, rozesłanie go e-mailem. W erze zalewu informacjami, starając się dotrzeć do odbiorców, musimy nie tylko zadbać o skondensowanie i zaprezentowanie informacji (wniosków, rekomendacji), tak by były one ciekawe, zrozumiałe, przejrzyste i użyteczne dla odbiorcy, ale również dostosować je do możliwości absorpcji wiedzy odbiorcy (por. Haber 2007: 56). Brak należytej dbałości

o jakość i użyteczność ewaluacji może powodować, że politycy, twórcy polityk publicznych i administratorzy publiczni będą skłaniać się ku jej ignorowaniu, zaś brak należytej dbałości o dotarcie do odbiorców, włączenie ich w proces badawczy, a także operowanie zrozumiałym dla nich komunikatem mogą powodować, iż nawet najlepsze ewaluacje zostaną niezauważone nie tylko przez otoczenie zewnętrzne dla danej interwencji, ale również środowisko wewnątrz instytucji zamawiającego.

Podsumowując, można powiedzieć, że jedną z głównych składowych zarządzania ewaluacją jest wiedza o eliminowaniu ryzyka porażki na poszczególnych etapach procesu ewaluacji. Mimo że ze względu na wielowymiarowość i złożoność procesu nie udało mi w powyższym rozdziale wyczerpać wszystkich kwestii związanych z tematem, mam nadzieję, że omówione zagadnienia ułatwią czytelnikowi sprawniejsze poruszanie się w procesie ewaluacji, jak również ułatwią konstruowanie najlepszych rozwiązań dla indywidualnych strategii badawczych.

BIBLIOGRAFIA

Babbie E. (2003), *Badania społeczne w praktyce*, Warszawa, WN PWN.

Bradley K., (2002), *Podstawy metodyki PRINCE 2*, Warszawa, Centrum Rozwiązań Menedżerskich S.A.

Bardach E. (2007), *Praktyczny poradnik do analiz polityk publicznych*, Kraków, Wydawnictwo Małopolskiej Szkoły Administracji Publicznej Uniwersytetu Ekonomicznego w Krakowie.

European Commision (1999), *Means collection: evaluating socio-economic programmes*, DG Regio, Office for Official Publications of the European Communities, Luxembourg.

Griffin R.W. (1996), *Podstawy zarządzania organizacjami*, Warszawa, WN PWN.

Haber A. (2007), *Proces ewaluacji ex-post – konceptualizacja, operacjonalizacja badania ewaluacyjnego*, w: *Ewaluacja ex-post. Teoria i praktyka badawcza*, Haber A. (red.), Warszawa, PARP; http://www.parp.gov.pl/index/more/2046/

Haber A., Szałaj M. (2008), *Kapitał społeczny – warunek i efekt udanych ewaluacji, w: Środowisko i warsztat ewaluacji*, Haber A., Szałaj M. (red.), Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.

Khakee A. (2006), O luce pomiędzy badaniami ewaluacyjnymi a praktyką, *Państwo i Rynek* **2**.

Leeuw F.L. (2008), Ewaluacja w Europie. Wyzwania przyszłości, *Państwo i Rynek* **1**.

Majone G. (2004), *Dowody, argumenty i perswazja w procesie politycznym*, Warszawa, Wydawnictwo Naukowe Scholar.

Nioche J.P., Ghertman M., Laroche H., (1996), *Strategie, Struktury, Decyzje, Tożsamość* w: *Zarządzanie firmą*, Warszawa, PWE.

Olejniczak K. (2008), *Mechanizmy wykorzystania ewaluacji*, Warszawa, Wydawnictwo Naukowe Scholar.

Opałka E., Bienias S. (red.) (2005), *Poradnik realizacji NPR i programów operacyjnych w Polsce*), Warszawa, KJO.

Piccito R. (2006), Ekonomia ewaluacji, *Państwo i Rynek* **2**.

Robson C. (1997), *Projektowanie ewaluacji*, w: *Ewaluacja w edukacji*, L. Korporowicz (red.), Warszawa, Oficyna Naukowa.

Smith P. (1995), On the Unintended Consequences of Public Performance Data in the public Sector, *International Journal of Public Administration* **18**.

Trocki M., Grucza B., Ogonek K., (2003), *Zarządzanie projektami*, Warszawa, PWE.

UKiE (2004), *Podręcznik ewaluacji okresowej programu Phare w Polsce*, Część I *Procedury, metodologia, raporty*, Warszawa.

ROZDZIAŁ 2.7.

Bartosz Ledzion

ZARZĄDZANIE KONTRAKTEM EWALUACYJNYM – PERSPEKTYWA WYKONAWCY

1. WSTĘP

Poniższy rozdział poświęcony jest omówieniu realizacji badań ewaluacyjnych z perspektywy wykonawcy. Prezentacja zaplecza realizacji badań wydaje się konieczna z uwagi na zgoła odmienny charakter niż z punktu widzenia zamawiającego. Różnego podejścia wykonawców należy upatrywać w każdym zakresie, poczynając od wyznaczonych celów badania, po kwestie dotyczące sposobu i zakresu jego realizacji. Bardzo często przyjęcie podejścia jest także uwarunkowane kalkulacjami finansowo-rzeczowymi projektu.

Zarządzanie z perspektywy wykonawców należy określić jako umiejętność zachowania równowagi między czterema krytycznymi aspektami: ograniczeniami operacyjnymi (czasem, budżetem), zakresem badania, podejściem badawczym oraz potrzebami i oczekiwaniami zamawiającego (Preskill 2005). Odpowiednio zaprojektowany system zarządzania badaniem powinien umożliwić wykonawcy:

- przygotowanie koncepcji badania,
- stałe monitorowanie prac oraz zaangażowanie personelem,
- monitorowanie kosztów i harmonogramu,
- bieżące informowanie klientów i interesariuszy o postępie prac badawczych oraz ewentualnych problemach.

Ponadto system zarządzania przyjęty przez wykonawcę musi być na tyle elastyczny, żeby umożliwić szybką reakcję na pojawiające się problemy.

Z jednej strony, rozdział ten jest zaadresowany przede wszystkim do przyszłych lub obecnych wykonawców, gdyż wskazówki w nim zawarte mogą pokazać nowy sposób realizacji projektów lub udoskonalić dotychczasowe podejście. Z drugiej zaś, może okazać się niezwykle cenny dla zamawiających, z uwagi na przedstawienie sposobu myślenia i podejścia wykonawców. Taka wiedza pozwoli zamawiającym na lepsze i bardziej efektywne projektowanie badań oraz koordynację ich działań w przyszłości.

W rozdziale uwaga jest skoncentrowana przede wszystkim na czterech etapach realizacji badania, w których udział wykonawcy jest najistotniejszy, w tym na: procesie przetargowym, uszczegółowieniu metodologii (strukturyzacji), realizacji badania oraz raportowaniu.

2. GŁÓWNE ETAPY ZARZĄDZANIA PROJEKTEM BADAWCZYM

2.1. Wybór wykonawcy ewaluacji – proces przetargowy

Proces wyboru wykonawcy jest istotnym etapem realizacji projektu, przede wszystkim z perspektywy zamawiającego. Zamawiający określa tryb postępowania przetargowego, definiuje kryteria oceny ofert, aby pod koniec dokonać wyboru wykonawcy, którego oferta cechuje się najlepszym stosunkiem jakości do ceny. Aby cel postępowania mógł zostać zrealizowany, zamawiający powinien mieć praktyczną wiedzę o wykonawcach – jakie czynniki motywują ich do uczestnictwa w postępowaniach, jaki mają potencjał i zdolności administracyjne (rzeczywiste, a nie te deklarowane), jakie kryteria są akceptowalne, a które wprost niemożliwe do spełnienia. Zaprezentowana poniżej wiedza pozwoli zamawiającym lepiej poznać perspektywę postrzegania postępowań, trafniej dobierać kryteria selekcji oraz efektywniej oceniać złożone propozycje.

Na wstępie analizowania procesu przetargowego należy zapoznać się z typologią podmiotów mogących uczestniczyć w postępowaniach przetargowych. Obecnie na rynku polskim występują cztery rodzaje podmiotów, które mogą efektywnie uczestniczyć w postępowaniach na projekty badawcze:

Małe firmy doradcze/badawcze

- Bardzo często duża część zespołu pochodzi spoza firmy, większość zadań jest zlecana ekspertom zewnętrznym.
- Jeśli są zaangażowane w wiele kontaktów badawczych w tym samym momencie, rzadko podejmują się kolejnych zleceń z uwagi na ograniczone zasoby techniczne i ludzkie.

- Bardzo często przygotowanie ofert zlecają ekspertom zewnętrznym.
- W pierwszym okresie działalności są bardzo zdeterminowane, aby uzyskać referencje, stąd oferowane przez nich ceny są konkurencyjne.
- Jakość wykonywanych badań w dużym stopniu zależy od klasy ekspertów zewnętrznych, którym zlecają wykonanie prac.
- Bardzo często przystępują do projektów we współpracy z firmami o większym potencjale wykonawczym z uwagi na swoje ograniczone zasoby.
- Wybór tematu badania uzależniają głównie od pól specjalizacji i dostępności ekspertów zewnętrznych.

Duże i średnie firmy doradcze

- Najczęściej są to filie dużych firm zagranicznych, które mają wieloletnie doświadczenie w zakresie badań ewaluacyjnych.
- Posiadają zespoły badaczy i ekspertów zatrudnionych na stałe, jednak coraz częściej dominują w nich także osoby wprost po studiach, z mniejszym doświadczeniem.
- Około 20–50% członków zespołów projektowych pochodzi z zewnątrz. Bardzo często angażują w proces badawczy pracowników naukowych.
- Badania ewaluacyjne stanowią zazwyczaj mniej niż 20% wszystkich zleceń; firmy te specjalizują się głównie w realizacji dużych projektów szkoleniowych, doradczych, finansowanych ze środków UE, ponadto duża część ich dochodów pochodzi z usług świadczonych na rzecz sektora prywatnego.
- Badanie ewaluacyjne realizują dla prestiżu oraz z nadzieją na bardziej rentowne projekty badawcze w przyszłości.
- Firmy zaczynają dzielić rynek, specjalizując się w określonych obszarach (np.: infrastruktura, zasoby ludzkie, innowacyjność, sektor MSP), co daje im większą przewagę konkurencyjną.
- Rzadko przystępują do badań o wartości poniżej 100 tys. PLN, bardzo istotne są dla nich rentowność i zakres projektu oraz zakres, jaki mogą zrealizować własnymi siłami.
- Stosują stałe stawki dzienne wyrażone w roboczodniach, zależne od stanowiska: stażyści, młodsi, starsi analitycy, konsultanci, starsi konsultanci, menedżerowie, starsi menedżerowie, dyrektor, partnerzy.

Ośrodki akademickie/naukowe

- Ośrodki akademickie i naukowe nie mają odpowiedniej wiedzy w zakresie przygotowania ofert pod kątem formalnym, stąd bardzo często odpadają podczas wstępnej weryfikacji.
- Ośrodki naukowe i akademickie bardzo często są zajęte działalnością statutową: wewnętrznymi projektami badawczymi, seminariami, konferencjami naukowymi; projekty komercyjne schodzą na drugi plan.

- Ośrodki przystępują wyłącznie do projektów, które bezpośrednio wynikają z ich specjalizacji bądź z tematów prac doktorskich pracowników naukowych.
- Uczestnictwo w projektach komercyjnych jest postrzegane jako źródło dodatkowych zarobków, poza bieżącym wynagrodzeniem.
- Część ośrodków nie jest w stanie realizować samodzielnie dużych projektów badawczych z uwagi na brak dostatecznej wiedzy w zakresie zarządzania skomplikowanymi projektami doradczo-badawczymi na warunkach komercyjnych. Dość często wchodzą więc w konsorcja z innymi podmiotami – firmami doradczymi.
- Mogą być bardzo konkurencyjne cenowo, gdyż utrzymują się głównie z dotacji uczelnianych, a projekty stanowią dla nich dodatkowe źródło dochodu, stąd rentowność badań nie jest dla nich priorytetowa (w odróżnieniu od pozostałych grup).
- Czasami ośrodki naukowe nie przystępują do projektów z uwagi na zbyt skomplikowane procedury uczestnictwa w takich projektach (np.: wymogi formalne).
- Często podmioty te nie mają odpowiednich referencji formalnych.

Firmy badawcze

- Są często zdeterminowane do wejścia w projekt pod warunkiem dużego zakresu badań ilościowych lub jakościowych.
- Część z nich buduje w swoich strukturach zespoły ewaluacyjne, jednak dominuje wśród nich tendencja do angażowania ewaluatorów zewnętrznych lub wchodzenia w konsorcja z firmami doradczymi.
- Mogą być konkurencyjne cenowo w badaniach o dużym zakresie badań ilościowych i jakościowych.

Na polskim rynku ewaluacyjnym mamy głównie do czynienia z trzema trybami przetargowymi stosowanymi przez zamawiających.

Najprostszym sposobem wyboru wykonawców ewaluacji jest tzw. zamówienie z wolnej ręki. Z perspektywy zamawiającego zamówienie to jest najłatwiejszą formą wyboru realizatorów badań. Zamawiający wysyłają zapytania ofertowe do wielu firm z zakresem i kryteriami wyboru, dając przy tym dwutygodniowy okres na złożenie oferty. Z punktu widzenia wykonawców takie postępowania mają swoje zalety i ograniczenia. Zaczynając od zalet, zazwyczaj takie postępowania nie są obwarowane dużymi kryteriami wejścia (referencje z badań), co umożliwia uczestnictwo wielu mniej doświadczonym firmom. Kolejnym plusem jest brak komplikacji formalnych: bardzo często takie oferty można wysłać tylko e-mailem, bez konieczności dostarczenia oryginałów dokumentów. Badania te nie należą do zbyt skomplikowanych, więc z powodzeniem mogą być wykonywane przez wykonawców o słabszym potencjale administracyjnym. Jednak należy pamiętać, że kwota do 14 tys. EUR może stanowić zachętę głównie dla podmiotów mniejszych.

Postępowanie „z wolnej ręki" ma też swoje wady. Pierwsza z nich polega na tym, że wykonawcy nie mogą monitorować takiego postępowania. Bardzo często dowiadują się, że przegrali, nie wiedząc przy tym, jak została oceniona ich oferta oraz oferty rywali – powoduje to częstokroć frustrację i zniechęcenie. Istnieje też pogląd wśród wykonawców, że rozstrzygnięcie takich postępowań jest bardzo często przesądzone, a brak informacji o wynikach oceny jeszcze bardziej pogłębia takie podejrzenia.

Kolejnym minusem tego trybu jest kwota ok. 50 tys. PLN – dla jednych firm wydaje się duża, jednak dla większości firm jest zbyt mała, aby projekt był rentowny, stąd realizują takie projekty bardziej ze względów promocyjnych, a nie finansowych. Należy jednak pamiętać, że wykonawcy takich projektów ze względów PR mogą zrealizować nie więcej niż jeden, dwa projekty. Ogłaszanie więc takich badań w długiej perspektywie raczej nie wpłynie na wzrost zainteresowania tymi projektami wśród poważniejszych firm badawczych.

Drugim trybem przetargowym jest tryb do 60 tys. EUR. Większość badań ewaluacyjnych jest ogłaszana w ten sposób. Z punktu widzenia wykonawców jedynym minusem tego podejścia jest brak możliwości skutecznego protestowania w sytuacjach, kiedy istnieje poczucie niesprawiedliwej oceny. Wymagane jest także od wykonawców duże doświadczenie badawcze, co stanowi dla części z nich sporą barierę wejścia.

Do pozytywnych stron tego trybu można zaliczyć na pewno wartości projektów: większość ogłaszanych badań ma wartość większą niż 100 tys. PLN. Kolejnym elementem postępowania przemawiającym *in plus* jest upublicznienie oceny, stąd każdy oferent ma dostęp zarówno do wyników oceny, jak i do innych ofert. Takie podejście ogranicza znacznie podejrzenia co do prawidłowości oceny.

Bardzo rzadko stosowaną na polskim rynku procedurą jest tryb powyżej 60 tys. EUR i powyżej 133 tys. EUR. W ten sposób ogłaszane były nieliczne badania. Taka procedura zazwyczaj zakłada bardzo wysoki próg wejścia w formie wymagań referencji kilku badań ewaluacyjnych, wysokiego wadium oraz doświadczonego zespołu badawczego. Takie wymogi dają praktycznie możliwość konkurencji jedynie firmom zagranicznym, które mogą posiłkować się referencjami z badań zagranicznych. Procedura ta stwarza wykonawcom możliwość dość skutecznego procesu protestowania, jakikolwiek błąd formalny, zauważony przez konkurencję, może spowodować wykluczenie z postępowania, opóźnienie w podpisaniu umowy, a nawet porażkę na etapie postępowania przed arbitrami.

Na przełomie 2005 i 2006 roku powszechną praktyką wśród oferentów było wykorzystywanie niedoskonałości Prawa Zamówień Publicznych w celu eliminacji konkurencji z postępowań przetargowych. Obecnie sytuacja uległa diametralnej zmianie na skutek zmiany ustawy.

Bardzo ważnym elementem oceny są wagi nadawane kryteriom wyboru ofert. Opierając się na doświadczeniach ostatnich lat, można wymienić

Tabela 1. Najistotniejsze kryteria badawcze

Kryterium	Zastosowanie praktyczne kryterium	Podejście wykonawców
Zespół badawczy	Jest to wyłącznie kryterium formalne, zazwyczaj zamawiający wymagają kierownika projektu z doświadczeniem, obejmującym co najmniej dwa projekty badawcze oraz ekspertów z doświadczeniem badawczym lub sektorowym.	Wśród wykonawców istnieje tendencja do obiecywania więcej w zakresie zespołu badawczego niż jest wymagane. Jest to celowy zabieg, mający na celu pokazanie, jak mocnym zespołem badawczym dysponują. Rzeczywistość jest odmienna, ponieważ duża część wskazanych ekspertów pracuje po kilka dni roboczych na kontrakcie, który trwa 3–4 miesiące; kontrakt prowadzony jest przez młodszych członków zespołu, starsi zazwyczaj włączają się w badanie na etapie panelów eksperckich, prezentacji wyników badań. Taką sytuację można łatwo zidentyfikować w przypadku ofert, w których ponad 70% członków zespołu stanowią tzw. *senior consultants*; ponadto w części opisującej podział obowiązków są zazwyczaj oddelegowani do nielicznych zadaniach. Aby kontrakt mógł być prawidłowo realizowany, ich obecność jest wymagana na każdym etapie realizacji projektów, co najmniej 2–3 dni w tygodniu.
Referencje wykonawcy	Zgodnie z PZP jest to wyłącznie kryterium formalne, jednakże wysoko postawione warunki stanowią narzędzie wstępnej selekcji wykonawców. Wykonawca gwarantuje, że podmioty uczestniczące w postępowaniu mają doświadczenie (przynajmniej instytucjonalne) w zakresie realizacji projektów badawczych. Należy jednak pamiętać, iż fakt, że firma zrealizowała 10 badań w innych częściach świata nie znaczy, że będzie potrafiła zrealizować poprawnie badanie w kraju realizacji. Badań nie realizują bowiem firmy, a ludzie, którzy są angażowani.	Należy pamiętać, że o ile można w ten sposób ograniczyć liczbę niedoświadczonych wykonawców krajowych, o tyle takie zabiegi nie są skuteczne w przypadku firm zagranicznych, mających swoje filie w Polsce. Zazwyczaj firma zagraniczna, chcąc wejść na rynek, posiłkuje się referencjami filii zagranicznych, wchodząc z nimi w sztucznie tworzone konsorcja. Zdaniem wykonawców, referencje oraz zespół badawczy są kryteriami o dużej wadze psychologicznej, wskazują bowiem na duże doświadczenie badawcze w zakresie danego sektora/obszaru oraz mocny zespół. Wierzą, że przełoży się to nieformalnie na finalną ocenę techniczną oferty.

Kryterium ceny	W większości postępowań waha się między 40–70% wagi całkowitej oceny. Waga tego kryterium jest bardzo często podyktowana systemem zamówień publicznych danej jednostki; niestety bardzo często komórki odpowiedzialne za postępowania nie widzą większej różnicy między badaniem ewaluacyjnym a zamówieniem sprzętu biurowego, co powoduje, że kryterium to odgrywa kluczową rolę przy wyborze wykonawcy. Przy nieprecyzyjnie dopracowanym opisie zamówienia umożliwia to niektórym wykonawcom przedstawianie rażąco niskich cen, co w efekcie powoduje, że oferty o niskiej wartości merytorycznej są wybierane dzięki cenie.	Kryterium ceny jest kryterium koniecznym, ale wymaga rozważnego i proporcjonalnego zastosowania względem kryteriów merytorycznych. Jeśli kryterium będzie wynosiło powyżej 50%, można być pewnym, że jakość ofert będzie średnia, a wykonawcy skupią się wyłącznie na konkurencji cenowej. Korzyścią takiego podejścia jest stosunkowo szybki proces oceny i podpisania umowy z wykonawcą, minusem zaś, że nie zawsze będzie to najlepszy wykonawca, ale na pewno najtańszy, co przy bardziej skomplikowanych badaniach znajdzie swoje odzwierciedlenie w jakości oferty. Natomiast im wyższe kryterium dotyczące metodologii kosztem kryterium ceny, tym na ciekawsze i bardziej innowacyjne oferty można liczyć – zwiększa to prawdopodobieństwo wyboru dobrej jakościowo oferty, tym samym – lepszego końcowego wyniku badania. Zaznaczyć jednak trzeba, że proces oceny może być znacznie dłuższy.
Propozycja metodologiczna	W zakresie większości SIWZ waga tego kryterium waha się między 40%–50%. Oceniane są adekwatność i innowacyjność podejścia badawczego, jakość zaproponowanych narzędzi, wielkość prób badawczych.	Im firma jest bardziej wyspecjalizowana w danej dziedzinie, ma zespół doświadczonych i dostępnych badaczy, tym łatwiej jej przygotować merytorycznie konkurencyjną ofertę w zakresie propozycji technicznej.

kilka najistotniejszych kryteriów formalnych i merytorycznych oceny, przekładających się na wybór oferentów.

Proces przetargowy z perspektywy wykonawcy rozpoczyna się z momentem ogłoszenia przetargu lub, co coraz częściej się zdarza, wraz z publikacją informacji o planowanych badaniach ewaluacyjnych. W momencie, gdy ogłoszony jest przetarg na realizację badania ewaluacyjnego, zespoły w pierwszej fazie podejmują wstępną weryfikację, dokonując przeglądu zakresu i poszczególnych zapisów opisu zamówienia. Można powiedzieć, że dokonują weryfikacji w kilku z góry określonych etapach.

W pierwszej kolejności wykonawcy identyfikują, czy mają wiedzę na temat przedmiotu badania, zarówno w organizacji, jak i poza nią. Zaangażowanie osób spoza firmy mniej się opłaca w przypadku średnich i dużych firm z uwagi na rosnący poziom podwykonawstwa; opłacalne jest to dla firm małych z uwagi na charakter ich działalności, głównie związany z angażowaniem dużej liczby ekspertów zewnętrznych. W części firm poziom podwykonawstwa jest określany na poziomie ok. 20%, to znaczy, że 80% prac projektowych musi wykonywać zespół stałych pracowników, resztę mogą wykonywać eksperci z zewnątrz. Nieco inaczej wygląda sytuacja w ośrodkach akademickich, gdzie głównym czynnikiem przystąpienia do przetargu jest stopień zgodności tematu przetargu z zakresem prac badawczych i statutowych danej instytucji. Bardzo rzadko ośrodki wchodzą w projekty, które nie mieszczą się w głównym nurcie ich naukowej działalności.

W kolejnym etapie wykonawcy precyzują podejście badawcze, określają strategię realizacji badania, dokonują doboru metod i technik badawczych. W praktyce ewaluacji metody są zwykle przypisywane do każdego z pytań ewaluacyjnych, a następnie agregowane w jeden pakiet metodologiczny. Przygotowanie strategii badania jest kluczowym momentem procesu przetargowego, pozwala wygenerować pomysł na realizację projektu, stąd im więcej czasu wykonawca ma na przygotowanie koncepcji badawczej, tym lepszych, bardziej dopracowanych propozycji metodologicznych można się spodziewać. W przypadku wygrania przetargu wstępna koncepcja określona na etapie oferty z reguły jest dopracowywana w raporcie metodologicznym.

W momencie opracowania koncepcji metodologicznej kolejnym krokiem jest opracowanie planu projektu. Przygotowanie planu pozwala zracjonalizować pomysł na projekt, określić, czy realizacja pomysłu będzie rentowna dla firmy. Odpowiednie przygotowanie planu projektu jest ważne z perspektywy realizacji badań, z uwagi na fakt, że:

- dobry plan projektu stanowi stan bazowy przed rozpoczęciem realizacji, powinien ułatwiać zarządzanie, dawać możliwość monitorowania postępu w realizacji projektu pod kątem finansowym i rzeczowym.
- dobry plan pozwala także określić strukturę zespołu badawczego, kolejność i czas podejmowanych działań.

- dobry plan badania stanowi podstawę do tworzenia budżetu badania, co jest kolejnym etapem działań wykonawców.

Sposobów przygotowania planów jest wiele, jednak opracowując plan, musimy pamiętać o trzech ważnych kwestiach, istotnych z punktu widzenia dalszych prac, w tym opracowania budżetu. Po pierwsze, plan musi określać odpowiednią sekwencję prac projektowych, w tym zależność, które etapy są zależne od siebie np.: badanie pilotażowe i badanie właściwe. Po drugie, musimy określić przybliżony czas realizacji poszczególnych zadań, zakładając datę podpisania umowy – tym samym weryfikujemy, czy uda się zrealizować projekt w czasie wyznaczonym przez zamawiającego. Ostatnią ważną kwestią jest to, że plan musi odzwierciedlać podział obowiązków zespołu pracowników oraz poziom ich zaangażowania, co jest szczególnie istotne ze względu na dalsze prace nad budżetem projektu.

Nieskomplikowanym narzędziem, które umożliwia opracowanie przejrzystych i logicznych planów zgodnych z powyższymi wytycznymi jest *one-page-project manager* (Cambell 2007). Za pomocą tego narzędzia można rozplanować w stosunkowo szybki sposób projekt badawczy.

Prawidłowo przygotowany plan badania powinien umożliwiać przede wszystkim weryfikację wykonalności badania w zakładanych przez zamawiającego terminach. Powinien też pomóc w określeniu podziału zadań na poszczególnych członków zespołu. Jest to szczególnie ważne przy budowie budżetu projektu (kolejny etap).

Następnym etapem jest przygotowanie kosztorysu badania, a zwłaszcza próba odpowiedzi na pytanie, czy projekt jest rentowny oraz wart wysiłku związanego z przygotowaniem oferty i realizacji badania? Firmy różnie podchodzą zarówno do zagadnień rentowności, jak i poziomu zaangażowania. Kiedy firmy są zaangażowane w realizację kilku projektów, a na rynku jest ogłoszonych wiele badań, determinacja wykonawców jest mniejsza, stąd oferowane ceny są wyższe, a oferty mogą być dalekie od ideału. Sytuacja może wyglądać podobnie w przypadku badań niszowych, np.: dotyczących sektora infrastruktury – wykonawcy, w obliczu małej konkurencji, będą raczej maksymalizować oferowane kwoty. Bardzo konkurencyjne propozycje cenowe można zauważyć wśród wykonawców chcących wejść na rynek. Wówczas firmy często decydują się na nierentowne projekty, gdyż ich celem jest zdobycie referencji i doświadczenia, a nie zysków.

W niektórych przypadkach, jeśli badanie jest ciekawe, a kwota określona przez zamawiającego jest niewielka, wykonawcy mają tendencję do szukania oszczędności przez upraszczanie struktury badania (redukcje, agregacje lub „prioryteryzacje" pytań badawczych), zmniejszanie próby

badawczej, przechodzenie z kosztownych metod ilościowych na metody jakościowe.

Budowę kosztorysu badania zaczynamy od określenia liczby roboczodni, jaką możemy alokować pomiędzy członków zespołu na realizację poszczególnych prac badawczych. Oczywiście, tworząc budżet, pamiętamy o podziale zadań i ich prioryteryzacji między członków zespołu z przygotowanego wcześniej planu. Im krótszy czas przeznaczony jest na wykonanie prac badawczych, tym więcej dni trzeba alokować na wykonanie danego zadania. Odnośnie do określenia liczby dni, które należy alokować na koordynację badania, warto trzymać się podejścia firm doradczych, które często zakładają 10% liczby wszystkich roboczodni na zarządzanie projektem.

W odniesieniu do kwestii opracowania kosztorysu można zidentyfikować istotny problem związany z określaniem stawek rynkowych przyjmowanych przez wykonawców. Nie da się w prosty sposób odpowiedzieć na to pytanie, zależy to bowiem od wielu czynników. Po pierwsze, od polityki wewnętrznej firm, po drugie, szczególnie w przypadków mniejszych firm, od obciążenia ekspertów dostępnych na rynku. Jeśli realizowanych i ogłoszonych jest wiele badań, to przy ograniczonych aktualnie zasobach ekspertów, stawki za realizację będą rosły w zastraszającym tempie. Wiele firm, żeby zaoszczędzić na kosztach personalnych, a tym samym zwiększyć rentowność przedsięwzięć badawczych, dużą część prac badawczych realizuje, wykorzystując młodszy personel projektowy, ograniczając udział doświadczonych badaczy do 3 kluczowych momentów: opracowania narzędzi badawczych, przygotowania raportu końcowego i udziału w spotkaniach z zamawiającym.

Tabela 2. Zakres stawek dziennych pracowników w firmach doradczych

Stanowisko	Stawka
Stażysta	20–50 EUR
Młodszy analityk	50–150 EUR
Analityk	150–250 EUR
Starszy analityk	250–450 EUR
Konsultant	450–600 EUR
Starszy konsultant	600–1000 EUR
Menedżer	1000–1500 EUR
Starszy menedżer	1500 EUR w górę
Dyrektor	2000 EUR w górę
Partner	3000 EUR w górę

Źródło: opracowanie własne.

Jeżeli chodzi o stawki ekspertów na polskim rynku badawczym, sytuacja wygląda bardzo różnie, gdyż jest uzależniona od polityki wewnętrznej firm. Najbardziej przewidywany i ustandaryzowany sposób wyceny prac ekspertów posiadają firmy doradcze, stąd możemy zaprezentować widełki płac konsultantów (przy czym trzeba zaznaczyć, że 80% prac projektowych jest wykonywanych przez pierwsze cztery grupy pracowników). Należy jednak przy tym zaznaczyć, że stawki się zmieniają wraz ze zmieniającymi się warunkami rynkowymi.

Stawki rynkowe kierowników projektów wahają się od 700 do 1500 PLN dziennie, eksperci sektorowi będą kosztować do nawet 2000 PLN, stawki ewaluatorów i socjologów mogą się wahać od 600 do 1500 PLN dziennie.

Jeśli koncepcja metodologiczna projektu jest ciekawa, plan badania wydaje się realny, a projekt przedstawia perspektywę zysku, wykonawcy przechodzą do kolejnego etapu: przygotowania oferty technicznej – doprecyzowania pomysłu realizacji projektu. W określeniu struktury oferty, wykonawcy kierują się analizą zapisów opisu zamówienia. Jeśli zamówienie jest określone precyzyjnie, z przedstawieniem dokładnego doboru prób oraz metod, wykonawca nie ma zbyt wiele miejsca na własne pomysły, stąd głównym czynnikiem zainteresowania firm staje się zysk. Natomiast, jeśli zamawiający pozostawi nieco miejsca wykonawcy na zastosowanie własnego podejścia, wówczas – poza pieniędzmi – wykonawcy znajdują także inne czynniki motywujące do podjęcia decyzji o badaniu. Ma to miejsce szczególnie w przypadku ośrodków akademickich i eksperckich, dla których pieniądze mają drugorzędne znaczenie.

Do przygotowania ofert firmy zazwyczaj angażują najbardziej doświadczonych pracowników lub współpracowników, zdarzają się także sytuacje, kiedy przygotowanie oferty zlecane jest na zewnątrz. Najczęściej firmy decydują się na takie rozwiązania w przypadku braku wystarczająco kompetentnych pracowników lub kiedy cały zespół jest zaangażowany w inne przedsięwzięcia.

Warto podkreślić, że oferty często są przygotowywane pod dużą presją czasową z uwagi na terminy składania (ostatnio w przypadku trybu do 60 tys. EUR, zamawiający dają nie więcej niż 14 dni, co rzeczywistości oznacza 10 roboczych). Rzadko ofertę przygotowuje się „od zera", często firmy wykorzystują schematy ze starych ofert, stąd zdarzają się sytuacje, że w ofercie można zauważyć pozostałości z poprzednich projektów.

Struktura ofert jest częściowo narzucana przez zamawiających, jednak każdy wykonawca posiada własne podejście budowy ofert. Na podstawie poczynionych obserwacji postępowań przetargowych można powiedzieć, że istnieją dwa sposoby przygotowania ofert na badania ewaluacyjne:

- W pierwszym podejściu wykonawcy dokonują we wstępnych rozdziałach oferty komentarza eksperckiego przedmiotu i zakresu badania,

później w tabelach przedstawiają pytania badawcze oraz propozycję metod ze stosownym uzasadnieniem celowości ich zastosowania.

- W drugim podejściu badacze w części wstępnej przedstawiają opis ekspercki badanego obszaru; później dla każdego pytania przedstawiają szczegółowe podejście metodologiczne i eksperckie, które umożliwi właściwą odpowiedź na postawione pytania badawcze. Taka metoda jest bardziej skuteczna, jednak jest również bardzo pracochłonna. O ile pierwsze lata rozwoju rynku badań zdominował pierwszy opisany wzorzec, o tyle ostatni rok pokazuje dominację ofert o charakterze nie tylko badawczym, ale także eksperckim. Dzieje się tak zgodnie z prognozą Jacka Szlachty (2007), który zauważa, że dobry ewaluator musi posiadać także wiedzę sektorową ocenianego obszaru. Opinia ta coraz częściej znajduje swoje odzwierciedlenie w ofertach poszczególnych wykonawców.

2.2. Strukturyzacja badania

Po ogłoszeniu zwycięskiego wyniku, kluczowym zadaniem wykonawcy jest jak najszybsza organizacja zespołu badawczego, weryfikacja i uszczegółowienie koncepcji badania, ocena dostępności danych oraz identyfikacja szczególnych potrzeb i oczekiwań zamawiającego. W przypadku krótkiego okresu oceny ofert, wykonawcy stosunkowo szybko organizują zespół projektowy. Problemy jednak pojawiają się, kiedy postępowanie przetargowe przedłuża się na skutek protestów czy procedury oceny. Powtórne zorganizowanie zespołu po 3–4 miesiącach od złożenia oferty bywa skomplikowane, bowiem część członków zespołu złożonego z ekspertów zewnętrznych może być zaangażowana w innych projektach, a to wymusza na wykonawcy dokonanie zmian w zespole.

Bardzo często wielu wykonawców po wygraniu projektu zaczyna weryfikację budżetu pod kątem optymalizacji zasobów względem wymogów zamówienia. W tej fazie bardzo często prowadzi się rozmowy z ekspertami o ich zaangażowaniu w projekt, określa się szczegółowo metodologię badania. Przyjęcie właściwego podejścia może pomóc w podniesieniu rentowności projektu nawet o 20% w porównaniu z budżetem ofertowym.

Istotnym elementem etapu strukturyzacji badania jest opracowanie raportu metodologicznego – uszczegółowienie metodologii badania. Podczas tej fazy wykonawcy opracowują narzędzia badawcze (opisane w poprzednich rozdziałach), dokonują rewizji pytań badawczych, czasami prosząc zamawiającego o ich priorytetyzację. Pomaga to wykonawcy skoncentrować się na najważniejszych zagadnieniach badaniach.

Ostatnim etapem strukturyzacji badania jest określenie podejścia do współpracy z zamawiającym. Podczas spotkania wstępnego zamawiający pokazuje nam swoje rzeczywiste, czasem ukryte potrzeby i oczekiwania odnośnie do badania i pozwala nam dostosować do nich tryb pracy.

Osiągniecie wysokiej jakości raportu z badań jest bardziej prawdopodobne w przypadku ścisłej współpracy z zamawiającym. Ponadto wspólna realizacja projektu pomoże szybciej zniwelować pojawiające się trudności, gdyż większości z nich nie można zlikwidować bez zaangażowania się zamawiającego. Zamawiający przez podejście partycypacyjne bierze częściową odpowiedzialność za końcowe wyniki badania, ale z korzyścią dla końcowej jakości procesu (Owen 2007).

W związku z tym, ważnym zadaniem wstępnej fazy badania jest określenie zasad wzajemnej współpracy między zamawiającym z wykonawcą. W pierwszych latach kształtowania się rynku ewaluacyjnego kontakty ewaluatora i zamawiającego ograniczały się wyłącznie do podpisania umowy oraz dwóch spotkań, na których wykonawca prezentował metodologię i wyniki badania. Obecnie coraz powszechniejsze jest podejście partycypacyjne realizacji badań ewaluacyjnych – zespół badawczy stosuje podejście interaktywne w procesie realizacji badań ewaluacyjnych, włączając zamawiającego w proces realizacji badania. Takie podejście gwarantuje przede wszystkim, że efekty procesu badawczego będą zgodne z oczekiwaniami zamawiającego. Jednakże praktyka pokazuje, że zamawiający realizujący dużą liczbę badań, raczej nie są zainteresowani zbyt częstymi kontaktami z ewaluatorami. Wynika to głównie z obciążeń innymi zadaniami.

2.3. Realizacja badania

W trakcie realizacji badania można wydzielić jego trzy zasadnicze fazy: 1) obserwację (zbieranie danych wtórnych i pierwotnych), 2) analizę i ocenę (opracowanie wstępnych wniosków), 3) konsultacje wniosków i opracowanie rekomendacji.

Celem fazy obserwacji jest zebranie danych pierwotnych i wtórnych wymaganych w części analitycznej procesu badawczego. W trakcie realizacji badacze gromadzą informacje o projektach, beneficjentach, procesach aplikacyjnych, danych kontekstowych i statystycznych; ponadto prowadzą szereg badań terenowych o charakterze jakościowym i ilościowym (szerzej o tym w rozdziałach 2.3 i 2.4).

Kolejnym etapem prac badawczych jest faza analizy i oceny. Celem tej fazy jest odpowiedź na postawione pytania badawcze. Uzyskanie odpowiedzi na pytania będzie możliwe przez zastosowanie szeregu metod analitycznych (szerzej o tym w rozdziałach 2.3, 2.4, 2.5 oraz części 6).

Ostatnim etapem realizacji badania są konsultacje wniosków i opracowanie rekomendacji, wieńczące cały proces przygotowaniem raportu. W trakcie oceny trafności wniosków część wykonawców podejmuje

próbę konsultacji wniosków z szerszą grupą ekspertów i interesariuszy, aby końcowa diagnoza nie była dla nikogo zaskoczeniem. Ponadto, żeby uwiarygodnić i umocnić wnioski, wykonawcy wykorzystują recenzje uznanych ekspertów w określonych dziedzinach.

Realizacja badań ewaluacyjnych niesie za sobą szereg wyzwań dla wykonawców. Za kluczowe należy uznać wyzwania o charakterze finansowym, czasowym, politycznym i technicznym (Bamberger 2006).

Ograniczenia budżetowe – istotą planowania ewaluacji jest jej budżet. Często zdarza się, że budżet badania jest zbyt mały do tego, aby przeprowadzić pełnowartościowe badanie, oparte na przeprowadzeniu prac terenowych na dużych próbach badawczych, odtworzeniu danych bazowych oraz analizach porównawczych. Ponadto brak odpowiednich środków bezpośrednio przekłada się także na uwarunkowania czasowe ewaluacji, badacze nie mają odpowiednio dużo czasu na przeprowadzenie badania, co z kolei przekłada się na poziom ich zaangażowania w proces badawczy. Przeciwdziałając ograniczeniom finansowym, wykonawcy dokonują modyfikacji struktury badania ewaluacji przez: koncentrację na najważniejszych pytaniach ewaluacyjnych, rewizje prób badawczych, przejście z metod ilościowych na jakościowe.

Ograniczenia czasowe – są najczęściej spotykanym problemem w realizacji badań ewaluacyjnych. Są one szczególnie odczuwalne w końcowej fazie badania. Odpowiednia ilość czasu jest niezbędna do prawidłowego przeprowadzenia badania ewaluacyjnego; presja czasowa często nie pozwala na przeprowadzenie pilotażu narzędzi badawczych, co przekłada się bezpośrednio na jakość wyników prowadzonych badań (szczególnie ilościowych). Czas potrzebny na konsultacje z partnerami, wizyty, pracę w terenie i analizę danych jest jasno określony terminami przekazywania określonych raportów. Ze szczególnym problemem presji czasowej borykają się eksperci i ewaluatorzy zagraniczni, którym często brakuje czasu na wnikliwe zaznajomienie się z warunkami panującymi w danym kraju.

Trudności w pozyskaniu i przetwarzaniu danych – częstym zjawiskiem spotykanym we wstępnym etapie cyklu realizacji badań jest problem z brakiem lub niewielką ilością dostępnych informacji wyjściowych (np. bazy teleadresowe beneficjentów, dane statystyczne). Nawet jeżeli wymagane dane są dostępne, to najczęściej nie są przygotowane w formie umożliwiającej ich wykorzystanie (dziesiątki zestawień tabelarycznych w różnych formatach). Istniejąca dokumentacja projektowa z systemu monitoringu często zawiera subiektywne, powtarzające się, bezużyteczne informacje. Bardzo duże problemy pojawiają się także przy identyfikacji i zbieraniu danych statystycznych. Zazwyczaj są publikowane przez GUS z dużym opóźnieniem, stąd szybko ulegają dezaktualizacji, a nawet, jeśli istnieją, ich koszt może być wysoki. Nawet jeżeli ilość środków przeznaczonych na badania jest wystarczająca, to często występują problemy z identyfikacją danych umożliwiających analizy porównawcze (szerzej

zob. rozdział 4.4), ponieważ wiele projektów ma swój unikalny charakter; rzadko występują standardy gromadzenia jednolitych danych.

Wyzwania natury politycznej – posługujemy się określeniem wpływu politycznego w odniesieniu do presji ze strony jednostek rządowych lub polityków na wnioski z badań końcowych. Często ewaluacje, które dotyczą programów pomocowych dotykają kontekstu politycznego, gdyż decyzje podejmowane na podstawie badań mogą powodować zmiany w alokacjach funduszy, zmiany w urzędach, nie zawsze po myśli interesów różnych grup politycznych (dobrym przykładem takiej sytuacji były badania typu *ex ante* programów operacyjnych, które przyczyniły się do rewizji listy projektów indykatywnych, co z kolei wywołało „burzę polityczną" wokół decyzji podejmowanych w Ministerstwie Rozwoju Regionalnego).

W przypadku kontrowersyjnych badań raczej nie da się uniknąć problemów natury politycznej. Jedną z metod przeciwdziałania takim praktykom może być włączenie zamawiającego w proces realizacji badania, przez co czyni się go współodpowiedzialnym za wyniki. Dzięki temu łatwiej będzie można bronić końcowych wyników prowadzonych badań. Można także w podejściu powołać się na sprawdzone naukowe paradygmaty, których nikt nie będzie mógł zakwestionować. Jeszcze jednym rozwiązaniem jest powołanie się na autorytet w danej dziedzinie, ale tu trzeba zwrócić szczególną uwagę na apolityczność takiego eksperta.

2.4. Raportowanie

Jedną z ostatnich faz badania jest przygotowanie raportu oraz promocja jego wyników. Po weryfikacji wniosków i rekomendacji należy je zawrzeć w raporcie końcowym z badania. Dobry raport należy napisać w stosunkowo krótkim czasie; powinien być logiczny, czytelny, przekonujący, a przede wszystkim zgodny z oczekiwaniami zamawiającego. Warto przed pisaniem raportu skonsultować z zamawiającym jego strukturę, dowiedzieć się, według jakich kryteriów dokona oceny tekstu. Posiadanie takiej wiedzy pozwoli uniknąć przykrych niespodzianek odrzucenia raportu lub żmudnej jego poprawy pod koniec projektu.

Bardzo interesujące podejście do efektywnego pisania raportów, w tym także badawczych, opracowała firma doradczo-szkoleniowa McLuhan & Davies Communications Inc., opracowując metodologię pisania tekstów pt.: *Writing Dynamics* (2005). Metoda ta umożliwia bardzo szybkie przygotowanie tekstów biznesowych, które charakteryzują się wysokim stopniem czytelności. Metoda opiera się w dużej mierze na mind-mappingu oraz powszechnie stosowanych technikach pisarskich.

Kluczowym zadaniem dobrego raportu jest przedstawienie wyników analizy i wypracowanych zaleceń w przystępnej, jasnej i operacyjnej (dającej wytyczne do zmian i korekt) formie. Aby to osiągnąć, w raporcie końcowym można zastosować szereg technik, które przyczynią się do zapewnienia komunikatywności wyników z badania. Po pierwsze, raport końcowy powinien zawierać krótkie, treściwe (3–5-stronicowe) podsumowanie – powtórzenie w skróconej formie głównych obserwacji, wniosków i rekomendacji. Po drugie, przygotowując raport, nie można zapomnieć o wizualizacji wyników badania na mapach i wykresach. Pozwala to na zobrazowanie najistotniejszych zależności i czynników, pokazanie efektów projektów w sposób całościowy, uwzględniający ich przestrzenny charakter. Po trzecie, tekst powinien zawierać tabele, schematy, zestawienia i wykresy, najciekawsze obserwacje. Zastosowanie tych różnorodnych technik pozwoli na osiągnięcie wysokiego poziomu czytelności raportu.

2.5. Upowszechnianie wyników badania

Można powiedzieć, że zamawiający ma szereg celów związanych z procesem upowszechnienia i wykorzystania wyników badań, począwszy od prezentacji rezultatów programu, wdrożenia działań naprawczych, kończąc na kształtowaniu kultury ewaluacyjnej i uczuleniu ludzi na wyniki badań.

Wykonawcy mają nieco bardziej utylitarne podejście w tym zakresie. Po pierwsze, ich celem jest to, aby ich raport został oceniony jak najlepiej, a wyniki były użyteczne. Pozytywna ocena raportu, artykułowana na różnych konferencjach i seminariach, przyczynia się do zwiększenia prestiżu zarówno firmy, jak i samych ekspertów. To z kolei znajduje swoje odzwierciedlenie w kolejnych postępowaniach przetargowych. Nieco inaczej wygląda sytuacja, kiedy wnioski z badania mają kontrowersyjny charakter. Wykonawcy często obawiają się, że dyskusyjne, „niepoprawne politycznie" wnioski mogą spowodować niechęć klienta, interesariuszy do firmy oraz próbę dyskredytacji przez podważenie kompetencji zespołu bądź trafności postawionych tez. Kiedy wystąpi taka sytuacja, wykonawcom nie zależy na specjalnym rozgłosie związanym z wynikami badania.

BIBLIOGRAFIA

Bamberger M., Rugh J., Mabry L. (2006), *RealWorld evaluation: working under budget, time, data, and political constraints*, Thousand Oaks, Calif., London, Sage Publication Inc.

Cambell C. (2007) *The One-page-project manager*, Hoboken, NY, John Wiley & Sons Inc.

McLuhan & Davies Communications Inc. (2005), *Writing Dynamics*, Toronto, McLuhan & Davies Communications Inc.

Owen J.M. (2007), *Program evaluation: forms and approaches*, 3rd edition, London, The Guilford Press.

Preskill H, Russ-Eft D. (2004), *Building Evaluation Capacity: 72 Activities for Teaching and Training*, Thousand Oaks, Calif., Sage Publication Inc.

Szlachta J., Zaleski J., Korporowicz L., Górniak J., Kowalczyk A. (2007), *Ewaluacja ex ante – podsumowanie doświadczeń administracji polskiej*, Ministerstwo Rozwoju Regionalnego, Warszawa.

CZĘŚĆ TRZECIA

BADANIE EWALUACYJNE – POMOC W PLANOWANIU

ROZDZIAŁ 3.1.

Bartosz Ledzion

WPROWADZENIE

Niniejsza część książki poświęcona jest jednej z trzech prezentowanych w podręczniku funkcji ewaluacji – racjonalizacji planowania. Funkcja ta wiąże się z zagwarantowaniem racjonalnego uzasadnienia dla interwencji publicznych przez zweryfikowanie jego spójności zewnętrznej, przystawalności celów do potrzeb i problemów. Ponadto celem ewaluacji wspomagających proces planowania jest zapewnienie wydajnej, racjonalnej i skutecznej alokacji zasobów przez zweryfikowanie logiki interwencji publicznych oraz racjonalności wyznaczonych celów (Olejniczak, Ferry 2008).

W tej części przedstawiamy cztery studia przypadków badań ewaluacyjnych, których wyniki zostały wykorzystanie w procesie planowania projektów oraz programów finansowanych ze środków Unii Europejskiej.

Pierwszy artykuł stanowi studium przypadku ewaluacji *ex ante* projektu pt. *Wykonanie koncepcji i dokumentacji technicznej budynku pn. Dom Gościnny „Hutnik" w Stalowej Woli*. Został on wybrany ze względu na temat związany z infrastrukturą kulturalną, jednocześnie obejmujący dość specyficzny zakres, jakim jest transgraniczna działalność kulturalna. Zaprezentowane studium stanowi uzupełnienie ewaluacji programów, gdyż pokazuje, że nawet na najniższym poziomie interwencji publicznej (na indywidualnym projekcie) można z powodzeniem stosować metody badań, stosowane dla dużych obszarów interwencji (programów) przy proporcjonalnie mniejszej skali.

Drugi artykuł dotyczy ewaluacji *ex ante* Programu Operacyjnego „Infrastruktura i Środowisko". Opisane studium przypadku przedstawia pogłębioną analizę procesu realizacji jednej z pierwszych ewaluacji typu *ex ante* zrealizowanych w Polsce. W stosunku do pozostałych przykładów ewaluacja POIiŚ stanowi ilustrację oceny programu wielosektorowego o zasięgu ponadregionalnym. Warto przy tym zaznaczyć, że unikalność badanego programu nie wynika jedynie ze złożoności struktury

wewnętrznej działań, ale także faktu największej wartości finansowej w historii Unii Europejskiej.

Przedmiotem trzeciego studium przypadku jest badanie *ex ante* Regionalnego Programu Operacyjnego Województwa Lubuskiego. Głównym powodem przeprowadzenia badania była potrzeba praktycznego przetestowania metod i narzędzi ewaluacji *ex ante*, wypracowanych w ramach projektu badawczego SRDTOOLS. Mimo że dla zespołu realizującego projekt SRDTOOLS nadrzędnym celem była ocena metod i narzędzi, wykonane badanie ewaluacyjne zostało tak zaplanowane i przeprowadzone, aby można je było wykorzystać jako „zwykłą" ewaluację *ex ante* programu.

Ostatnie studium przypadku dotyczy ewaluacji *ex ante* Programu Operacyjnego „Zasoby Ludzkie i Zatrudnienie 2007–2013" finansowanego ze środków Europejskiego Funduszu Społecznego (EFS) w Republice Czeskiej. Powyższy przypadek został wybrany ze względu na interaktywność procesu badawczego oraz znaczące wykorzystanie potencjału, jaki ewaluacja wniosła ze sobą w proces planowania wydatkowania funduszy publicznych.

ROZDZIAŁ 3.2.

Paweł Kościelecki

EWALUACJA *EX ANTE* NA POZIOMIE PROJEKTU. STUDIUM PRZYPADKU PROJEKTU MUZEUM REGIONALNEGO W STALOWEJ WOLI

1. WSTĘP

Niniejszy artykuł stanowi studium przypadku ewaluacji *ex ante* projektu pt. *Wykonanie koncepcji i dokumentacji technicznej budynku pn. Dom Gościnny „Hutnik" w Stalowej Woli*[1]. Został on wybrany nieprzypadkowo. Po pierwsze, jego temat związany jest z infrastrukturą kulturalną i jednocześnie obejmuje dość specyficzny zakres, jakim jest transgraniczna działalność kulturalna. Po drugie, projekt ma charakter *pipe-line*, nie dotyczy zatem samej inwestycji, ale działań poprzedzających proces budowlany (co ma znaczenie przy budowie logiki interwencji). Po trzecie, zakres badania był bardzo ograniczony, co pozwala uznać je za mikroewaluację. Zaprezentowane studium stanowić ma, według mojego zamiaru, uzupełnienie ewaluacji programów lub komponentów programowych.

[1] Niniejszy artykuł bazuje na pracy Buchacz, Kościelecki 2007.

Okazuje się bowiem, iż nawet na najniższym poziomie interwencji publicznej (na indywidualnym projekcie) można z powodzeniem stosować metodologię oraz metody badań, stosowane dla całych dużych obszarów interwencji (programów) przy odpowiednio mniejszej skali.

2. KONTEKST I PRZEDMIOT EWALUACJI

2.1. Opis przedmiotu badania

Poddany ewaluacji projekt stanowić miał kontynuację działań, realizowanych przez Muzeum Regionalne w Stalowej Woli. Od 2002 roku Muzeum wdraża coraz więcej inicjatyw społeczno-kulturowych z partnerami z Ukrainy oraz Słowacji. W związku z powyższym, Muzeum w Stalowej Woli jest aktywnym beneficjentem, zainteresowanym wykorzystywaniem środków UE, dostępnych w ramach Programów współpracy przygranicznej w województwie podkarpackim.

Poza przedsięwzięciami o charakterze miękkim, w zakresie kultury i rozwoju zasobów ludzkich, Muzeum z czasem zaczęło interesować się wdrażaniem projektów infrastrukturalnych. Zainteresowanie wzrasta tym bardziej, że jednym z podstawowych problemów tej instytucji kultury jest bardzo mała baza lokalowa. Muzeum Regionalne dysponuje dwoma obiektami: tzw. Zamkiem Lubomirskich (zbudowanym w latach 1782–1786) przy ul. Sandomierskiej oraz Domem Gościnnym „Hutnik" (w stylu art déco z 1938 roku) przy ul. Wyszyńskiego. Aby można było poszerzyć ofertę kulturalną, w tym podjąć w jeszcze szerszej skali działania transgraniczne, dyrekcja Muzeum podjęła decyzję o realizacji projektu rozbudowy Domu Gościnnego i aplikowanie o współfinansowanie tego zadania ze środków Programu Sąsiedztwa Programu Polska – Białoruś – Ukraina INTERREG IIIA/Tacis CBC 2004–2006.

Program Sąsiedztwa Polska – Białoruś – Ukraina jest jednym z siedmiu programów współpracy transgranicznej realizowanych z udziałem Polski w okresie programowania 2004–20062 .

Program Sąsiedztwa Polska – Białoruś – Ukraina INTERREG IIIA/ Tacis CBC 2004–2006 (nr CCI 2004 CB 16 0 PC 001) został zatwierdzony przez Komisję Europejską w drodze decyzji nr K(2004) 4366, w dniu 5 listopada 2004 roku. Uzupełnienie Programu zostało przyjęte na I posiedzeniu Komitetu Monitorującego, w Lublinie, 26 stycznia 2005 roku oraz nowelizowane przez ten Komitet odpowiednio: 7 października 2005 roku oraz w 2006 roku: 22 maja, 2, 30 czerwca i 20 grudnia.

---

[2] Ze względu na różny okres realizacji projektów, finansowanych ze środków IW INTERREG IIIA i Tacis CBC, program formalnie zostanie zakończony dopiero w 2010 roku.

Tabela 1. Program Sąsiedztwa Polska-Białoruś-Ukraina INTERREG IIIA/Tacis CBC 2004–2006. Priorytety i działania

Priorytety	Działania
Priorytet 1. Wzrost konkurencyjności regionów przygranicznych przez modernizację i rozbudowę infrastruktury transgranicznej	Działanie 1.1. Modernizacja i rozbudowa istniejących systemów transportowych w celu poprawy dostępności regionu
	Działanie 1.2. Rozwój wspólnego transgranicznego systemu ochrony środowiska naturalnego
	Działanie 1.3. Rozwój infrastruktury okołobiznesowej i turystyki
Priorytet 2. Rozwój kapitału ludzkiego i instytucjonalnych form współpracy transgranicznej oraz poprawa bezpieczeństwa na granicach Unii Europejskiej	Działanie 2.1. Wzmocnienie instytucjonalnej współpracy transgranicznej oraz podniesienie jakości kapitału ludzkiego
	Działanie 2.2. Wsparcie inicjatyw społeczności lokalnych (Fundusz Mikro-Projektów)
Priorytet 3. Pomoc techniczna	Działanie 3.1. Zarządzanie, wdrażanie oraz kontrola
	Działanie 3.2. Ewaluacja i promocja programu

Zasadniczym celem Programu jest **podniesienie poziomu życia i społeczno-gospodarczej integracji regionów sąsiadujących**. Do osiągnięcia tego celu służą priorytety i działania przedstawione w tabeli 1.

Składany przez Muzeum Regionalne w Stalowej Woli projekt miał być złożony do działania 1.3 *Rozwój infrastruktury okołobiznesowej i turystyki*.

Celem głównym Priorytetu 1 jest „poprawa konkurencyjności obszaru przygranicznego" (Program 2004), zaś celem głównym działania 1.3 jest z kolei „rozwój sektora turystycznego" (ibidem). „Realizacja działania przyczyni się do wzmocnienia atrakcyjności regionu i będzie stymulowała rozwój społeczny i gospodarczy przy użyciu mocnych stron i korzyści zidentyfikowanych w profilu kulturalnym i turystycznym" (Uzupełnienie 2005).

Badany projekt Muzeum Regionalnego w Stalowej Woli dotyczył dofinansowania ze środków Inicjatywy Wspólnotowej INTERREG IIIA (pochodzących z Europejskiego Funduszu Rozwoju Regionalnego – EFRR) w ramach Programu Sąsiedztwa dokumentacji technicznej rewitalizacji Domu Gościnnego „Hutnik" (obiekt z okresu międzywojennego, stanowiący przed II wojną światową obiekt socjalny zakładów Centralnego Okręgu Przemysłowego). Według beneficjenta:

„Celem projektu jest wzrost konkurencyjności miasta Stalowa Wola poprzez rozbudowę infrastruktury służącej współpracy transgranicznej. Zakłada on stworzenie w budynku Dom Gościnny «Hutnik», należącym do Muzeum, transgranicznego centrum kultury *art déco*. Zabytkowy budynek z elementami stylu *art déco* z okresu 20-lecia międzywojennego, kiedy to powstała Stalowa Wola i wiele ulic we Lwowie, po modernizacji i rozbudowie stanie się miejscem rozwoju polsko-ukraińskich inicjatyw w zakresie kultury i turystyki w oparciu o zasoby dziedzictwa kulturowego. W budynku powstaną nowe sale ekspozycyjne prezentujące sztukę w stylu *art déco*, centrum konferencyjne, polsko-ukraińskie Centrum Informacji Turystycznej o *art déco*".

Na wybór takiego, a nie innego rodzaju przedsięwzięcia, istotny wpływ miała specyfika konkursu (zaproszenie do zgłaszania), ogłoszonego przez Wspólny Sekretariat Techniczny (dalej WST) w dniach 19 marca–18 maja 2007 roku (IV nabór wniosków)[3]. Wnioski o dofinansowanie projektów ze środków EFRR w priorytecie 1 mogły dotyczyć wyłącznie dokumentacji technicznej, analiz, strategii etc. (projekty typu *pipe-line*, wspierające właściwy proces inwestycyjny). Było to podyktowane postępem rzeczowym Programu i stanem wskaźników realizacji projektów.

2.2. Dlaczego podjęto badanie?

Ewaluację projektu rozbudowy i modernizacji Domu Gościnnego „Hutnik" podjęto przede wszystkim po to, by zweryfikować i dokonać korekt w kluczowych jego elementach, takich jak: logika interwencji, wskaźniki interwencji, weryfikacja niezbędnych zasobów finansowych do jego przeprowadzenia, czas niezbędny do realizacji; oraz ocenić, czy rzeczywiście pozwoli on przynieść określone rezultaty twarde (produkty), rezultaty miękkie oraz być użyteczny i wpłynąć pozytywnie na zniwelowanie zidentyfikowanego problemu i jego negatywnych następstw dla otoczenia beneficjenta.

2.3. Kim byli kluczowi interesariusze?

Ewaluacja miała wymiar bardzo praktyczny dla Zamawiającego (Muzeum Regionalnego w Stalowej Woli) przez badanie przystawalności projektu do Programu Sąsiedztwa Polska–Białoruś–Ukraina oraz badanie efektu transgranicznego. Ewaluacja miała podnieść szansę projektu

[3] WST odpowiada za ich rejestrację oraz ocenę formalną i techniczną, stworzenie listy projektów rekomendowanych do wyboru przez Komitet Sterujący i dalej przygotowania ich do podpisania umów o dofinansowanie.

na uzyskanie dofinansowania ze wspomnianego programu. Kluczowym interesariuszem tego badania był zatem potencjalny beneficjent. Jak zostanie to wykazane poniżej, badanie miało również innego interesariusza (dyskretnego, pośredniego), czyli WST.

2.4. Jaki to był typ ewaluacji?

Ewaluacja ta miała charakter *ex ante*. Definicje ewaluacji zostały omówione w innych rozdziałach książki, stąd nie będę tym zagadnieniem zbyt obszernie się zajmował (por. także Olejniczak 2007; 2008, tam obszerna literatura przedmiotu). Istotne było to, że, zgodnie z dostępnymi dokumentami metodologicznymi, celem ewaluacji *ex ante* jest poprawienie i wzmocnienie ostatecznej jakości przygotowywanego planu lub programu[4]. Poddawano ocenie przedsięwzięcie, które dopiero starało się o dofinansowanie.

3. WKŁAD

3.1. Jakie były potrzeby informacyjne i jakie pytania postawił Zamawiający?

Cel badania dla Zamawiającego był bardzo jasny. Chciał zweryfikować wyłącznie przygotowany projekt (konkretnie przygotowany wniosek aplikacyjny). W trakcie kontaktu z Zamawiającym, badacz przedstawił szereg argumentów merytorycznych, które pozwoliły rozszerzyć zakres prac. W konsekwencji zamiast standardowego sprawdzenia wniosku, udało się przeprowadzić badanie projektu jako pewnej zapisanej idei (koncepcji) działań, która ma przynieść określone rezultaty twarde (produkty), rezultaty miękkie oraz być użyteczna i wpłynąć pozytywnie na zniwelowanie zidentyfikowanego problemu i jego negatywnych następstw dla otoczenia.

3.2. Jakie postawiono warunki operacyjne? (budżet, czas badania, zespół, dostępność danych)

Jedynym wymogiem przeprowadzenia tej ewaluacji było uzyskanie konkretnych produktów przed 18 maja 2007 roku, tak aby wyniki

[4] Komisja Europejska stawia przed ewaluacją *ex ante* szereg konkretnych celów. Są to: dokonanie oceny, czy strategia działania lub program jest właściwym narzędziem do sprostania wyzwaniom, przed którymi stają państwo, region czy sektor; czy strategia lub program dobrze określiły osie strategiczne, priorytety i cele, a także czy są one właściwe i czy w danej sytuacji mogą zostać osiągnięte; przyczynienie się do lepszej kwantyfikacji celów (European Commission 2000a; 2000b; 2006).

pracy można było wykorzystać do przeprowadzania korekt we wniosku aplikacyjnym, załącznikach przed złożeniem dokumentacji do WST. Koszt projektu znalazł się poniżej progów przetargowych, a praca była wykonywana przez jedną osobę. W tym przypadku możemy zatem mówić o mikro-ewaluacji. Ponadto badanie było zamówione u osoby, która wówczas nie pracowała już w strukturach instytucji, zaangażowanej w realizację Programu, a która miała jeszcze aktualną wiedzę na temat warunków formalnych i merytorycznych.

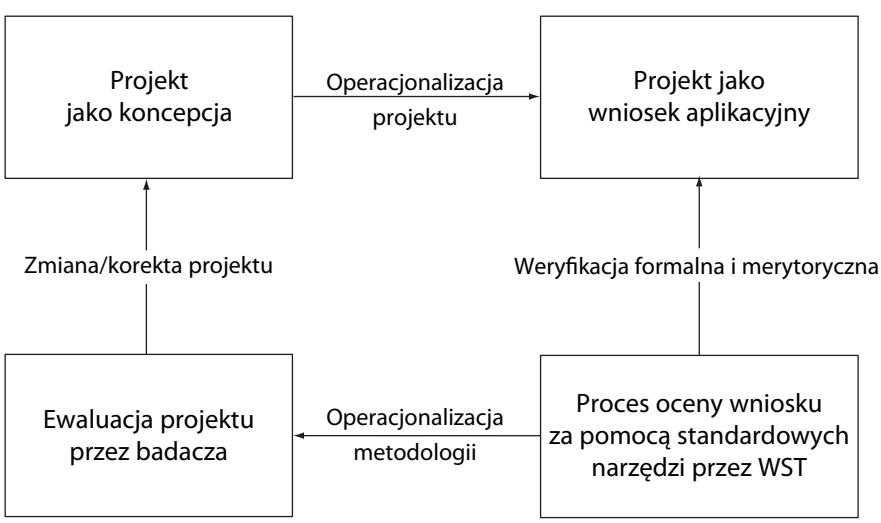
Zamawiający dostarczył swój projekt wniosku aplikacyjnego i zadeklarował pełną współpracę przy działaniach ewaluacyjnych. Treści zawarte w tym wniosku były podstawą do określenia przedmiotu projektu (rozbudowa i modernizacja obiektu). Zamawiający nie precyzował żadnych metod badawczych, nie ingerował w sposób prowadzenia prac.

4. PROCES BADAWCZY

4.1. Jakie podejście badawcze i metodologię przyjął ewaluator?

Cały proces badawczy został podporządkowany praktycznemu aspektowi ewaluacji (rysunek 1).

Jak wspomniano, ewaluacja miała na celu weryfikację projektu i w podstawowym założeniu miała po prostu poprawić jakość projektu transgranicznego. Ta poprawiona koncepcja (opis projektu) stała się



Rysunek 1. Ewaluacja projektu i ewaluacja wniosku. Wzajemne relacje

podstawą do sporządzenia wniosku (co nazwać można operacjonalizacją projektu). Wniosek aplikacyjny był następnie poddany weryfikacji formalnej i merytrycznej przez WST. Projekt był zatem oceniany dwukrotnie. W pierwszym przypadku była możliwość zmian w samej koncepcji działań lub samej narracji, w którym te zamierzenia opisywano. W merytorycznej ocenie wniosku (jako zamkniętej, skodyfikowanej przez szablon narracji) takiej możliwości już nie było.

Wybór kryteriów ewaluacyjnych był podyktowany z jednej strony zakresem badań *ex ante*, z drugiej zaś musiał uwzględniać kluczowe kwestie projektu, będące obiektem zainteresowania WST w trakcie oceny technicznej (merytorycznej) aplikacji. Można to nazwać operacjonalizacją metodologii. Zaproponowano następujący model oceny *ex ante* projektu.

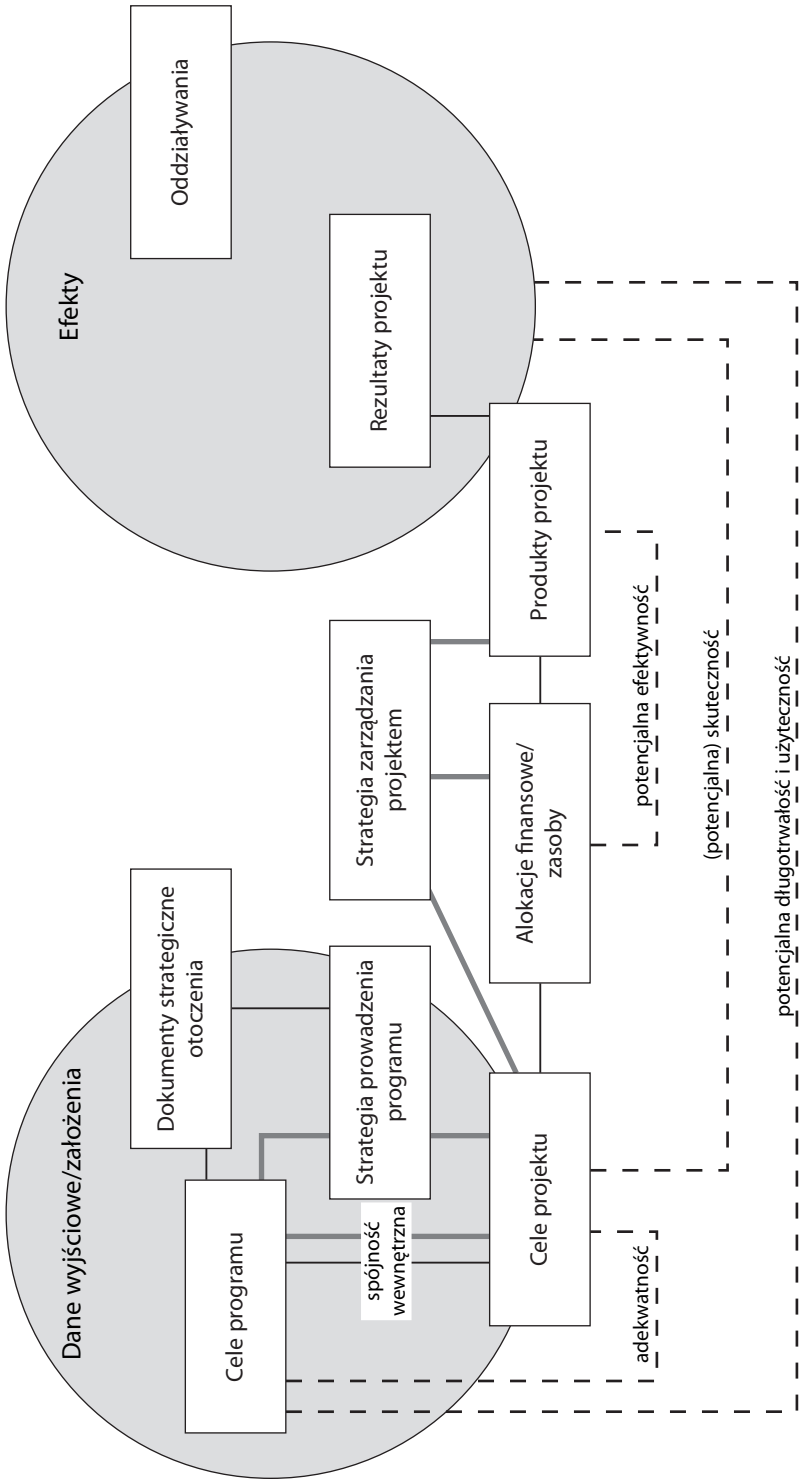
Kryterium odpowiedniości (adekwatności) jest równoznaczne ze zgodnością celów projektów z danymi wyjściowymi (założeniami). Drugim istotnym elementem danych wyjściowych jest kontekst pozaprogramowy, wyrażany w dokumentach strategicznych na poziomie UE (polityki horyzontalne), kraju i regionu (kryterium to jest tożsame z badaniem znaczenia projektu w trakcie oceny technicznej aplikacji). Jednym z pobocznych wątków badania adekwatności projektu jest badanie jego spójności wewnętrznej, czyli zgodności celów projektu z wyborami strategicznymi na poziomie projektu (wyborem działań, określeniem grup docelowych, parterów etc.).

Badaniu podlegać może również kryterium efektywności. Oznacza to analizę zasadności wydatków, wysokości i umiejscowienia alokacji finansowych oraz innych zasobów, niezbędnych do przeprowadzenia projektu, w tym także liczby personelu, harmonogramu działań etc.

W badaniu zwrócono również uwagę na jakość i logikę wewnętrzną interwencji, czyli powiązanie produktów, rezultatów i oddziaływań. Wraz z tym badaniem przeprowadza się analizę odpowiedniości i wartości wskaźników. W konsekwencji można odpowiedzieć, czy projekt ma szansę być skuteczny (WST, badając wykonalność projektu, przeprowadza ocenę efektywności i właśnie skuteczności projektu).

Badanie potencjalnej długotrwałości polegało na powiązaniu efektów projektu z możliwościami beneficjentów w ich utrzymaniu oraz kosztami tego utrzymania. Należy podkreślić, iż nie jest to tożsame z ewaluacyjnym kryterium (potencjalnej) trwałości. W stosunku do niego zadaje się przede wszystkim pytania o trwałość samych efektów, mniej zaś o wsparcie instytucjonalne i administracyjne w zarządzaniu nimi.

Badaniu poddano także użyteczność projektu, która pozostaje poza analizą techniczną WST. W kontekście ewaluacji projektu, jest to ocena, czy efekty projektu odpowiadają na potrzeby, czyli w konsekwencji z celami programu (rysunek 2).



Rysunek 2. Kryteria ewaluacji *ex ante* projektu

4.2. Jak przebiegał proces badawczy?

Na podstawie tak skonstruowanych kryteriów sporządzono pytania badawcze (tabela 2) oraz przygotowano narzędzia badawcze.

Do przeprowadzenia analiz dokumentów wykorzystano Program Sąsiedztwa, Uzupełnienie Programu, wytyczne dla beneficjentów, a także dotychczasowe wnioski aplikacyjne, sprawozdania z realizacji projektów, informacje o przedsięwzięciach na stronie internetowej Muzeum. Ponadto zostały udostępnione dokumenty dotyczące bilansu, rachunku zysków i strat beneficjenta oraz wstępne założenia do Domu Gościnnego „Hutnik", w którym Muzeum opisało zamierzenia w zakresie programu kulturalnego i społecznego w perspektywie najbliższych kilku lat.

Badanie rozpoczęto od weryfikacji logiki interwencji. Po zidentyfikowaniu słów kluczowych i zdań bazowych (typu: „celem projektu jest…", „projekt ma przyczynić się do…", „mamy zamiar…"), zidentyfikowano ich wzajemną pierwotną zależność i uporządkowano przy pomocy drzewa logicznego (rysunek 3). Wymagała ona znacznych korekt i poprawy przede wszystkim w ciągach przyczynowo-skutkowych. Efektem tego była nowa logika interwencji w projekcie (rysunek 4).

Warto przy tym zwrócić uwagę, że w przypadku projektów typu *pipe-line* zasadniczym produktem jest dokumentacja przedinwestycyjna. Zatem rezultatem (sposobem wykorzystania produktu) w projekcie byłaby właściwa infrastruktura (rysunek 4). Tym samym, cała logika interwencji takiego projektu jest „przesunięta w dół" względem typowego projektu infrastrukturalnego. W przypadku tego ostatniego produktem jest infrastruktura zaś rezultatem sposób jej wykorzystania.

Zastosowana macierz przystawalności (*Assessment Matrix*) służyła do oceny spójności zewnętrznej dokumentu, a więc ocenie powiązania pomiędzy celami interwencji na poziomie Programu Sąsiedztwa a celami projektu. Cele porównywanych programów/dokumentów z projektem wpisano w matryce (cele programu w poziomie, cele projektu w pionie) i analizując treść zapisów dokumentów, zaznaczono na matrycy pola rozbieżności/zbieżności/konfliktów (tabela 3).

Po zakończeniu tych analiz przeprowadzono badania jakościowe. Polegały one na przeprowadzeniu długiego wywiadu swobodnego, w trakcie którego ustalono jego wizję projektu, zidentyfikowano jego rzeczywiste intencje, zweryfikowano cele. Przy okazji, w trakcie wizyty w Muzeum dokonano oglądu obiektu i wspólnie z przedstawicielami placówki rozważano roboczo inne warianty rozbudowy dostępnej infrastruktury lub alternatywnego wykorzystania dostępnej powierzchni w obiektach Muzeum. Wynik tej wizyty potwierdzał, że wariant rozbudowy projektu jest w chwili badania najlepszy. Badanie w terenie pozwoliło na zweryfikowanie, czy problem został prawidłowo zdefiniowany, a zatem jakie są bezpośrednie i pośrednie cele projektu. Rozmowy pozwoliły także na stwierdzenie, czy w koncepcji projektu został w sposób wyczerpujący omówiony

Tabela 2. Zestawienie pytań badawczych dla projektu *Wykonanie koncepcji i dokumentacji technicznej budynku pn. Dom Gościnny „Hutnik" w Stalowej Woli*

Kryterium ewaluacyjne	Zagadnienie	Pytania badawcze	Zastosowane narzędzia badawcze
Trafność	Zagadnienie 1: Adekwatność celu projektu do celu działania 1.3 Programu Sąsiedztwa	Pytanie 1: Jakie są cele programu/ priorytetu działania?	Analiza dokumentów
		Pytanie 2: Jak zostały zdefiniowane potrzeby beneficjenta w zakresie współpracy przygranicznej?	Badania jakościowe: wywiad swobodny z przedstawicielami beneficjenta, wizja lokalna Analiza projektu
		Pytanie 3: Czy cele programu/priorytetu działania są zgodne z celami projektu?	Matryce przystawalności
	Zagadnienie 2: Adekwatność celu projektu do celów dokumentów strategicznych	Pytanie 1: Jakie są cele strategii/priorytetu działania?	Analiza dokumentów
		Pytanie 2: Czy cele strategii działania są zgodne z celami projektu?	Matryce przystawalności
	Zagadnienie 3: Adekwatność projektu z politykami UE	Pytanie 1: Czy projekt jest zgodny z kwestiami ochrony środowiska?	Analiza projektu
		Pytanie 2: Czy projekt jest zgodny z kwestiami równych szans?	Analiza projektu
		Pytanie 3: Czy projekt jest zgodny z kwestiami polityki ICT?	Analiza projektu

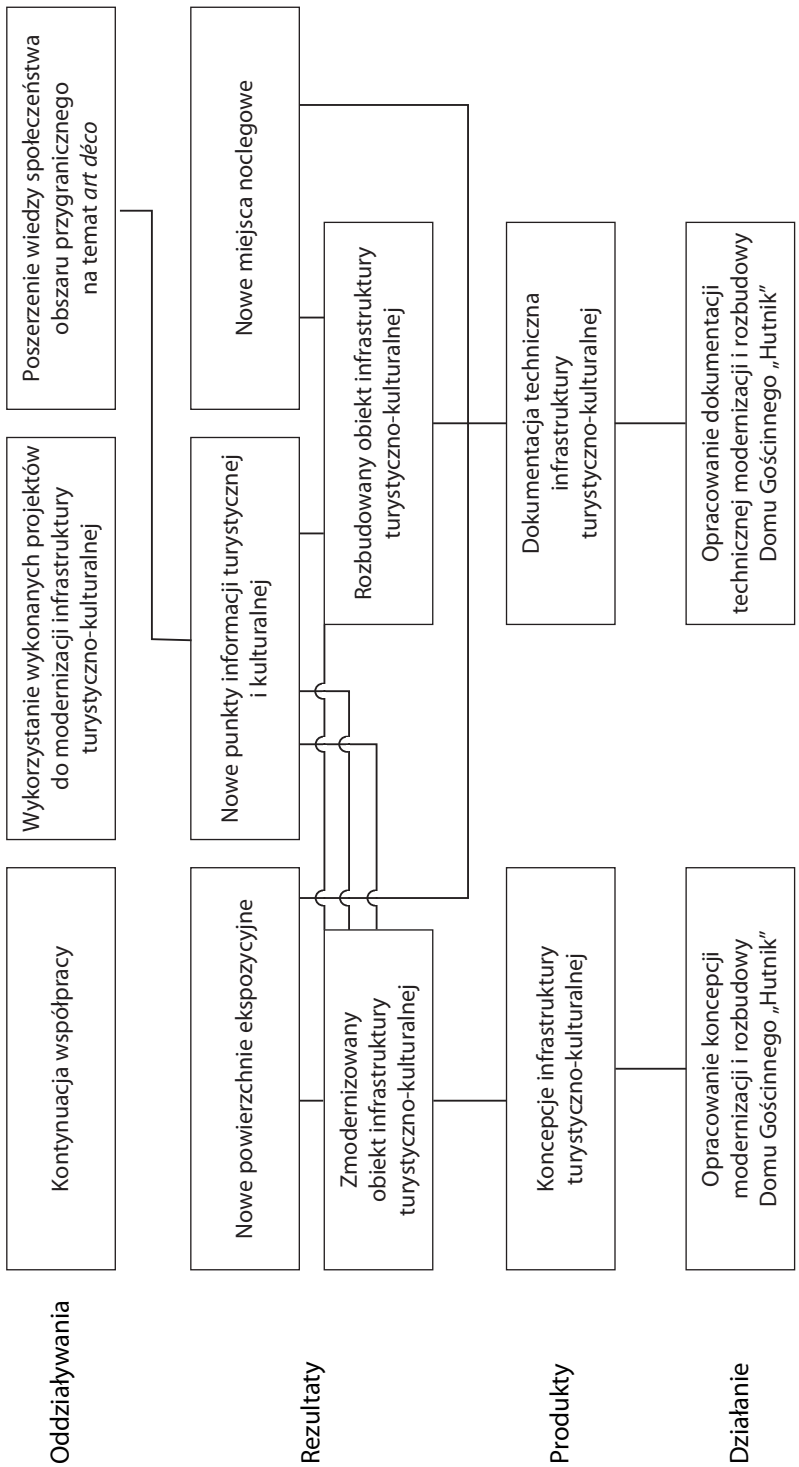
Trafność (cd.)	Zagadnienie 4: Adekwatność problemów grupy docelowej z celami projektów	Pytanie 1: Jakie są rzeczywiste problemy beneficjentów/ grupy docelowej?	Analiza dokumentów
		Pytanie 2: Czy cele działania wobec grupy docelowej są zgodne z celami projektu?	Analiza projektu
	Zagadnienie 5: Powiązanie projektu z innymi projektami, współfinansowanymi ze środków UE	Pytanie 1: Czy istnieje powiązanie badanego projektu z innymi projektami, współfinansowanymi ze środków UE?	Analiza dokumentów, wywiad swobodny z przedstawicielami beneficjenta
Skuteczność	Zagadnienie 1: Spójność logiki interwencji	Pytanie 1: Czy produkty, rezultaty i działania są względem siebie adekwatne?	Drzewo logiczne
	Zagadnienie 2: Jakość systemu wskaźników na poziomie projektu	Pytanie 1: Jak wygląda system wskaźników na poziomie projektu ?	Analiza projektu
		Pytanie 2: Czy wskaźniki są odpowiednio dobrane do produktów, rezultatów i oddziaływań projektu?	Analiza SMART, analiza matrycy logicznej
		Pytanie 3: Czy system wskaźników na poziomie projektu da się agregować na wyższy poziom?	Analiza dokumentów
	Zagadnienie 3: Oczekiwane osiągnięcie efektów projektu	Pytanie 1: Czy wskaźniki produktu są możliwe do osiągnięcia?	Weryfikacja w trakcie kontaktu z beneficjentem, analiza dokumentów
		Pytanie 2: Czy wskaźniki rezultatu są możliwe do osiągnięcia?	Weryfikacja w trakcie kontaktu z beneficjentem, analiza dokumentów

Tabela 2. cd.

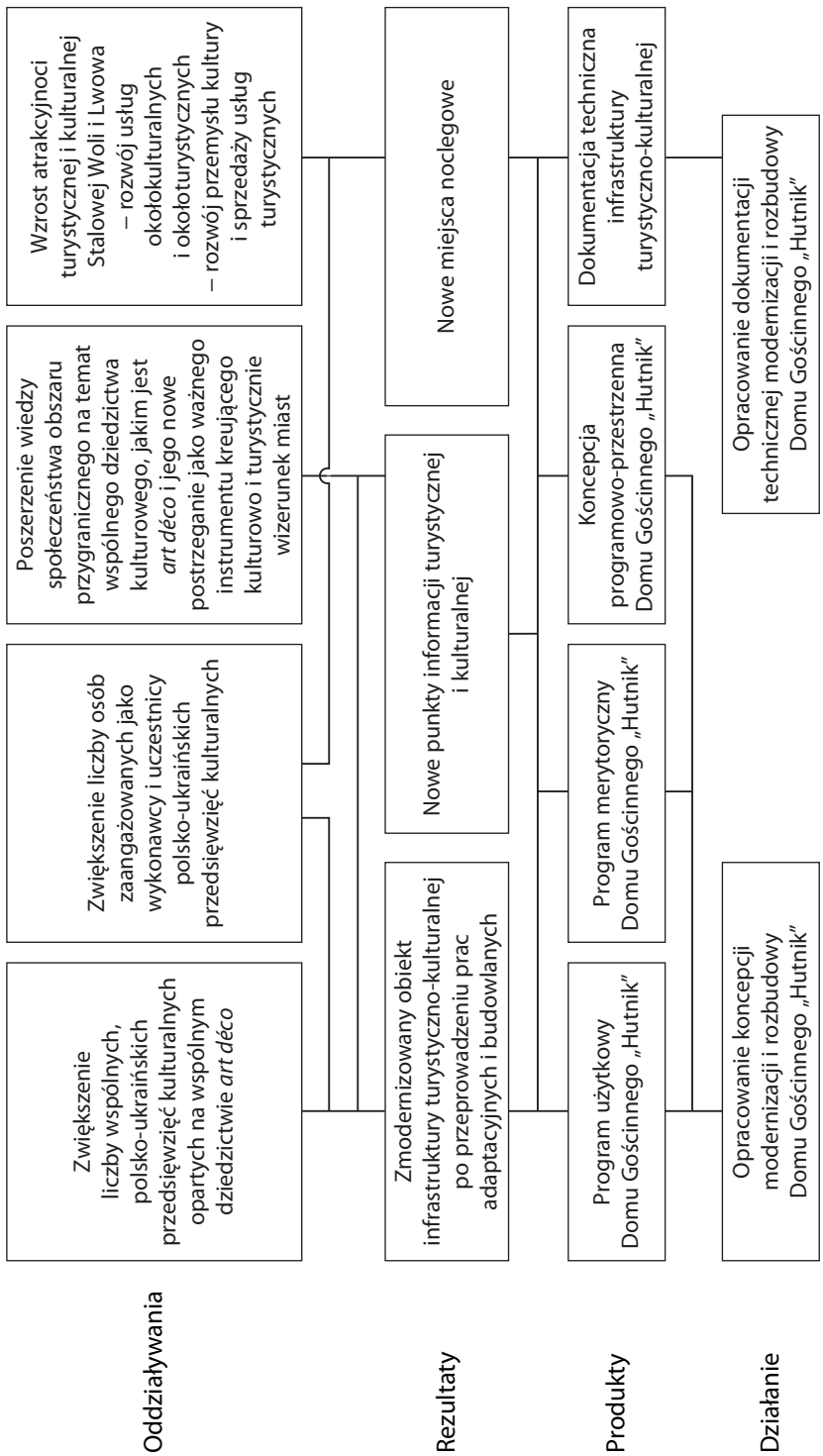
Kryterium ewaluacyjne	Zagadnienie	Pytania badawcze	Zastosowane narzędzia badawcze
Skuteczność (cd.)	Zagadnienie 4: Kwestia budowy polsko-ukraińskiej sieci instytucjonalnej	Pytanie 1: Jak projekt będzie wpływał na rozwój współpracy z partnerami ukraińskimi	Weryfikacja w trakcie kontaktu z beneficjentem, analiza dokumentów
Efektywność	Zagadnienie 1: Efektywność finansowa	Pytanie 1: Czy projekt zawiera wszystkie niezbędne wydatki ?	Analiza projektu
		Pytanie 2: Czy projekt zawiera wydatki zbędne?	Analiza projektu
	Zagadnienie 2: Efektywność instytucjonalna	Pytanie 1: Jakie jest doświadczenie partnerów polskich?	Weryfikacja w trakcie kontaktu z beneficjentem, analiza dokumentów
		Pytanie 2: Jakie jest doświadczenie partnerów zagranicznych?	Weryfikacja w trakcie kontaktu z beneficjentem, analiza dokumentów
Trwałość	Zagadnienie 1: Szansa na trwałość struktur współpracy	Pytanie 1: Jaką formę ma mieć przyszła współpraca?	Badania jakościowe: wywiad swobodny z przedstawicielami beneficjenta
		Pytanie 2: Czy beneficjent planują kontynuację współpracy z partnerami ukraińskimi?	Badania jakościowe: wywiad swobodny z przedstawicielami beneficjenta
	Zagadnienie 2: Trwałość instytucjonalna i finansowa projektu	Pytanie 1: Jaka jest trwałość instytucjonalna projektu?	Weryfikacja w trakcie kontaktu z beneficjentem, analiza dokumentów
		Pytanie 2: Jaka jest trwałość finansowa projektu?	Weryfikacja w trakcie kontaktu z beneficjentem, analiza dokumentów

Tabela 3. Matryca przystawalności projektu *Wykonanie koncepcji i dokumentacji technicznej budynku pn. Dom Gościnny „Hutnik" w Stalowej Woli* z wybranymi dokumentami strategicznymi. Przykład

Program/strategia / Projekt	NSRK 2004–2013 1. Rozbudowa i modernizacja istniejących instytucji kultury i ich adaptację do nowych funkcji	Program Strategii Rozwoju województwa podkarpackiego na lata 2007–2020. W priorytecie 3: *Rozwój kultury*, w kierunku działania 1*:* Zwiększenie możliwości dostępu do kultury	Program rozwoju kultury województwa podkarpackiego na lata 2004–2009
Wzrost konkurencyjności miasta Stalowa Wola i Lwowa przez rozbudowę infrastruktury służącej współpracy transgranicznej oraz aktywizację społeczną	●	●	◖
Poszukiwanie nowych elementów dziedzictwa kulturowego łączących oba kraje, na których można budować transgraniczny produkt kulturowy i nowe usługi turystyczno-kulturowe	◖	◖	◖
Wykorzystanie istniejącego potencjału zasobów dziedzictwa kulturowego przez stworzenie nowej oferty programowej promującej wspólne ponadregionalne wartości w zakresie kultury i rozwoju turystyki	◖	●	●
Rozwój wiedzy na temat słabej znajomości dziedzictwa kulturowego okresu 20-lecia międzywojennego na pograniczu polsko-ukraińskim i jej popularyzacja	○	◖	◖



Rysunek 3. Logika projektu *Wykonanie koncepcji i dokumentacji technicznej budynku pn. Dom Gościnny „Hutnik" w Stalowej Woli*. Wersja pierwotna



Rysunek 4. Logika projektu *Wykonanie koncepcji i dokumentacji technicznej budynku pn. Dom Gościnny „Hutnik" w Stalowej Woli*. Wersja po ewaluacji

model współpracy z partnerami z Ukrainy i czy jest on spójny ze wstępnymi założeniami programu kulturowego. Pomocne w rozstrzygnięciu tych zagadnień były także sprawozdania roczne, które Muzeum sporządza na potrzeby swojego organizatora, czyli miasta Stalowa Wola, w których opisano projekty transgraniczne oraz wymieniano instytucje kultury, współpracujące stale lub okresowo z Muzeum. Badania jakościowe pozwoliły zatem rozstrzygnąć część zagadnień, związanych ze skutecznością projektu (szanse na realizację wskaźników i uzyskania oddziaływań), a także uzyskać odpowiedzi na pytanie o potencjalną długotrwałość projektu (rozumianej jako szanse na długotrwałość współpracy).

Ostatnia część pracy znów koncentrowała się na badaniu tak dokumentacji zewnętrznej (np. badanie dokumentów założycielskich, bilansów oraz struktur finansowania Muzeum), jak i badaniach samego wniosku (np. kwestie związane z realizacją przez projekt polityk horyzontalnych UE).

Warto przy okazji zwrócić uwagę na specyficzne badanie kryterium skuteczności, które w pierwszej kolejności koncentrowało się na sprawdzeniu systemu wskaźników. W wyniku badania okazało się, że wskaźniki produktu oraz rezultatu były niepełne lub źle ustalone. Ostatecznie zweryfikowano je i poprawiono za pomocą wspomnianej już analizy SMART. Posegregowano je przede wszystkim według jednostki miar. W przypadku odpowiedzi na pytanie 1 i 2 w analizie SMART przystawiano je do właściwego wskaźnika i na tej podstawie dokonywano już ostatecznej oceny (tabela 4).

Tabela 4. Analiza SMART dla wskaźników projektu *Wykonanie koncepcji i dokumentacji technicznej budynku pn. Dom Gościnny „Hutnik" w Stalowej Woli*

Poziom interwencji	Wskaźniki	Kryteria					
		1	2	3	4	5	6
Produkt	Sztuka	+	+	+	+/-	+	+
	Komplet	+	+	+	+/-	+	+
Rezultat	Metr kwadratowy	+	+	+	+/-	+	+
	Sztuka	+	+	+	+/-	+	+

1 – czy wskaźnik w bezpośredni sposób odzwierciedla realizację produktu/rezultatu?
2 – czy wskaźnik jest precyzyjny?
3 – czy wskaźnik jest możliwy do zebrania?
4 – czy wskaźnik zaspokaja potrzeby informacyjne beneficjenta i zarządzających programem?
5 – czy wskaźnik odzwierciedlający produkt i rezultat są łatwe do zebrania w okresach raportowania?
6 – czy są szanse na ich osiągnięcie przy zaproponowanej wartości?

4.3. Jakie były wyzwania i ograniczenia procesu?

Odnośnie do barier i wyzwań, w trakcie ewaluacji nie zidentyfikowano praktycznie żadnych przeszkód. Najbardziej istotne ryzyko leżało u samych podstaw badania, w momencie efektywnego wykorzystania jego wyników.

Jak już wskazano, gotowy produkt badania, czyli zweryfikowany i poprawiony opis projektu został przekształcony we wniosek (zoperacjonalizowany). Wyniki tego etapu pracy miały zostać ostatecznie zweryfikowane nie przez Zamawiającego, ale przez struktury WST. W konsekwencji ewaluator miał świadomość, że projekt, a w efekcie wniosek aplikacyjny, nie tylko w warstwie techniczno-formalnej, ale przede wszystkim narracyjnej, będzie zderzony z poglądami, postawami i kompetencjami pracowników WST (co do których ewaluator miał rozeznanie, ale nie miał na nie wpływu).

Ocena wniosku aplikacyjnego w WST dokonywana była w sposób sformalizowany za pomocą kart oceny formalnej i technicznej. Pytania na karcie oceny technicznej koncentrują się jednak na aspektach przede wszystkim wykonalności projektu, nie aspektach merytorycznych, choć te ostatnie są bardzo wysoko punktowane[5]. Przy dużej dowolności w ocenie, powodzenie zależało od odpowiedniego sformułowania narracji we wniosku na podstawie spójnego katalogu słów-kluczy, oparcia projektu na kontekście Programu Sąsiedztwa. Miało to zapobiec swoistej „przemocy strukturalnej"[6]. Owa „przemoc (władza) strukturalna" polega na kreowaniu reguł gry, jaką jest relacja: oceniany (beneficjent i jego wniosek) – oceniający (WST) przez tego ostatniego. Reguły oceny sprowadzają się przy tym do stosowania w trakcie trwania tej relacji (oceny wniosku) jednolitych narzędzi (karty ocen i katalog pytań, jakie zadaje się treści wniosków) przy indywidualnej, odmiennej interpretacji treści, zawartych we wnioskach. W konsekwencji, co naturalne, każdy z pracowników odmiennie postrzega i interpretuje oceniany wniosek i tym samym odmiennie ocenia narracje we wnioskach, które są do siebie podobne. Dlaczego tak się dzieje? Na podstawie moich dotychczasowych doświadczeń mogę wyrazić pogląd, iż w procesie oceny zachodzi z jednej strony proces przeformalizowania, z drugiej zaś niedoformalizowania (Kożuch 2004). Przefornalizowanie nastąpiło przy badaniu elementów

---

[5] Aby projekt mógł być zatwierdzony na Komitecie Sterującym, musiał mieć minimum 65 punktów na 100 możliwych. W praktyce okazywało się, iż zdobycie za ocenę efektu transgranicznego projektu mniej niż 15 punktów powodowało negatywną ocenę merytoryczną. Szczegółowość pytań odnośnie do zasadniczych części wniosku była jednak bardzo niska. Dla przykładu: w ramach badania transgraniczności projektu (łącznie 35 pkt) badano dwa zagadnienia: 1) wpływ transgraniczny i poziom współpracy partnerskiej (25 pkt) oraz 2) czy projekt stwarza warunki do współpracy transgranicznej (10 pkt).

[6] Pojęcie inspirowane i zaadaptowane za: Staniszkis 2005.

wniosku o mniejszym znaczeniu technicznym, zaś niedoformalizowanie – przy elementach kluczowych. Niedoformalizowanie rozumiane jest tutaj jako brak odpowiedniej liczby informacji dających podstawę do podejmowania decyzji. Decyzje, szczególnie tak kluczowe jak decyzja o przyznaniu dofinansowana i uruchomieniu interencji publicznej, muszą zaś być podejmowane przy określonej liczbie przesłanek (Gasparski 2007). Brak przesłanek i danych do odpowiedzi na pytania zawarte w arkuszach ocen rodzi wątpliwości. W nawale wniosków i ich silnej konkurencji w procesie oceny przez pracowników WST zachodzi zjawisko opisane przez J. Scandurę. Człowiek, nie mogąc rozwiązać problemu (rozstrzygną wątpliwości przy ocenie wniosku), za pomocą reguły elementarnej odnosi się do metareguły (Scandura 1980; Strzałecki 2007). W procesie oceny wniosku metaregułą jest zakładanie błędu popełnionego przez beneficjentów i obniżenie mu punktacji w trakcie oceny.

4.4. Jaką rolę odegrali poszczególni interesariusze w procesie badawczym?

Badanie miało charakter partycypacyjny. Ewaluator, przez bieżący kontakt, współtworzył z Zamawiającym projekt. Nie była to więc ewaluacja z klasycznymi pisemnymi rekomendacjami, na które zamawiający opowiada pisemnie. Relacje z beneficjentem oparte były nie tylko na interakcjach merytorycznych, co jest nie do przecenienia w sytuacji, kiedy ewaluacja przebiega w sposób dynamiczny, a zmiany w opisie projektu zachodzą w sposób ciągły. Otwartość obu stron, wynikająca z wzajemnego szacunku i pozytywnych emocji, znacznie ułatwiała przyjmowanie racji drugiej strony. Przebieg badania pozwalał na budowanie poczucia współwłasności projektu, co stanowi w moim przekonaniu istotną wartość dodaną.

5. PRODUKTY BADANIA

5.1. Jaką wiedzę wytworzono w badaniu?

Stworzona w wyniku badania wiedza miała charakter techniczny i operacyjny. Dawała konkretne rekomendacje, wykorzystywane na bieżąco w tworzeniu projektu. W odniesieniu do samego badania projektu raczej nie zaskoczyła ewaluatora, potwierdziła natomiast intuicyjne założenia, iż można z powodzeniem wykorzystywać katalog kryteriów i pytań badawczych ewaluacji *ex ante* z poziomu programu do poziomu projektu. Ponadto potwierdziła, iż ewaluacja projektu, bazująca głównie na analizie różnych dokumentów jako zbiorów różnych narracji (zbioru informacji o rzeczywistości), ma silne zabarwienie konstruktywistyczne (por. Greene 2003) i jest oceną narracji przez drugą narrację.

Z punktu widzenia dalszego wykorzystania ewaluacji zaskakujące dla badacza były efekty przejścia z pozycji osoby oceniającej projekty i wnioski aplikacyjne do pozycji osoby ocenianej (ewaluator w konsekwencji silnej partycypacji był bowiem współautorem projektu) i uświadomienie sobie istniejących wad metodologii oceny wniosków w Programie Sąsiedztwa (o czym już była mowa w rozdziale 3.2).

W przypadku samego Zamawiającego, w wyniku wykorzystanych technik (drzew logicznych, matryc przystawalności, analiz SMART) mógł on sobie przyswoić narzędzia niezbędne do poprawy jakości swojej pracy. Zatem ewaluacja ma za zadanie wprowadzenie nie tyle kultury ewaluacyjnej (myślenia w kategorii ewaluacji), ile kultury projektowej (myślenia w kategorii odpowiedniości, użyteczności, skuteczności, efektywności długotrwałości samego projektu) i ostatecznie podnieść kompetencje zainteresowanych osób w Muzeum Regionalnym w Stalowej Woli (por. Cousins, Withmore 1998).

5.2. Jak wyglądał proces rozpowszechniania wyników?

Ewaluacja miała charakter wewnętrzny i służyła jedynie bezpośrednio Zamawiającemu. Szczegółowy opis tego badania został zawarty w pracy dyplomowej (Buchacz, Kościelecki 2007); jest on dostępny w Akademii Leona Koźmińskiego (za zgodą Zamawiającego).

6. WNIOSKI ZE STUDIUM PRZYPADKU

6.1. Jakie były efekty, rzeczywista użyteczność badania?

Początkowe założenia badawcze zrealizowano połowicznie. W wyniku badania udało się poprawić jakość projektu, uporządkować jego narrację, strukturę działań, celów, dobrać odpowiednie wskaźniki, ocenić, zweryfikować i zmienić ewentualne wszystkie istotne parametry projektu, w tym czas jego realizacji, koszty oraz zakres rzeczowy, doprecyzować grupę docelową projektu. Wnioski z badania wykorzystano w trakcie sporządzenia właściwej aplikacji. Porażka w konkursie przekreśliła jednak możliwość zastosowania wyników badania w praktyce. Wniosek bowiem nie otrzymał dofinansowania i projekt nie wszedł w fazę realizacji. Niemożliwe jest także ostateczne stwierdzenie, czy ewaluacja była poprowadzona źle, czy tekst wniosku został odczytany niezgodnie z intencjami beneficjenta, biorąc pod uwagę, że autor niniejszego tekstu byłby stroną w tym sporze.

Można jednak powiedzieć, że istnieje pewne długofalowe oddziaływanie tej ewaluacji. Po badaniu koncepcji przedsięwzięcia Muzeum otrzymało ściśle określoną, wypracowaną ścieżkę działania i gotowy projekt do aplikowania ze środków Programu Współpracy Transgranicznej

ENPI Polska – Białoruś – Ukraina 2007–2013 w obecnym okresie programowania. Identyfikacja produktów i rezultatów oraz nadanie im wartości wskaźników dyscyplinują Muzeum w dalszych działaniach na rzecz Domu Gościnnego „Hutnik" i jego wykorzystania na rzecz współpracy transgranicznej oraz w działaniach sektora kultury.

6.2. Jakie były silne i słabe strony analizowanej ewaluacji?

Do silnych stron ewaluacji *ex ante* projektu zaliczyłbym: jej relatywnie niskie koszty przy dość dużej wartości dodanej w postaci uzyskanej wiedzy dla Zamawiającego oraz możliwość przetestowania pewnych hipotez przez Wykonawcę; jasno sprecyzowany cel i bardzo praktyczny charakter badania, partycypacyjny charakter ewaluacji oraz potwierdzenie przydatności szerokiego spektrum badań gabinetowych, które są kluczowe przy pracy nad narracjami. Do słabych stron zaliczyłbym przede wszystkim niepewność co do weryfikacji wyników prac (wynikającą z przemocy strukturalnej po stronie WST), a z punktu widzenia metodycznego – ograniczoną użyteczność narzędzi jakościowych i ilościowych omawianego badania.

BIBLIOGRAFIA

Buchacz M., Kościelecki (2007), *Metody badań typu desk research w ewaluacji ex-ante projektu o charakterze transgranicznym*, Warszawa, praca dyplomowa w Akademii Leona Koźmińskiego, archiwum autorów.

Cousins J.B., Whithmore E. (1998), Framing participatory evaluation, *New Directions for Evaluation* **80**, 5–23.

European Commission (2000a), *Methodological Working Paper 2 – The Ex-Ante Evaluation of the Structural Funds interventions*, DG Regio, Luxembourg: The New Programming period 2000–2006.

European Commission (2000b), *Methodological Working Paper 7 – Ex Ante Evaluation and Indicators for INTERREG (Strand A and B)*, DG Regio, Luxembourg: The New Programming period 2000–2006.

European Commission (2006), *Working Document 1 – Indicative Guidelines on Evaluation Methods: Ex ante evaluation*, DG Regio, Luxembourg: The new Programming period 2007–2013.

Gasparski W. (2007), *Herberta A. Simona teoria administracji*, w: *Etos urzędnika,* D. Bąk (red.), Warszawa, Wydawnictwo WSPiZ.

Greene J.C (2003), *Understanding Social Programs through Evaluation*, w: *Collecting and interpreting qualitative materials,* Denzin N.K., Lincoln Y.S. (ed.), Thousand Oaks, Sage.

Kożuch B. (2004), *Zarządzanie publiczne w teorii i praktyce polskich organizacji,* Warszawa, Agencja Placet.

Olejniczak K. (2007), *Teoretyczne podstawy ewaluacji ex-post*, w: *Ewaluacja ex-post. Teoria i praktyka badawcza,* A. Haber (red.), Warszawa, PARP.

Olejniczak K. (2008) *Mechanizmy wykorzystania ewalkuacji. Studium ewaluacji średniookresowych INTERREG III,* Warszawa, Wydawnictwo Naukowe Scholar.

Program (2004), *Program Sąsiedztwa Polska – Białoruś – Ukraina INTERREG IIIA/Tacis CBC 2004–2006*, (nr CCI 2004 CB 16 0 PC 001), K(2004) 4366 (opublikowany: Warszawa 2005, WWPWP Phare).

Scandura J. (1980), Theoretical Foundations of Instruction: A Systems Alternative to Cognitive Psychology, *Journal of Structural Learning* **4**, 347–366.

Staniszkis J. (2005) *Postkomunizm. Próba opisu*, Gdańsk, Wyd. słowo/obraz/terytoria.

Strzałecki A. (2007), *Style rozwiązywania problemów praktycznych. Uwarunkowania psychologiczne*, w: *Współczesne wyzwania nauk praktycznych*, A. Lewicka-Strzałecka (red.), Warszawa, Wydawnictwo WSPiZ.

Uzupełnienie (2005), *Uzupełnienie Programu Sąsiedztwa Polska – Białoruś – Ukraina INTERREG IIIA/Tacis CBC 2004–2006*, zatwierdzony na I posiedzeniu Komitetu Monitorującego (opublikowany: Warszawa 2005, WWPWP Phare).

ROZDZIAŁ 3.3.

Bartosz Ledzion
Elżbieta Kozłowska

STUDIUM PRZYPADKU EWALUACJA *EX ANTE* PROGRAMU OPERACYJNEGO INFRASTRUKTURA I ŚRODOWISKO 2007–2013

1. KONTEKST I PRZEDMIOT EWALUACJI

Ewaluacja *ex ante* Programu Operacyjnego Infrastruktura i Środowisko (dalej POIiŚ) została wybrana do niniejszego opracowania bardziej ze względu na charakter programu niż na możliwość wskazania badania jako przykładu dobrych praktyk. Opisany przypadek przedstawia przebieg jednej z pierwszych ewaluacji typu *ex ante* zrealizowanych w Polsce, a co za tym idzie, pokazuje trudności związane z realizacją badań, wypracowaniem wniosków oraz nowymi doświadczeniami – zarówno po stronie zamawiającego, jak i wykonawcy. W stosunku do pozostałych przykładów ewaluacja POIiŚ stanowi ilustrację oceny programu wielosektorowego o zasięgu ponadregionalnym. Ponadto, warto przy tym zaznaczyć, że unikalność badanego programu nie wynika jedynie ze złożoności jego struktury wewnętrznej, ale także faktu największej alokacji

finansowej w historii Unii Europejskiej. Przed realizacją badania łączna wielkość środków finansowych zaangażowanych w realizację Programu Operacyjnego Infrastruktura i Środowisko na lata 2007–2013 wyniosła ok. 30 mld EUR.

Powyższe studium może stanowić interesujący przypadek dla osób zajmujących się badaniami inicjatyw o charakterze infrastrukturalnym oraz wielosektorowym. Wiedza wynikająca z tego projektu może być niezwykle cenna w perspektywie planowania podobnych badań w przyszłości.

2. OPIS PRZEDMIOTU BADANIA

Opis procesu planowania i projektowania badania warto rozpocząć od zarysowania kontekstu politycznego powstania samego Programu Operacyjnego Infrastruktura i Środowisko. Idea jednego programu operacyjnego, scalającego wszystkie ponadregionalne działania w zakresie infrastruktury, zarządzanego na poziomie krajowym, powstała wraz ze zmianą rządu w 2005 roku i powołaniem nowego ministerstwa, tj. właściwego ds. rozwoju regionalnego (Rozporządzenie Rady Ministrów z dnia 31 października 2005 roku w sprawie utworzenia Ministerstwa Rozwoju Regionalnego). Fakt ten miał ogromny wpływ na perspektywę instytucjonalną badania. System instytucjonalny POIiŚ przewidywał zaangażowanie kilkudziesięciu instytucji w proces wdrażania programu, poczynając od Instytucji Zarządzającej, Instytucji Pośredniczących dla poszczególnych priorytetów, na Instytucjach wdrażających przydzielonych do określonych działań kończąc. Tak rozbudowany system wdrażania stanowił istotne ograniczenie w procesie realizacji samego badania, jak i konsultacji wniosków końcowych.

Obecnie Program prowadzi interwencję w 6 sektorach (środowisko, transport, energetyka, kultura, zdrowie i szkolnictwo wyższe), na łączną wielkość środków finansowych 37,6 mld EUR.

3. DLACZEGO PODJĘTO BADANIE?

Konieczność podjęcia tego badania wynikała wprost z zapisów rozporządzenia Rady WE nr 1083/2006, ustanawiającego przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego oraz Funduszu Spójności i uchylającego rozporządzenie (WE) nr 1260/1999, zgodnie z którymi państwa członkowskie zostały zobowiązane do przeprowadzenia oceny *ex ante*. W kontekście czasowym oraz instytucjonalnym, w jakim badanie było podejmowane,

oraz biorąc pod uwagę szerokie tło, tj. kulturę administracyjną, w której brakuje tradycji badań ewaluacyjnych, trudno mówić w przypadku ewaluacji *ex ante* POIiŚ o świadomej samodzielnej decyzji decydentów. Również główny cel badania, tj. optymalizacja alokacji zasobów budżetowych w ramach programu i poprawa jakości programowania, został zaczerpnięty wprost z zapisów wspomnianego rozporządzenia.

4. PLANOWANIE I PROJEKTOWANIE EWALUACJI

Badanie zostało zaprojektowane, zlecone oraz nadzorowane przez Instytucję Zarządzającą POIiŚ. Instytucje Pośredniczące zaopiniowały projekt badania, opracowały katalog strategicznych dokumentów sektorowych, niezbędnych do oceny spójności zewnętrznej oraz uczestniczyły w konsultacjach wyników badań, zmieniając na bieżąco zapisy programowe.

Na etapie planowania ewaluacji *ex ante* została powołana grupa sterująca, w skład której wchodził jeden przedstawiciel IZ zajmujący się wyłącznie ewaluacją oraz po kilku przedstawicieli sześciu Instytucji Pośredniczących. Ze względu na fakt, iż w czasie przeprowadzania ewaluacji *ex ante* w IP POIiŚ nie wyznaczono osób odpowiedzialnych za ewaluację, w pracach grupy uczestniczyły osoby zajmujące się bezpośrednio programowaniem. Z jednej strony ułatwiło to bieżące wprowadzanie rekomendacji w trakcie prac nad programem, z drugiej jednak ograniczyło aktywną rolę Instytucji Pośredniczących w samym procesie projektowania oraz nadzoru nad ewaluacją.

Celem ewaluacji *ex ante* POIiŚ były optymalizacja alokacji zasobów budżetowych oraz poprawa jakości programowania. Cele te zostały osiągnięte przez realizację celów cząstkowych, takich jak:

- ocena alokacji środków na poszczególne działania,
- ocena kwantyfikacji wartości docelowych wskaźników osiągnięcia celów PO,
- weryfikacja zidentyfikowanych obszarów problemowych dla poszczególnych sektorów,
- weryfikacja spójności wewnętrznej i zewnętrznej PO,
- ocena projektu systemu wdrażania PO.

W pracach ewaluacyjnych rolę odegrał również powołany przez Krajową Jednostkę Oceny Zespół koordynujący prace nad ewaluacją *ex ante* dokumentów strategicznych na lata 2007–2013. Podczas posiedzeń tego zespołu Wykonawca prezentował wyniki, uzyskane na danym etapie prowadzonego badania. Członkowie zespołu przyjęli entuzjastycznie raport metodologiczny zespołu badawczego. Problemy pojawiły się na etapie prezentowania wstępnych wyników, podczas którego rozgorzały

dyskusje nad częścią środowiskową. Komentarze dotyczyły pojedynczych stwierdzeń zespołu badawczego nie kwestionowano metodologii. W toku prac ewaluacyjnych zorganizowano również spotkanie zespołów przygotowujących ewaluację *ex ante* oraz prognozę oddziaływania POIiŚ na środowisko.

Podsumowując zaangażowanie poszczególnych podmiotów, można stwierdzić, że ewaluacja *ex ante* POIiŚ była zarządzana raczej z poziomu operacyjnego niż strategicznego. Zlecający ewaluację, określając jej zakres, oparł się na dokumentach roboczych Komisji Europejskiej. Poszczególne komponenty badania zostały zaczerpnięte dosłownie z ww. dokumentu. Zamawiający nie wskazał szczególnych potrzeb informacyjnych, wykraczających poza wymienione w wytycznych Komisji Europejskiej, natomiast określił bardzo dużą liczbę pytań ewaluacyjnych, co skutkowało m.in. tym, że zespół badawczy skupił się na odpowiedzi na poszczególne pytania, tracąc czasami obraz całości.

Zamawiający nie określił szczegółowo metodologii badania, wskazał jedynie, że ewaluatorzy muszą oprzeć się na analizie dokumentów oraz wyników poprzednich ewaluacji, posiłkując się dodatkowymi metodami, takimi jak: analizy porównawcze, panele ekspertów, analiza SWOT, zogniskowane wywiady grupowe, wywiady indywidualne, badania ankietowe.

Zamawiający oszacował badanie na ok. 200 tys. PLN oraz zaplanował jego realizację na trzy i pół miesiąca od momentu podpisania umowy. Wyznaczony czas badania przy tak złożonym przedmiocie badania

Priorytetowe pytania ewaluacyjne

- Czy program przedstawia właściwą strategię, pozwalającą sprostać wyzwaniom, przed którymi stoi dany sektor?
- Czy strategia programu jest właściwe zdefiniowana, ma przejrzyste cele i priorytety?
- Czy realne jest osiągnięcie tych celów przy danej alokacji zasobów finansowych na poszczególne priorytety?
- Czy strategia jest spójna z politykami sektorowymi, krajowymi i Wspólnoty?
- Czy zidentyfikowano właściwe wskaźniki dla poszczególnych celów?
- Czy zaproponowane wskaźniki i ich wartości docelowe tworzą podstawę przyszłego sytemu monitoringu i ewaluacji wdrażania danej strategii programu?
- Jakie będzie – w kategoriach ilościowych – oddziaływanie strategii programu?
- Czy system wdrażania jest zaprojektowany w sposób umożliwiający realizację programu i osiągnięcie jego celów?

okazał się niewystarczający. Jeśli chodzi o zapewnienie dostępności danych, to Zamawiający przedstawił w opisie zamówienia katalog dokumentów, z którymi powinien zapoznać się ewaluator. Problemem okazała się natomiast weryfikacja rzetelności danych zapisanych w części diagnostycznej obszarów interwencji POIiŚ.

Ze względu na multisektorowy charakter programu, Zamawiający, określając warunki, które musiał spełnić zespół badawczy, skupił się na wymaganiach wobec poszczególnych członków zespołu. Do badania musiał zostać wyznaczony minimum sześcioosobowy zespół, składający się z osób z kwalifikacjami lub doświadczeniem analitycznym w obszarach: ochrony środowiska, transportu, energetyki, infrastruktury ochrony zdrowia oraz infrastruktury kultury. Wymóg posiadania eksperta w zakresie infrastruktury szkolnictwa wyższego nie pojawił się w opisie zamówienia, gdyż na etapie zlecania ewaluacji *ex ante* kwestie szkolnictwa wyższego nie były włączone w POIiŚ. Ewaluacja priorytetu szkolnictwa wyższego została zlecona jako zamówienie dodatkowe z wolnej ręki, ze względu na fakt, iż nie można było przewidzieć na etapie planowania zamówienia, że wystąpią takie zmiany w zakresie programu.

W odpowiedzi na wyzwania związane z realizacją badania Wykonawca zaproponował 13-osobowy zespół badawczy, składający się zarówno z ekspertów sektorowych, jak i badaczy mających doświadczenie ewaluacyjne. W stosunku do propozycji ofertowej liczba osób zaangażowanych w badanie wzrosła trzykrotnie.

5. CZYNNIKI WPŁYWAJĄCE NA KONTEKST REALIZACJI BADANIA

Brak doświadczenia podmiotu zlecającego

Dotychczasowe ewaluacje *ex ante* dokumentów programowych były zlecane horyzontalnie, tj. w przypadku dokumentów na lata 2004–2006 zostały one włączone w umowę bliźniaczą Ministerstwa Gospodarki z francuską instytucją DATAR, natomiast projekt Narodowego Planu Rozwoju na lata 2007–2013 (tj. poprzednika Narodowej Strategii Spójności 2007–2013) przygotowany przed zmianą rządu został poddany ocenie szacunkowej zleconej przez Krajową Jednostkę Oceny (*Raport końcowy z oceny szacunkowej wstępnego projektu Narodowego Planu Rozwoju na lata 2007–2013*). Oznaczało to, iż Instytucje Zarządzające (IZ) poszczególnymi programami operacyjnymi nie miały doświadczenia w przeprowadzaniu ewaluacji *ex ante* dokumentów programowych. Co więcej, w przypadku IZ POIiŚ brakowało doświadczenia w zlecaniu jakichkolwiek badań ewaluacyjnych, gdyż instytucja ta, jako zarządzająca w perspektywie finansowej 2004–2006 Funduszem Spójności, nie dysponując środkami z Pomocy Technicznej, nie powołała jednostki ewaluacyjnej.

Ograniczony czas procesu programowania i ewaluacji

Drugim czynnikiem, mającym wpływ na proces planowania i projektowania ewaluacji *ex ante* POIiŚ, obok braku doświadczenia ewaluacyjnego IZ, był brak czasu. Zmiana koncepcji dokumentów programowych na lata 2007–2013 pod koniec 2005 roku ograniczała niewątpliwie ramy czasowe zarówno samego procesu programowania, jak i realizacji ewaluacji *ex ante*. W takim terminie niemożliwe było zastosowanie się do wytycznych KE, zalecających „proces interakcyjny i interaktywny, w którym ewaluatorzy realizują poszczególne komponenty ewaluacji w miarę postępu przygotowywania programu" (*Nowy okres programowania 2007–2013. Przewodnik po metodach ewaluacji: ewaluacja ex ante. Dokument roboczy nr 1*). Mimo że w procesie ewaluacji *ex ante* POIiŚ nie zastosowano *explicite* ww. wskazówki, zapewniono dwojakiego rodzaju interaktywność. Z jednej strony, ze względu na fakt, iż prace nad POIiŚ trwały równolegle do procesu oceny, ewaluator był zobowiązany do uaktualniania wyników badania w celu uwzględnienia ewentualnych zmian i uzupełnień dokonanych w zapisach programu. Z drugiej zaś, uwagi do zapisów programu były uwzględniane na bieżąco podczas procesu programowania.

Co więcej, w przypadku ewaluacji *ex ante* dokumentów programowych na lata 2007–2013 zachodziła obawa, że raporty końcowe nie powstaną do czasu rozpoczęcia negocjacji z KE, planowanych pierwotnie na początek października 2006 roku. W związku z powyższym, w opisie przedmiotu zamówienia umieszczono zapis zobowiązujący Wykonawcę, w przypadku zaistnienia sytuacji opisanej powyżej, do przedstawienia dodatkowego materiału (w języku polskim i angielskim) opisującego stan prac nad oceną oraz jej wstępne wyniki. Proces programowania rozciągnął się jednak w czasie i negocjacje z KE rozpoczęły się po zakończeniu ewaluacji *ex ante*, tj. po listopadzie 2006 roku.

Kontrowersyjny charakter działań będących przedmiotem ewaluacji

Kolejnym czynnikiem sytuującym ewaluację *ex ante* POIiŚ był wspomniany charakter programu. Po pierwsze, liczba badań ewaluacyjnych zleconych w sektorze infrastruktury była znikoma, co oznaczało, iż nie tylko IZ nie dysponowała doświadczeniem ewaluacyjnym, ale również podmioty funkcjonujące na rynku ewaluacji nie miały doświadczenia w zakresie badań w tym sektorze. Zgodnie z informacjami zawartymi w bazie badań ewaluacyjnych, przed 2006 rokiem zrealizowano tylko trzy oceny działań infrastrukturalnych. Badania te zleciła Krajowa Jednostka Oceny (szerzej zob. rozdział 1.4), a dotyczyły one: struktury i spójności projektów z zakresu infrastruktury drogowej; przewidywanego wpływu inwestycji infrastrukturalnych realizowanych na obszarach wiejskich i wiejsko-miejskich oraz przewidywanego wpływu wybieranych projektów

z zakresu modernizacji i rozbudowy infrastruktury społecznej na rozwój społeczno-gospodarczy kraju. Wartość tych badań nie przekraczała 30 tys. PLN. Oznaczało to, iż ewaluacja *ex ante* POIiŚ była pierwszym tak dużym badaniem w sektorze infrastruktury (zarówno dla zlecających, jak i wykonujących). Ponadto, w czasie, gdy troska o środowisko naturalne jest przedmiotem szczególnego zainteresowania Komisji Europejskiej oraz opinii publicznej, działania infrastrukturalne budzą wiele obaw i kontrowersji, co również miało niebagatelny wpływ na sam proces ewaluacji i konsultacji z Komisją Europejską.

Multisektorowy charakter przedmiotu ewaluacji

Dodatkowym wyzwaniem w procesie planowania i realizacji badania był multisektorowy charakter programu, który, po pierwsze, wpłynął na koszty badania oraz szerokie spektrum zaangażowanych instytucji, po drugie – na skład zespołu ewaluacyjnego oraz koordynację jego prac, a po trzecie – na wyniki samego badania w postaci raportu. O wszystkich ww. skutkach będzie mowa w dalszej części rozdziału.

Waga przedmiotu ewaluacji w dyskusji politycznej i społecznej

Ostatnią determinantą kontekstu ewaluacji *ex ante* POIiŚ była skala środków nań przeznaczonych, a co za tym idzie – waga programu w dyskusji politycznej oraz społecznej w kraju i za granicą. Otóż program ten budził szerokie zainteresowanie, jako największy (pod względem wartości) w historii polityki spójności UE program operacyjny. Z drugiej strony, przy stanie infrastruktury w kraju, skala potrzeb w tym zakresie w Polsce przewyższa dostępne środki, co oznacza, iż decyzje, mające wpływ na ich podział, zapadają na najwyższych szczeblach. Fakt ten nie pozostał bez wpływu na sposób wykorzystania wyników ewaluacji, o czym będzie mowa pod koniec niniejszego rozdziału.

6. PROCES BADAWCZY

Jakie podejście badawcze i metodologię przyjął ewaluator?[1]

W wyniku przeprowadzonych konsultacji z Zamawiającym i po wstępnej analizie programu oraz zestawu pytań badawczych, za obopólną zgodą (IZ POIiŚ) określono zestaw pytań i obszarów priorytetowych. Ewaluatorzy zwrócili szczególną uwagę na analizę przedstawionych poniżej obszarów badawczych.

[1] *Ewaluacja ex-ante Programu Operacyjnego Infrastruktura i Środowisko*, Kantor Doradcy w Zarządzaniu Sp. z o.o., Ministerstwo Rozwoju Regionalnego, s. 35–36.

- **W ramach analizy obszaru 1: Weryfikacji analizy społeczno-ekonomicznej oraz trafności rankingu głównych zidentyfikowanych zróżnicowań** badacze wykonali analizę dysproporcji rozwojowych i potencjału badanych sektorów; utworzono ranking zróżnicowań, dokonano także weryfikacji trafności diagnoz programowych dla poszczególnych sektorów. Jednocześnie eksperci – wykonawcy prowadzili ocenę adekwatności priorytetów (celów, działań) w stosunku do zidentyfikowanych, rzeczywistych potrzeb społeczno-gospodarczych. Efektem prac badaczy było uzupełnienie istniejących analizy sektorowych, wraz z poprawą trafności strategii programu.
- **W ramach analizy obszaru 2: Oceny uzasadnienia ekonomicznego oraz spójności wewnętrznej strategii** badacze podjęli próbę weryfikacji zasadności ekonomicznej strategii programu na podstawie analizy celów i priorytetów pod kątem niedoskonałości rynku oraz kontekstu obszaru interwencji.
- **W ramach obszaru 3: Oceny spójności strategii z regionalnymi i krajowymi politykami oraz Strategicznymi Wytycznymi Wspólnoty** badacze przeanalizowali spójność strategii programu, opierając się na analizie dokumentów z wykorzystaniem matryc zgodności celów.
- **Kolejny ważny obszar badawczy dotyczył „Oceny oczekiwanego rezultatu i oddziaływania programu".** W ramach prowadzonych prac badacze dokonali weryfikacji jakości zaproponowanego systemu wskaźników; ponadto podjęto próbę weryfikacji wartości bazowych i docelowych wskaźników programu z uwzględnieniem alokacji środków finansowych na poszczególne działania. W tym obszarze badacze głównie opierali się na wiedzy eksperckiej oraz nielicznych opracowaniach sektorowych.

Ostatnim, aczkolwiek bardzo istotnym obszarem ewaluacji *ex ante*, była **ocena projektu systemu wdrażania programu**. Badacze podjęli się oceny zasad zarządzania, monitorowania i ewaluacji programu, oceniono ryzyko wystąpienia wąskich gardeł w procesie wdrażania, ponadto dokonano analizy potencjału instytucjonalnego organizacji zaangażowanych we wdrażanie POIiŚ, wykorzystując przy tym metodologię stosowaną do audytów organizacyjnych.

Jakie były wyzwania i ograniczenia procesu badawczego?

W trakcie realizacji badania, ewaluatorzy natknęli się na wiele ograniczeń i trudności, które w sposób pośredni i bezpośredni przełożyły się na efekt końcowy procesu ewaluacyjnego.

Istotna słabość całego procesu wynikała ze specyfiki przedmiotu badania. Zbyt duży zakres niektórych działań programu spowodował większą koncentrację badaczy na ocenie poszczególnych części programu (priorytetów) niż na całościowej ocenie. Takie podejście spowodowało,

że raport miał zbyt rozbudowaną część analityczną, która utrudniała w znacznym stopniu czytanie i zrozumienie dokumentu. Ostatecznie raport miał większą objętość od dokumentu programowego.

Kolejnym elementem przekładającym się na końcową jakość raportu były rozbieżności między wynikami ewaluacji *ex ante* a prognozą oddziaływania na środowisko. W efekcie Komisja Europejska zaleciła ponowne przeprowadzenie prognozy.

Pewne problemy wynikały także z braku doświadczenia badaczy w badaniach typu *ex ante*. W fazie początkowej określono zbyt ambitne założenia na poziomie raportów metodologicznych, których nie można było zrealizować ze względu na krótkie terminy i ograniczone zasoby.

Jakie były efekty, rzeczywista użyteczność badania?

Rekomendacje z ewaluacji POIiŚ można podzielić na te dotyczące *stricte* zapisów programu operacyjnego – postulujące zarówno merytoryczne, jak i redakcyjne zmiany w treści programu, oraz rekomendacje dotyczące systemu zarządzania i wdrażania.

Pierwsza z wymienionych grup rekomendacji była przedstawiana na bieżąco przez ewaluatora i analizowana przez Instytucję Zarządzającą oraz Instytucje Pośredniczące, równolegle do procesu programowania. Takie podejście zaowocowało dużym stopniem uwzględnienia uwag dotyczących części diagnostycznej programu – praktycznie we wszystkich sektorach zostały poprawione dane statystyczne, lepiej opisano ich źródła, jak również uporządkowano analizę głównych problemów, na które POIiŚ odpowiada oraz analizę słabych i mocnych stron danych sektorów. Na bieżąco wykorzystywano również rekomendacje uspójniające wewnętrznie cały dokument.

Z najważniejszych rekomendacji merytorycznych IZ uwzględniła następujące:

- wprowadziła zmiany w liście beneficjentów interwencji w sektorze transportu rozszerzając ich katalog,
- dopuściła finansowanie dużych budowli hydrotechnicznych wyłącznie w przypadku wykluczenia zastosowania alternatywnych rozwiązań,
- ograniczyła finansowanie inwestycji wodociągowych wyłącznie do projektów obejmujących swym zakresem również infrastrukturę ściekową,
- ograniczyła wskaźnik głównego celu programu, tj. zajmowanego przez Polskę miejsca w światowym rankingu konkurencyjności do obszaru infrastruktury.

Natomiast wiele zasadniczych rekomendacji merytorycznych nie zostało uwzględnionych. W sektorze transportu, pomimo rekomendacji ewaluatora, aby zweryfikować zasadność finansowania działań z zakresu żeglugi śródlądowej ograniczonymi w stosunku do potrzeb środkami,

pozostawiono zapisy bez zmian. Nie zwiększono również alokacji na priorytet dotyczący rozwoju przyjaznych środków transportu, jak również nie zrezygnowano z finansowania projektów dotyczących lotnisk, które w opinii ewaluatora są projektami generującymi dochód pokrywający koszty budowy i utrzymania infrastruktury, oraz infrastruktury drogowej w Polsce wschodniej, gdzie ewaluator proponował zrezygnować z odrębnego działania lub rozszerzyć je również na inne gałęzie transportu. W sektorze środowiska nie zwiększono środków na wypełnianie zobowiązań akcesyjnych Polski oraz na poprawę bioróżnorodności. Nie wyłączono również finansowania infrastruktury szkolnictwa wyższego z POIiŚ, postulowanego przez ewaluatora, który proponował przeniesienie wsparcia dla tych działań do Programu Operacyjnego *Innowacyjna Gospodarka*. W przypadku podjęcia decyzji odnośnie do pozostawienia wsparcia dla ww. przedsięwzięć w ramach POIiŚ, ewaluator wnosił o ograniczenie finansowania infrastruktury szkolnictwa wyższego wyłącznie do wydziałów nauk ścisłych i technicznych, co nie zostało uwzględnione.

Negatywne stanowisko IZ wobec wymienionych rekomendacji mogło wynikać z opisanego na wstępie niniejszego rozdziału charakteru badanego programu, tj. jego kontrowersyjności oraz niewystarczającej liczby środków wobec potrzeb rozwojowych Polski, jak również z braku tradycji ewaluacyjnych w procesie decyzyjnym.

Jeśli chodzi natomiast o rekomendacje wybiegające w przyszłość, dotyczące aspektów zarządzania programem, to należy zaznaczyć, iż ewaluator przedstawił wnioski wykraczające poza wymagania opisu zamówienia, ponadto przeprowadził dwa dodatkowe szkolenia z zakresu zarządzania oraz opracowywania studiów wykonalności, co stanowiło wartość dodaną całego procesu ewaluacyjnego.

ROZDZIAŁ 3.4.

Adam Płoszaj

EWALUACJA *EX ANTE* RPO LUBUSKIE 2007–2013

1. KONTEKST I PRZEDMIOT EWALUACJI

1.1. Opis przedmiotu badania

Przedmiotem badania był Regionalny Program Operacyjny (RPO) województwa lubuskiego. RPO jest dokumentem strategicznym, opisującym sposób rozdysponowania części środków pomocowych UE przeznaczonych dla regionu lubuskiego. Dokument ten określa priorytety i cele, na których realizację mają być przeznaczone fundusze z okresu programowania 2007–2013. Ostateczna wersja RPO obejmuje szeroki zakres interwencji w ramach 5 priorytetów:

- Priorytet I. Rozwój infrastruktury wzmacniającej konkurencyjność regionu.
- Priorytet II. Stymulowanie wzrostu inwestycji w przedsiębiorstwach i wzmocnienie potencjału innowacyjnego.
- Priorytet III. Ochrona i zarządzanie zasobami środowiska przyrodniczego.
- Priorytet IV. Rozwój i modernizacja infrastruktury społecznej.
- Priorytet V. Rozwój i modernizacja infrastruktury turystycznej i kulturowej.

Badanie przeprowadzono wiosną 2006 roku, w momencie gdy była gotowa wstępna wersja RPO i gdy było jeszcze dużo czasu na nawet bardzo znaczące zmiany jego zapisów (program został ostatecznie zatwierdzony 2 października 2007 roku decyzją Komisji Europejskiej). Był to zatem bardzo dobry moment na przeprowadzenie ewaluacji typu *ex ante* – jej wyniki mogły być łatwo (przynajmniej teoretycznie) wprowadzone

do dokumentu. Na marginesie należy zaznaczyć, że w polskiej praktyce bardzo często mamy do czynienia z sytuacją, gdy ewaluacja *ex ante* jest zlecana i realizowana w taki sposób, że jej wyniki nie mogą mieć już znaczącego, czy nawet żadnego wpływu na kształt programu, ponieważ po prostu nie ma czasu na zmianę zawartości danego programu.

1.2. Dlaczego podjęto badanie?

Głównym powodem przeprowadzenia badania ewaluacyjnego omawianego w niniejszym rozdziale była potrzeba praktycznego przetestowania metod i narzędzi ewaluacji *ex ante* wypracowanych w ramach projektu badawczego SRDTOOLS – *Metody i techniki oceny wpływu polityki spójności na rozwój trwały*[1], finansowanego w ramach VI Ramowego Programu Badań UE.

Istotą badania nie była zatem ewaluacja sama w sobie, lecz ocena użytych narzędzi, metod, podejścia badawczego. Mimo że dla zespołu realizującego projekt SRDTOOLS nadrzędnym celem była ocena metod i narzędzi, wykonane badanie ewaluacyjne zostało zaplanowane i przeprowadzone tak, aby można je było wykorzystać jako „zwykłą" ewaluację *ex ante* programu.

1.3. Kim byli kluczowi interesariusze?

Omawiana ewaluacja nie była zlecona przez instytucje odpowiedzialne za kształt RPO. Nie była też finansowana jako typowa ewaluacja. Została wymyślona i przeprowadzona jako element (zresztą bardzo istotny) szerszego projektu o czysto naukowym charakterze. Mimo to, główni odbiorcy ewaluacji (Urząd Marszałkowski i inni interesariusze zaangażowanych w tworzenie RPO) byli aktywnie włączeni w jej realizację. Co więcej, aktywny udział interesariuszy był wręcz nieodzowny dla przeprowadzenia badania, którego kluczowym element stanowią warsztaty z udziałem przedstawicieli instytucji uczestniczących w pracach nad RPO.

1.4. Jaki to był typ ewaluacji?

Rozpatrując omawiane tu badanie jako ewaluację i przypisując je do poszczególnych typów ewaluacji, należy stwierdzić, że była to:

- ewaluacja *ex ante* (badanie prowadzone było w momencie kluczowym dla ostatecznego sformułowania tekstu programu),
- ewaluacja programu (przedmiotem badania był projekt Regionalnego Programu Operacyjnego województwa lubuskiego),

---

[1] W realizacji polskiej części projektu SRDTOOLS prowadzonego w Centrum Europejskich Studiów Regionalnych i Lokalnych EUROREG (Uniwersytet Warszawski) brali udział: Grzegorz Gorzelak, Elżbieta Kozłowska, Karol Olejniczak, Adam Płoszaj, Maciej Smętkowski. Więcej informacji na temat projektu można znaleźć na stronie: www.srdtools.info; por. też Ekins 2008a.

- ewaluacja zewnętrzna (głównymi wykonawcami badania były osoby niezwiązane z formułowaniem programu).

2. WKŁAD

2.1. Jakie były potrzeby informacyjne i jakie pytania postawił Zamawiający?

Głównym celem projektu było praktyczne przetestowanie metod i narzędzi badawczych, jednak do tego celu wybrano taki przypadek, aby wyniki przeprowadzonego badania miały jak największe praktyczne zastosowanie. Wybór do badań lubuskiego RPO był niejako automatyczny, gdyż wcześniejsze badania w ramach projektu dotyczyły właśnie tego województwa.

W przypadku niniejszego badania nie było instytucji zamawiającej (zlecającej), więc Wykonawca mógł w pełni decydować o wszystkich aspektach badania, także o pytaniach badawczych. Podstawowe pytanie w sferze metodologicznej zostało sformułowane w następujący sposób: czy analizowane metody (cztery kapitały, trendy i progi krytyczne) są skutecznym narzędziem ewaluacji *ex ante*? Aby umożliwić odpowiedź na to pytanie, należało jednak wcześniej odpowiedzieć na pytanie dotyczące wprost RPO: w jaki sposób realizacja RPO wpłynie na stan czterech kapitałów w województwie lubuskim?

2.2. Jak sprecyzowano zakres badania (przedmiotowy, przestrzenny, czasowy)

Przedmiotem badania był Regionalny Projekt Operacyjny województwa lubuskiego na lata 2007–2013. Tak określony przedmiot badania w oczywisty sposób zdefiniował również zakres przestrzenny (województwo lubuskie) oraz czasowy badania (potencjalny wpływ RPO na sytuację w regionie w okresie programowania: lata 2007–2013).

2.3. Jakie założenia metodologiczne postawiono w OPZ?

Jak już podkreślano, badanie nie było zamawiane, zatem nie była też w tym przypadku przeprowadzana procedura przetargowa.

2.4. Jakie postawiono warunki operacyjne (budżet, czas badania, zespół, dostępność danych)?

Warunki operacyjne badania były ustalone przez zespół badawczy; w zasadzie jedynymi ograniczeniami były tutaj ramy projektu SRDTOOLS, które jednak pozostawiały dość dużą swobodę – szczególnie wobec tego, że ewaluacja RPO była tylko jednym z elementów o wiele szerszego

projektu. Ewaluacja RPO była tak bardzo związana z innymi działaniami podejmowanymi w ramach projektu SRDTOOLS, że trudno wydzielić ją jako odrębną całość, szczególnie w sensie czasu realizacji badania, a przede wszystkim w odniesieniu do budżetu przeznaczonego na realizację omawianego badania ewaluacyjnego (głównymi kosztami badania były wynagrodzenia zespołu badawczego, które wyznaczano na podstawie zaangażowania w całość projektu SRDTOOLS, a nie jedynie w część dotyczącą ewaluacji RPO).

Ewaluację realizowali pracownicy Centrum Europejskich Studiów Regionalnych i Lokalnych EUROREG. W skład zespołu weszli: Grzegorz Gorzelak, Karol Olejniczak, Adam Płoszaj oraz Maciej Smętkowski.

3. PROCES BADAWCZY

3.1. Jakie podejście badawcze i metodologię przyjął ewaluator?

Metodologię badania oparto na koncepcji czterech kapitałów, progów i trendów krytycznych oraz sprzeczności rozwojowych. Jak już wspominano, praktyczne przetestowanie tych koncepcji było jednym z głównych celów projektu SRDTOOLS, w ramach którego przeprowadzono omawianą ewaluację. Ponieważ koncepcje te są stosunkowo nowe i wiedza o nich nie jest jeszcze szeroko rozpowszechniona i łatwo dostępna, poniżej przedstawiono ich ogólną charakterystykę.

Cztery kapitały

Koncepcja czterech kapitałów sformułowana przez Paula Ekinsa (1992) mówi, że wszystkie czynniki istotne dla trwałego rozwoju gospodarczego można przypisać jednemu z czterech kapitałów: gospodarczemu, naturalnemu, ludzkiemu i społecznemu, oraz że trwałość rozwoju wymaga zachowania równowagi między tymi kapitałami. Poszczególne kapitały rozumiane są w tej koncepcji w następujący sposób (Ekins 2008; por. też Płoszaj 2007):

- **Kapitał gospodarczy** (czy też wytworzony przez człowieka) – obejmuje wszystko to, co tradycyjnie rozumiane jest jako kapitał (wytworzone zasoby, wykorzystywane do produkcji innych dóbr i usług). Do tego kapitału zalicza się np.: maszyny, narzędzia, budynki, infrastrukturę.
- **Kapitał naturalny** – obejmuje wszystkie formy ekosystemu i zasobów naturalnych, które uczestniczą w tworzeniu społecznego dobrobytu. A zatem, poza tradycyjnie rozumianymi zasobami naturalnymi (takimi jak drewno, woda, energia i zasoby mineralne), do kapitału naturalnego zalicza się także zasoby naturalne, które nie dają się łatwo wycenić, np.: bioróżnorodność, gatunki zagrożone, czy też ekosystem, który jest dostarczycielem ekologicznych usług (takich jak filtrowanie wody i powietrza). Kapitał naturalny może być rozpatrywany

jako elementy natury, które mogą być powiązane bezpośrednio lub pośrednio ze społecznym dobrobytem.

- **Kapitał ludzki** – odnosi się do zdrowia, dobrobytu i potencjału produkcyjnego indywidualnych ludzi. Elementami tego kapitału są np. zdrowie fizyczne i psychiczne człowieka, wykształcenie, motywacja i umiejętności. Elementy te nie tylko przyczyniają się do budowy szczęśliwego, zdrowego społeczeństwa, lecz także, przez dostarczenie produktywnej siły roboczej, stwarzają możliwości dla rozwoju gospodarczego.
- **Kapitał społeczny** – podobnie jak kapitał ludzki, odnosi się do społecznego dobrobytu, lecz z perspektywy całych społeczności, a nie poszczególnych jednostek. Kapitał społeczny jest systemem relacji bazujących na zaufaniu oraz formalnych i nieformalnych zasadach, które umożliwiają ludziom lub grupom osób dostęp do zasobów. Kapitał ten budują sieci kontaktów społecznych, które wspomagają tworzenie sprawnego i spójnego społeczeństwa oraz ułatwiają społeczne i intelektualne interakcje między członkami tego społeczeństwa. Kapitał społeczny odnosi się do tych zasobów społecznego zaufania, norm i relacji, z których ludzie mogą korzystać, by rozwiązywać wspólne problemy i tworzyć spójność społeczną. Przykładowe elementy tego kapitału to: stowarzyszenia sąsiedzkie i obywatelskie oraz spółdzielnie. Polityczne i prawne struktury, które sprzyjają politycznej stabilności, demokracji, wydajnemu zarządzaniu oraz sprawiedliwości społecznej (wszystkie one są tak samo dobre dla produktywności jak i są pożądane same w sobie) są także częścią kapitału społecznego.

Progi krytyczne, trendy krytyczne, sprzeczności rozwojowe

Próg krytyczny można zdefiniować jako punkt między dwoma stanami danego układu. W momencie gdy zostanie on przekroczony, układ podlega jakościowej zmianie. Zmiana ta, zazwyczaj związana ze stratą pewnych cennych wartości układu, waloryzowana jest negatywnie oraz nieodwracalna czy też bardzo trudna (np. kosztowna) do odwrócenia (Ten Brink et al. 2008). Progi krytyczne w oczywisty sposób występują w przyrodzie i są ściśle wyrażane przez nauki biologiczne, chemiczne, fizyczne itd.) – przykładem może tu być choćby minimalne stężenie tlenu w powietrzu, potrzebne człowiekowi do życia, które można wyrazić za pomocą prostego wskaźnika ilościowego. Natomiast progi krytyczne w świecie społecznym czy gospodarczym są o wiele trudniejsze do precyzyjnego określenia i, niejako ze swej natury, mają bardziej płynny, nieostry charakter (np. można przypuszczać, że przekroczenie pewnego poziomu przestępczości może mieć znaczący negatywny wpływ na rozwój gospodarczy, jednak trudno, przynajmniej *a priori*, określić wartość tego progu).

Trend krytyczny można rozumieć jako proces (trend), który zmierza w kierunku progu krytycznego.

Sprzeczność rozwojowa (*trade-off*) to taka sytuacja, gdy rozwój (wzrost) jednego rodzaju kapitału, powoduje stratę w innym kapitale.

3.2. Jak przebiegał proces badawczy?

Badanie było realizowane w trzech głównych etapach:

- Wstępne analizy
  - Podsumowanie oceny stanu czterech kapitałów w województwie lubuskim (na podstawie wcześniejszych wyników projektu SRDTOOLS).
  - Podsumowanie oceny wpływu Strategii na stan czterech kapitałów w województwie lubuskim (na podstawie wcześniejszych wyników projektu SRDTOOLS).
  - Ocena przystawalności projektu RPO do Strategii.
  - Analiza alokacji środków w projekcie RPO w odniesieniu do stanu czterech kapitałów w regionie.
  - Opracowanie scenariusza warsztatów oraz narzędzi animujących dyskusję.
- Warsztaty
  - 1. blok warsztatowy – ocena stanu czterech kapitałów, oraz obecnych i potencjalnych trendów i progów krytycznych.
  - 2. blok warsztatowy – ocena RPO w świetle oceny stanu czterech kapitałów, oraz obecnych i potencjalnych trendów i progów krytycznych. Dyskusja na temat wyników oraz kierunki możliwych zmian w RPO.
- Podsumowanie wyników
  - Podsumowanie wyników badania i sformułowanie ich jako części raportu końcowego projektu SRDTOOLS.

3.3. Jakie były wyzwania i ograniczenia procesu?

Największym wyzwaniem w realizacji zaplanowanej procedury były warsztaty, których efektywne przeprowadzenie uzależnione było od zainteresowania i aktywności regionalnych interesariuszy. Dzięki pomocy i zaangażowaniu Urzędu Marszałkowskiego, który zajął się techniczną organizacją warsztatów oraz zaproszeniem uczestników, udało się je przeprowadzić zgodnie z pierwotnymi założeniami.

3.4. Jaką rolę odegrali poszczególni interesariusze w procesie badawczym?

Bardzo istotną cechą omawianej ewaluacji było interaktywne, partycypacyjne podejście. Badanie zostało zaprojektowane tak, że aktywny udział interesariuszy był konieczny dla jego realizacji. Kluczowym elementem były animowane przez zespół badawczy warsztaty, w których interesariusze sami oceniali RPO. Takie podejście wynikało z założenia, że to właśnie regionalni aktorzy z jednej strony najlepiej znają regionalne uwarunkowania i potrzeby, a z drugiej, że aktywne zaangażowanie ich w proces ewaluacji spowoduje, iż łatwiej przyswoją wyniki ewaluacji.

4. PRODUKTY BADANIA

4.1. Jaką wiedzę wyprodukowano w badaniu?

Realizacja poszczególnych etapów badania dostarczała cząstkowych wyników, na podstawie których prowadzone były dalsze analizy (kolejne etapy). Przyjęto zasadę, że każdy wynik cząstkowy będzie prezentowany w syntetycznej, przejrzystej i atrakcyjnej graficznie formie. Wiązało się to z potrzebą łatwej i skutecznej prezentacji wyników podczas warsztatów z udziałem interesariuszy. Należy podkreślić, że takie podejście okazało się bardzo skuteczne i pozytywnie odebrane przez interesariuszy. W dalszej części niniejszego rozdziału zostaną przedstawione wspominane wyniki poszczególnych etapów.

4.2. Etap 1. Wstępne analizy

Pierwszym krokiem było podsumowanie analiz stanu czterech kapitałów w województwie lubuskim prowadzonych w pierwszym roku realizacji projektu SRDTOOLS. Szczegółowe analizy (opisy, statystyki, wykresy) skondensowano w formie jednej prostej tabeli, odnoszącej stan czterech kapitałów w regionie do stanu czterech kapitałów w kraju i UE (tabela 1).

Następnie opracowano podsumowanie oceny wpływu Strategii Rozwoju Województwa Lubuskiego na stan czterech kapitałów. Należy podkreślić, że ocena ta także powstała przy aktywnym udziale regionalnych interesariuszy (w ramach warsztatów zorganizowanych w Urzędzie Marszałkowskim w roku 2005)[2]. Syntetyczne wyniki warsztatów przedstawia tabela 2. Ocena jest wynikiem dyskusji i uzyskanego w jej toku konsensu.

Tabela 1. Ocena stanu czterech kapitałów w województwie lubuskim

Stan zasobów / Kapitał	bardzo zły	zły	średni	dobry	bardzo dobry
gospodarczy	UE		PL		
ludzki		UE	PL		
społeczny		UE PL			
naturalny				PL	UE

Źródło: raport końcowy projektu SRDTOOLS.

[2] Szczegółowy opis tego etapu projektu SRDTOOLS jest przedstawiony w artykule: Gorzelak et al. 2006.

Tabela 2. Ocena wpływu Strategii Rozwoju Województwa Lubuskiego na stan czterech kapitałów („–" – wpyw negatywny; „·" – brak wpywu; „+" – wpyw pozytywny; „++" –wpyw bardzo pozytywny; „?" — trudno powiedzieć)

Cel/Cel operacyjny	Kapitał gospodarczy			Kapitał ludzki			Kapitał społeczny			Kapitał naturalny		
	Potencjany	Rzeczywisty	Ostrzeżenie	Potencjalny	Rzeczywisty	Ostrzeżenie	Potencjalny	Rzeczywisty	Ostrzeżenie	Potencjalny	Rzeczywisty	Ostrzeżenie
Zapewnienie przestrzennej, gospodarczej społecznej spójności regionu												
Zapewnienie sprawnego systemu komunikacyjnego	++	·	–		·		·	·		·	·	
Wykorzystanie warunków do rozwoju transportu lotniczego	++	·		·	·		·	·		·	·	
Udoskonalenie i rozbudowa infrastruktury technicznej, komunalnej i społecznej	++	+		++	+		+	·		++	++	
Wszechstronna współpraca transgraniczna i międzyregionalna	++	+		+	+		+	+		+	+	
Osiągnięcie wysokiej umiejętności korzystania ze środków Unii Europejskiej	+	+		++	++		·	·		·	·	

Tabela 2. cd.

Cel/Cel operacyjny	Kapitał gospodarczy			Kapitał ludzki			Kapitał społeczny			Kapitał naturalny		
	Potencjany	Rzeczywisty	Ostrzeżenie	Potencjalny	Rzeczywisty	Ostrzeżenie	Potencjalny	Rzeczywisty	Ostrzeżenie	Potencjalny	Rzeczywisty	Ostrzeżenie
Podniesienie poziomu wykształcenia społeczeństwa i zwiekszenie potencjału innowacyjnego nauki i gospodarki												
Podniesienie poziomu wykształcenia na poziomie ponadgimnazjalnym i wyższym	++	+		++	++		++	+		·	·	
Dostosowanie kształcenia do potrzeb rynku pracy i uwarunkowań wynikających z procesu integracji z UE	++	·		++	+		+	·		·	·	
Wyrównanie szans edukacyjnych dzieci i młodzieży	·	·		++	·	–	++	·	–	·	·	
Wzmocnienie i ustabilizowanie kadry naukowej oraz rozwój bazy naukowo-badawczej uczelni lubuskich	++	+		+	+		+	+		·	·	
Utworzenie Uniwersytetu Lubuskiego i wzmocnienie sieci szkół wyższych	+	+		++	++		+	+		·	·	
Rozwój przedsiębiorczości												
Aktywna promocja gospodarcza i pozyskiwanie inwestorów	++	+	–	+	·		+	·		·	·	

Stworzenie regionalnego systemu wsparcia innowacji i transferu technologii	++	·	–	++	·	–	+	·		·	·	
Rozwój instytucjonalnego i kapitałowego otoczenia biznesu	++	++		++	+		++	+		·	·	
Pozarolniczy rozwój terenów wiejskich	+	·	–	+	·	–	+	·		+	·	
Restrukturyzacja i reorientacja towarowych gospodarstw rolnych i zakładów przetwórstwa żywności	+	+		+	·		·	·		+	·	–
Efektywne wykorzystanie zasobów środowiska naturalnego i kulturowego												
Rozwijanie świadomosci proekologicznej	·	·		+	+		+	+		++	+	
Wykorzystanie walorów środowiska i dziedzictwa kulturowego dla rozwoju turystyki	+	·		·	·		+	·		+	·	
Skuteczna promocja walorów turystycznych i system informacji turystycznej	+	+		·	·		·	·		+	·	
Rozwój usług kulturalnych, zdrowotnych i sportowych dla mieszkańców regionu i gości zagranicznych	+	+		+	+		+	·		·	·	

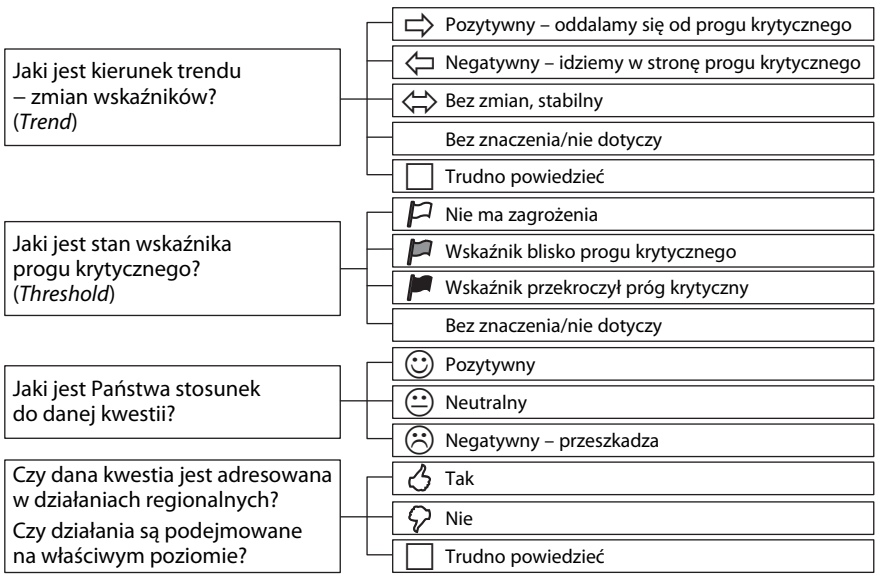
Źródło: raport końcowy projektu SRDTOOLS.

Kolejnym krokiem było sprawdzenie przystawalności projektu RPO do Strategii Rozwoju województwa. Wykorzystano w tym celu standardową matrycę przystawalności, opartą na eksperckiej analizie tekstów obu dokumentów. Wyniki tej analizy w zasadzie potwierdziły przystawalność obu dokumentów. Następnie przygotowano zestawienie alokacji środków przewidywanych w projekcie RPO. Wyniki tej części, zestawione z oceną wpływu strategii na cztery kapitały, pozwoliły wskazać obszary RPO, których realizacja jest szczególnie ważna lub zagrożona (por „ostrzeżenia" w tabeli 2), a przewidziane na nią nakłady w RPO są relatywnie małe (lub niewystarczające).

4.3. Etap 2. Warsztaty

W warsztatach zorganizowanych na terenie Urzędu Marszałkowskiego udział wzięło ok. 20 osób – marszałek województwa, dyrektorowie poszczególnych departamentów tematycznych Urzędu Marszałkowskiego, przedstawiciele partnerów społeczno-gospodarczych, przedstawiciele regionalnych agend instytucji centralnych. Pierwszym krokiem realizacji warsztatów było zaprezentowanie, omówienie i dyskusja wyników etapu 1. Następnie przeprowadzono ocenę stanu czterech kapitałów w regionie. Na potrzeby tej oceny wykorzystano 18 wskaźników przypisanych do poszczególnych kapitałów (por. rysunek 2). Podczas dyskusji interesariusze oceniali dla każdego wskaźnika:

- kierunek trendu,
- aktualny stan zjawiska w odniesieniu do progu krytycznego,



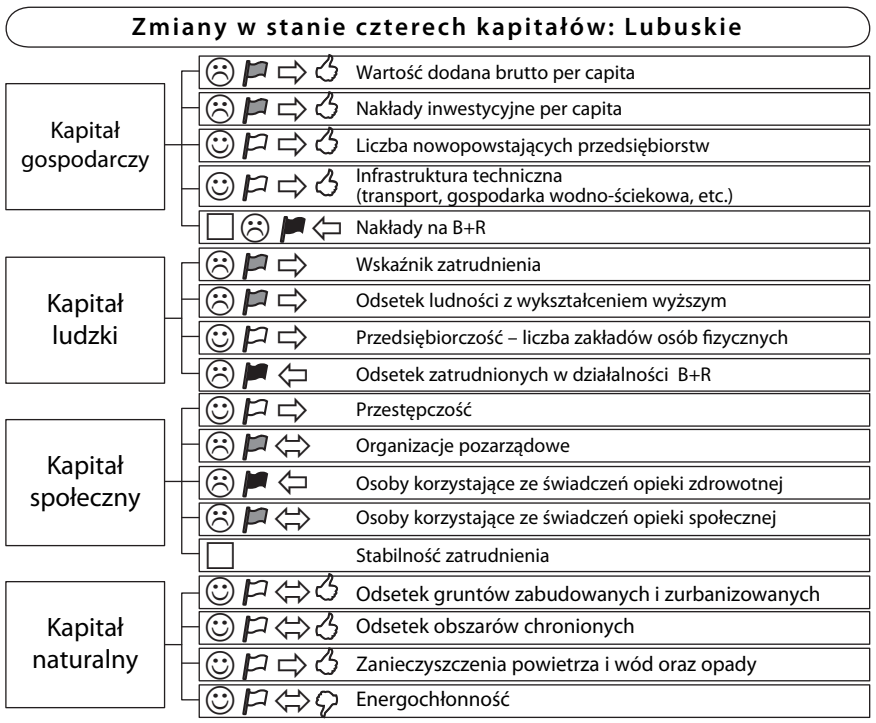
Rysunek 1. System symboli wykorzystanych do oceny stanu czterech kapitałów

Źródło: raport końcowy projektu SRDTOOLS.

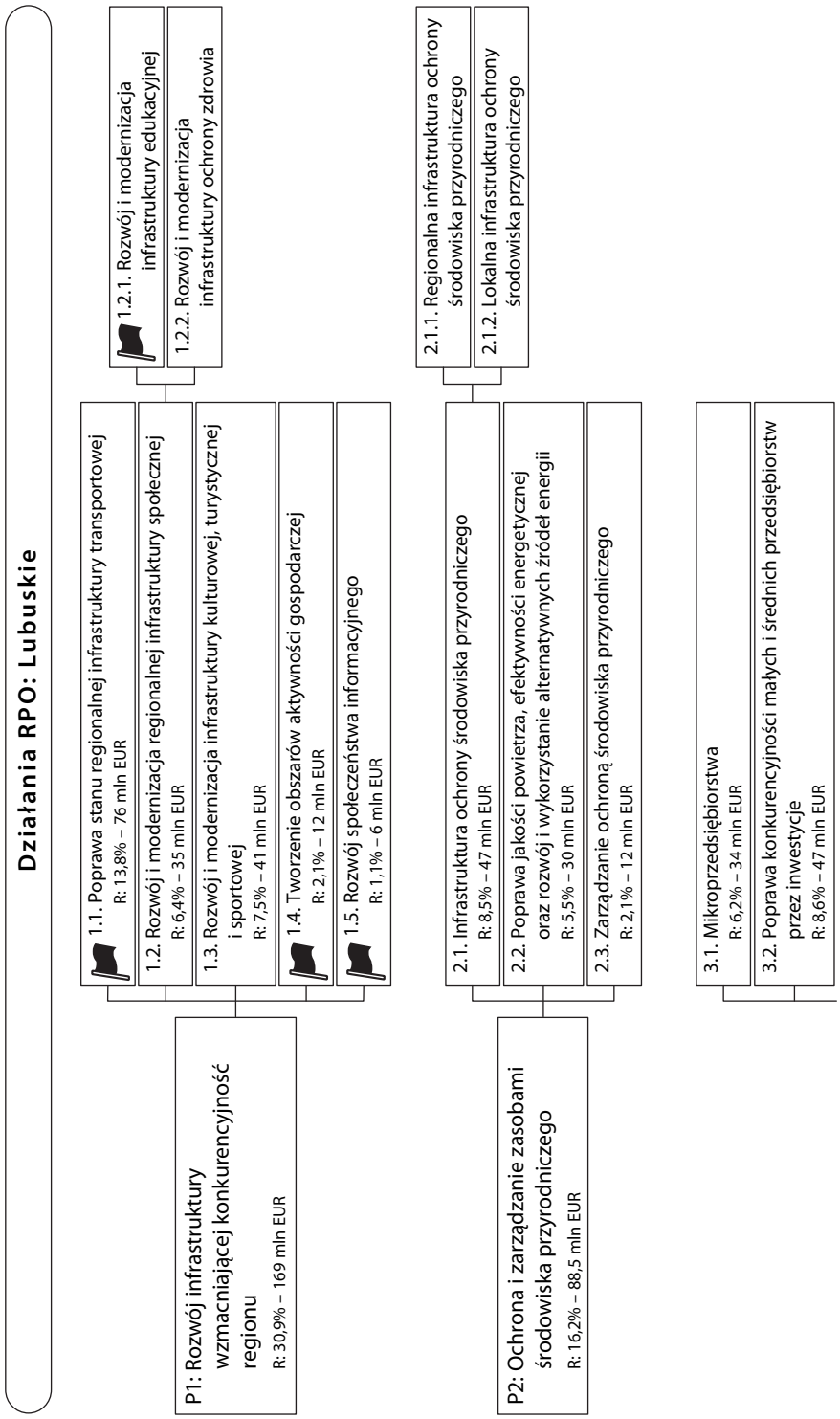
- własny (instytucjonalny) stosunek do obserwowanych zmian,
- obecność (lub nie) danego zagadnienia w działaniach podejmowanych przez władze regionalne.

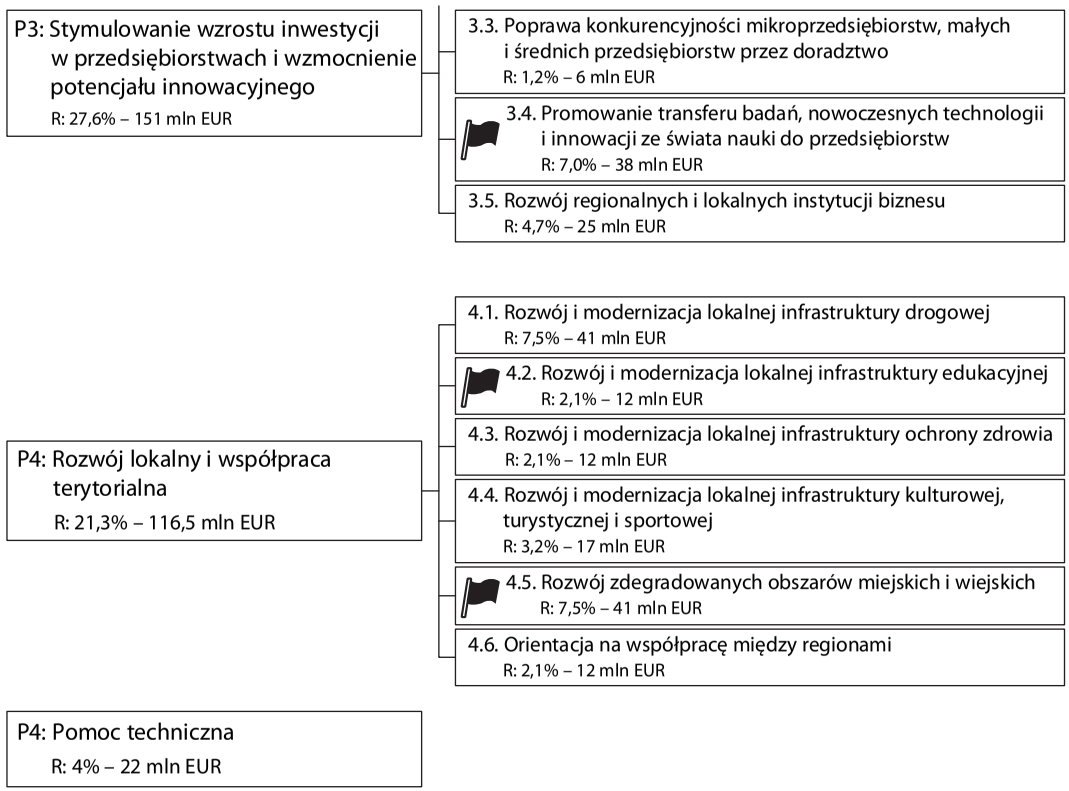
Do animowania dyskusji została wykorzystana metoda graficznych symboli (por. rysunek 1). Wyniki oceny stanu czterech kapitałów były na bieżąco wstawiane do wcześniej przygotowanego schematu (uczestnicy sami decydowali, niejednokrotnie podczas burzliwej dyskusji, jaki ostatecznie symbol przypisać danemu wskaźnikowi – rolą ewaluatorów było moderowanie tej dyskusji). Wyniki tej części badania przedstawia rysunek 2.

Ostatnim elementem warsztatów było sprawdzenie i porównanie, na ile zamierzony sposób podziału środków finansowych odpowiada zidentyfikowanym wcześniej obszarom zagrożeń. W tym celu na drzewo struktury priorytetów i działań RPO nałożono „gorące punkty" (czarne flagi) – a więc obszary problemowe, wynikające z analizy trendów i progów krytycznych oraz z wcześniejszej analizy dotyczącej aktualizacji Strategii Rozwoju Województwa Lubuskiego. Wynikiem dyskusji wśród interesariuszy moderowanej przez ewaluatorów było wskazanie tych obszarów interwencji RPO, w których należałoby zwiększyć planowaną alokację. Wyniki tej dyskusji przedstawia rysunek 3 (flagami oznaczono działania, na których realizację rekomenduje się zwiększenie planowanej alokacji).



Rysunek 2. Wyniki oceny czterech kapitałów przeprowadzonej przez interesariuszy
Źródło: raport końcowy projektu SRDTOOLS.


Działania RPO: Lubuskie
P1: Rozwój infrastruktury wzmacniającej konkurencyjność regionu
R: 30,9% – 169 mln EUR
1.1. Poprawa stanu regionalnej infrastruktury transportowej
R: 13,8% – 76 mln EUR
1.2. Rozwój i modernizacja regionalnej infrastruktury społecznej
R: 6,4% – 35 mln EUR
1.3. Rozwój i modernizacja infrastruktury kulturowej, turystycznej i sportowej
R: 7,5% – 41 mln EUR
1.4. Tworzenie obszarów aktywności gospodarczej
R: 2,1% – 12 mln EUR
1.5. Rozwój społeczeństwa informacyjnego
R: 1,1% – 6 mln EUR
1.2.1. Rozwój i modernizacja infrastruktury edukacyjnej
1.2.2. Rozwój i modernizacja infrastruktury ochrony zdrowia
P2: Ochrona i zarządzanie zasobami środowiska przyrodniczego
R: 16,2% – 88,5 mln EUR
2.1. Infrastruktura ochrony środowiska przyrodniczego
R: 8,5% – 47 mln EUR
2.2. Poprawa jakości powietrza, efektywności energetycznej oraz rozwój i wykorzystanie alternatywnych źródeł energii
R: 5,5% – 30 mln EUR
2.3. Zarządzanie ochroną środowiska przyrodniczego
R: 2,1% – 12 mln EUR
2.1.1. Regionalna infrastruktura ochrony środowiska przyrodniczego
2.1.2. Lokalna infrastruktura ochrony środowiska przyrodniczego
3.1. Mikroprzedsiębiorstwa
R: 6,2% – 34 mln EUR
3.2. Poprawa konkurencyjności małych i średnich przedsiębiorstw przez inwestycje
R: 8,6% – 47 mln EUR



Rysunek 3. Wyniki oceny czterech kapitałów przeprowadzonej przez interesariuszy

Źródło: opracowanie własne.

4.4. Etap 3. Podsumowanie wyników

Etap ten polegał na opracowaniu wyników fazy przygotowawczej oraz warsztatów w formie rozdziału raportu końcowego projektu SRDTOOLS. Materiały wypracowane w toku realizacji badania zostały przekazane do Urzędu Marszałkowskiego, tak aby mogły być wykorzystane w toku dalszych prac nad RPO.

Wiedza wyprodukowana w badaniu dostarczyła z jednej strony dość szczegółowego opisu badanego regionu oraz jego dokumentów strategicznych, z drugiej natomiast mogła być łatwo wykorzystana do poprawy efektywności interwencji planowanych w RPO. Szczególnie istotnym aspektem badania było to, że jego wyniki w znacznej mierze wytworzyli interesariusze z regionu w toku dyskusji podczas warsztatów, co sprzyjało ich zrozumieniu, akceptacji oraz dużej chęci uwzględnienia w toku dalszych prac nad RPO.

W odniesieniu do kontekstu projektu SRDTOOLS należy natomiast powiedzieć, że przeprowadzone badanie spełniło pokładane w nim oczekiwanie, tzn. umożliwiło przetestowanie analizowanych metod i narzędzi, które okazały się w tym przypadku skuteczne oraz pozytywnie odbierane przez interesariuszy.

4.5. Jak wyglądał proces rozpowszechniania wyników?

Najważniejszym kanałem rozpowszechniania wyników badania były omawiane warsztaty. Ponadto szczegółowy opis badania zawarty został w rozdziale raportu końcowego projektu SRDTOOL (dostępnego w języku angielskim) oraz w kilku artykułach opublikowanych w czasopismach naukowych lub jako rozdziały publikacji zwartych (por. przede wszystkim Gorzelak et al. 2007; 2008). Wyniki badania prezentowane były także podczas otwartego seminarium naukowego przeprowadzonego na Uniwersytecie Warszawskim na początku 2007 roku oraz na seminarium naukowym (Institutional Analysis & Development Seminar) w Indiana University (Bloomington) w USA pod koniec 2006 roku.

5. WNIOSKI ZE STUDIUM PRZYPADKU

5.1. Jakie były efekty, rzeczywista użyteczność badania?

Początkowe założenia badawcze zostały w pełni zrealizowane. Wszystkie etapy badania zrealizowane zostały terminowo i z oczekiwanymi rezultatami. Wybór Regionalnego Programu Operacyjnego do testowania analizowanych w projekcie metod i narzędzi okazał się słuszny. Tak jak zakładano, przeprowadzona ewaluacja *ex ante* RPO pozwoliła

wszechstronnie przetestować wybrane metody i narzędzia, dowodząc ich poprawności metodologicznej, skuteczności oraz atrakcyjności dla interesariuszy. Testowane metody okazały się użyteczne przede wszystkim w kontekście weryfikacji *a priori* programów operacyjnych pod kątem:

- spójności zewnętrznej (ang. *external coherence*), a więc przystawania do innych kluczowych dokumentów strategicznych, całościowej wizji rozwoju regionalnego,
- spójności wewnętrznej (ang. *internal coherence*), a więc współgrania ze sobą celów i działań w ramach badanego dokumentu,
- trafności (ang. *relevance*) tego, na ile cele programu odpowiadają zidentyfikowanym potrzebom, i czy adresują główne siły i słabości danego układu.

Omawiane badanie ewaluacyjne spełniło przede wszystkim funkcję racjonalizacji planowania – wypracowane w jego toku rezultaty zostały w pewnej mierze wykorzystane w dalszych pracach nad RPO. Równie istotną funkcją było budowanie partnerstwa i współwłasności – tutaj szczególnie ważną rolę odegrały przeprowadzone warsztaty. Natomiast w kontekście projektu SRDTOOLS najważniejsza była funkcja poznawcza – ewaluację wykorzystano do testowania metod i narzędzi oceny trwałości rozwoju regionalnego analizowanych w tym projekcie.

5.2. Jakie były silne i słabe strony analizowanej ewaluacji?

Niewątpliwą zaletą wykorzystanego podejścia jest interaktywność, oparcie na konsultacjach, partycypacyjny charakter. Wykorzystanie w ewaluacji moderowanej dyskusji podczas warsztatów pozwoliło zaktywizować interesariuszy oraz zwiększyć ich identyfikację z tworzonymi planami. Metoda wpisuje się więc dobrze w nowy model badań ewaluacyjnych, podkreślających konieczność włączania interesariuszy w proces ewaluacji, a przede wszystkim w sam moment oceny (tzw. podejście partycypacyjne). Takie podejście gwarantuje szybkie przełożenie wniosków z ewaluacji na proces decyzyjny (zmiany w programie), co sprawdziło się również w badanym przypadku.

Testowana metoda okazała się także przydatna do pokazaniu interesariuszom nieco innej, szerszej perspektywy programu oraz zwiększenia ich świadomości idei rozwoju trwałego.

Od strony technicznej, ustrukturalizowanie wymiany poglądów wokół czterech kapitałów okazało się skuteczną i jasną metodą dyskusji grupowej, weryfikacji dokumentów i podejmowania decyzji zespołowych w oparciu o konsensus.

Stosowanie metod wykorzystanych w omawianej ewaluacji i prezentowanych w niniejszym rozdziale niesie za sobą pewne wymogi

i obostrzenia. Przede wszystkim należy pamiętać, że jest to metoda prospektywna. Użycie jej do oceny efektów już wdrażanych strategii, wymaga dalszych modyfikacji i połączenia z innymi metodami oceny (ocena przez beneficjentów, panel ekspertów, analiza danych regionalnych).

Zastosowanie prezentowanych tu metod wymaga przede wszystkim:

- wypracowania jasnej relacji między kapitałami, wskaźnikami poszczególnych kapitałów a celami i działaniami ocenianych dokumentów strategicznych i programowych,
- ograniczenia liczba dyskutowanych wskaźników i tematów (do 20),
- użycia jasnego, przejrzystego (najlepiej w formie graficznej) systemu oceny (w przypadku tego badania sprawdził się system ikon – czerwonych, zielonych kartek i symboli), tak by całościowa ocena był łatwa do interpretacji przez uczestników panelu,
- połączenia z dobrą, kompetentną animacją dyskusji (w tym z przygotowaniem przeglądowych materiałów kontekstowych i porównawczych, przedstawiających sytuację w danym regionie w perspektywie ogólnokrajowej czy europejskiej).
- Podczas prowadzenia badania zaobserwowano również, że interesariusze mają trudności z jednoznacznym oszacowaniem trendów i progów krytycznych w kapitałach innych niż naturalny. Ocena jest tu bardziej względna, a punkt krytyczny zależy od percepcji poszczególnych aktorów. Ten aspekt wymaga szczególnej uwagi i ostrożności podczas realizacji badań ewaluacyjnych wykorzystujących koncepcję trendów i progów krytycznych.

BIBLIOGRAFIA

Ekins P. (1992), *A four-capital model of wealth creation*, w: Ekins P., Max-Neef M. (eds.), *Real-life economics: understanding wealth creation*, London, New York, Routledge.

Ekins P. (2008a), Introduction to the Special Issue, *European Environment* **18**, 59–62.

Ekins P. (2008b), The Four-Capital Method of Sustainable Development Evaluation, *European Environment* **18**, 63–81.

Gorzelak G., Płoszaj A., Smętkowski M. (2006), Ocena strategii rozwoju regionu – wykorzystanie modelu czterech kapitałów, *Studia Regionalne i Lokalne* **3(25)**, 67–82.

Gorzelak G., Olejniczak K., Płoszaj A., Smętkowski M. (2008), Applying the Four-Capital Model to the Assessment of Regional Development in the Lubuskie Region of Poland, *European Environment* **18**, 110–122.

Gorzelak G., Olejniczak K., Płoszaj A., Smętkowski M. (2007), *Building Regional & Local Strategies for Sustainable Development – A Participative Approach*, referat prezentowany przez K. Olejniczaka na: *Institutional Analysis & Development Seminar Indiana University*, Bloomington, US, 11.12.2006.

Płoszaj A. (2007), *„Cztery kapitały" w strategiach lokalnych*, w: Gorzelak G. (2007) *Polska regionalna i lokalna w świetle badań EUROREG-u*, Warszawa, Wydawnictwo Naukowe Scholar.

Ten Brink P., Miller C., Kettunen M., Ramsak K., Farmer A., Hjerp P., Anderson J. (2008), Critical Thresholds, Evaluation and Regional Development, *European Environment* **18**, 81–95.

ROZDZIAŁ 3.5.

Agnieszka Sternik

EWALUACJA *EX ANTE* PROGRAMU OPERACYJNEGO „ZASOBY LUDZKIE I ZATRUDNIENIE 2007–2013 W CZECHACH"

Niniejszy rozdział prezentuje studium przypadku ewaluacji *ex ante* Programu Operacyjnego *Zasoby Ludzkie i Zatrudnienie 2007–2013* (dalej jako PO ZLZ) finansowanego ze środków Europejskiego Funduszu Społecznego (EFS) w Republice Czeskiej. Powyższy przypadek wybrany został spośród wielu ewaluacji *ex ante* programów operacyjnych przeprowadzonych przez kraje członkowskie w okresie 2006–2007 ze względu na interaktywność procesu badawczego oraz znaczące wykorzystanie potencjału, jakie tego typu ewaluacja niesie za sobą w procesie planowania wydatkowania funduszy publicznych.

Badanie rozpoczęło się w styczniu 2006 roku, pół roku po zainicjowaniu prac nad PO ZLZ. Ostateczna wersja raportu ewaluacyjnego została przyjęta w październiku 2006, wraz z zakończeniem prac nad Programem i przedłożeniem go oficjalnie do Komisji Europejskiej. Głównym odbiorcą ewaluacji było Ministerstwo Pracy i Polityki Społecznej, Departament Zarządzania EFS, pełniące funkcję Instytucji Zarządzającej PO ZLZ. Wykonawcą badania była firma konsultingowa DHV CR wraz z zespołem zewnętrznych ekspertów, głównie badaczy akademickich, zatrudnionych w celu prowadzenia analizy poszczególnych obszarów

polityki PO ZLZ. Badanie zostało sfinansowane ze środków pomocy technicznej przeznaczonych na wsparcie realizacji PO Rozwoju Zasobów Ludzkich 2004–2006.

1. KONTEKST INSTYTUCJONALNY I PRZEDMIOT REALIZACJI EWALUACJI

Od wstąpienia do Unii Europejskiej w maju 2004 roku Republika Czeska korzysta ze wsparcia Funduszy Strukturalnych i Funduszu Spójności. W poprzednim okresie programowania Europejski Fundusz Społeczny wdrażany był za pomocą dwóch programów operacyjnych: PO Rozwój Zasobów Ludzkich (cel 1), Jednolity Dokument Programowy dla Pragi (cel 3) oraz poprzez Inicjatywę Wspólnotową EQUAL. Implementacja powyższych programów stanowiła cenne doświadczenie, wykorzystane na etapie przygotowywania PO ZLZ. Do najważniejszych błędów poprzedniego okresu programowania, zidentyfikowanych dzięki prowadzonej wówczas ewaluacji *on-going*, należały: powolny start wdrażania programów spowodowany niedoświadczeniem pracowników administracji publicznej oraz brakiem przygotowania potencjalnych beneficjentów do aplikowania o środki, skomplikowanie systemu monitorowania, niezadowalająca jakość składanych projektów oraz brak komplementarności i synergii pomiędzy poszczególnymi programami krajowymi.

Bazując na tych doświadczeniach oraz pogłębionej analizie sytuacji społeczno-gospodarczej, Ministerstwo Pracy i Polityki Społecznej przygotowało PO *Zasoby Ludzkie i Zatrudnienie 2007–2013*. Celem nadrzędnym PO ZLZ jest podniesienie stopy zatrudnienia do wysokości średniej z 15 krajów o najwyższej wartości tego wskaźnika. PO ZLZ definiuje pięć głównych priorytetów:

- Wzrost adaptacyjności pracowników i pracodawców.
- Poprawa dostępu do zatrudnienia i przeciwdziałanie bezrobociu.
- Wzmocnienie integracji osób wykluczonych i zagrożonych wykluczeniem społecznym.
- Wzmocnienie zdolności administracyjnych i skuteczności sektora publicznego.
- Rozwój współpracy transnarodowej w obszarze rozwoju zasobów ludzkich i zatrudnienia.

Budżet PO ZLZ na lata 2007–2013 przewiduje ponad 2 miliardy EUR, z czego 85% stanowią środki EFS, a 15% wkład z budżetu narodowego.

Celem ewaluacji *ex ante* PO ZLZ była analiza programu prowadząca do bezpośredniego podniesienia jakości dokumentu programowego oraz pośredniego wpłynięcia na efektywność i spójność planowanych interwencji, przez zoptymalizowanie budżetu, celów i planowanych rezultatów. Przede wszystkim jednak, przeprowadzenie ewaluacji *ex ante* było wymogiem formalnym Komisji Europejskiej zapisanym w, wówczas jeszcze będącym w fazie projektu, rozporządzeniu Rady, ustanawiającym

ogólne przepisy dla Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego i Funduszu Spójności[1]. Zgodnie z rozporządzeniem odpowiedzialność za przeprowadzenie ewaluacji ponosiła instytucja zajmująca się przygotowaniem programu operacyjnego w kraju członkowskim. Wyniki ewaluacji *ex ante* miały być włączone do dokumentu programowego w postaci odrębnego załącznika i przekazane do Komisji Europejskiej.

Warto zauważyć, że Rozporządzenie 1081/2006 wprowadziło bardziej strategiczne podejście do planowania i wykorzystywania środków z funduszy strukturalnych, kładąc większy nacisk na rzeczywiste rezultaty interwencji w latach 2007–2013. Powyższa zmiana wynikała w głównej mierze z potrzeby wprowadzenia skuteczniejszych działań w celu osiągnięcia szerszych celów Unii Europejskiej zapisanych w Strategii Lizbońskiej. Komisja Europejska postrzegała ewaluację *ex ante* jako punkt wyjścia dla całego procesu wykorzystania przyznanych środków oraz wypracowania skutecznych metod oceny i systemu wskaźników pozwalających mierzyć ilościowe oraz jakościowe efekty programów.

2. TYP EWALUACJI

Zgodnie z zamówieniem badanie było ewaluacją *ex ante*. Jej celem było zweryfikowanie spójności interwencji, jej przystawalności do zidentyfikowanych problemów, tego, czy oczekiwane oddziaływanie jest realistyczne, a także oszacowanie rezultatów oraz zapewnienie jasnych i kwantyfikowalnych celów. Ewaluacja prowadzona była równolegle do procesu przygotowywania PO ZLZ. Wyniki cząstkowe ewaluacji, były wykorzystywane na poszczególnych etapach przygotowania Programu. Na zakończenie procesu przedstawiono raport końcowy z ewaluacji, zawierający podsumowanie procesu badawczego, głównych wniosków i rekomendacji oraz opis, w jakim stopniu wcześniejsze rekomendacje zostały uwzględnione przez osoby odpowiedzialne za przygotowanie Programu.

3. ZAKRES BADANIA I PYTANIA EWALUACYJNE

Ewaluacja *ex ante* jest, obok ewaluacji *ex post*, badaniem, którego przeprowadzenie oraz zakres zostały prawnie określone w Rozporządzeniu 1081/2006. Fakt ten, oraz istnienie stosunkowo szczegółowych wytycznych Komisji Europejskiej, w postaci dokumentu roboczego nr 2 dt.

---

[1] Rozporządzenie (WE) Nr 1081/2006 Parlamentu Europejskiego i Rady z dnia 5 lipca 2006 roku w sprawie Europejskiego Funduszu Społecznego i uchylające rozporządzenie (WE) nr 1784/1999.

ewaluacji *ex ante*[2] spowodowały, iż ewaluacja *ex ante* przebiega w sposób stosunkowo podobny we wszystkich krajach członkowskich, które realizują interwencje współfinansowane z funduszy strukturalnych. Celem nadrzędnym ewaluacji *ex ante* jest „optymalizacja podziału środków budżetowych w ramach programów operacyjnych oraz poprawa jakości programowania" (artykuł 46(2))[3].

Ewaluacja *ex ante* PO *Zasoby Ludzkie i Zatrudnienie* zarówno ze względu na sposób przebiegu procesu badawczego, jak i swój zakres odpowiadała zapisom Rozporządzenia oraz dokumentu roboczego nr 2 KE. Badanie podzielone było na trzy fazy; do każdej z nich Zamawiający sformułował listę pytań, stanowiących jednocześnie kryteria ewaluacyjne. Najważniejsze z nich zostały przedstawione poniżej.

Pierwszy etap badania skupił się na ocenie jakości analizy kontekstowej zaprezentowanej w Programie. Głównymi kryteriami ewaluacyjnymi były:

- adekwatność analizy społeczno-gospodarczej do zakresu Programu,
- spójność analizy Programu z analogiczną analizą przeprowadzoną w Narodowych Strategicznych Ramach Odniesienia (NSRO),
- właściwość zidentyfikowanych problemów i potrzeb oraz jakość analizy SWOT.

Na tym etapie uwagi wykonawców dotyczyły niewystarczającego analitycznego podejścia do danych zastanych i zbyt opisowego charakteru analizy, która zawierała również elementy opisu strategii Programu. Odrębny zestaw rekomendacji został przygotowany dla analizy SWOT w pierwszej wersji PO ZLZ, która zdaniem wykonawców była niespójna, zawierała zbyt rozbudowane wyjaśnienia rezultatów oraz niejednokrotnie raczej banalne wnioski. Na podstawie powyższych uwag, Zamawiający dokonał modyfikacji dokumentu, uwzględniając zdecydowaną większość z przedstawionych uwag. Wyniki analizy tej fazy zostały przedstawione w pierwszym z dwóch raportów cząstkowych przedłożonych przez wykonawcę.

Druga faza badania poświęcona była ocenie spójności wewnętrznej oraz zewnętrznej Programu oraz ewaluacji adekwatności strategii wobec zidentyfikowanych w analizie kontekstowej problemów. Jednymi z kilkunastu kryteriów ewaluacyjnych były:

- adekwatność strategii Programu wobec zidentyfikowanych problemów i potrzeb, przejrzystość zdefiniowania celów,
- komplementarność i synergia poszczególnych priorytetów dla osiągnięcia nadrzędnych celów Programu,
- spójność Programu z NSRF i innymi dokumentami programowymi oraz strategiami krajowymi,

---

[2] *The New Programming Period*, 2007–2013: Methodological Working Papers, Working Paper No. 2 on Ex Ante Evaluation, European Commission, 2005.

[3] Rozporządzenie (WE) nr 1081/2006.

- adekwatność strategii dla osiągania celów określonych w Strategii Lizbońskiej, innych polityk wspólnotowych oraz podstawowych zasad wspólnotowych, m.in. partnerstwa, równych szans kobiet i mężczyzn, antydyskryminacji i współpracy transnarodowej,
- właściwość alokacji zasobów finansowych pomiędzy poszczególne priorytety z punktu widzenia zidentyfikowanych problemów i potrzeb.

W drugim raporcie cząstkowym wykonawcy podważyli sposób sformułowania celu głównego Programu, jako metodologicznie błędnego, tj. definicja wykraczała poza obszar, w którym można było zarejestrować bezpośrednie skutki oddziaływania Programu. Także definicje celów szczegółowych, zdaniem wykonawców, były zbyt opisowe, często przyjmowały raczej charakter opisu działań możliwych do sfinansowania ze środków Programu. W kwestii spójności zewnętrznej PO ZLZ wykonawcy zwrócili uwagę na brakujący opis komplementarności Programu wobec konkretnych programów krajowych z tego samego obszaru. Także w tym przypadku, wiele z rekomendacji wykonawców zostało wdrożonych przez Zamawiającego już na etapie przygotowywania kolejnej wersji dokumentu.

Ostatnia faza badania dotyczyła oceny przyjętego systemu zarządzania Programem, ze szczególnym uwzględnieniem systemu monitorowania. Zakres badawczy podzielono na trzy grupy tematyczne, wymienione poniżej wraz z głównymi kryteriami ewaluacyjnymi:

- Ewaluacja przewidywanych bezpośrednich i pośrednich efektów programu:
  - zachowanie logicznego związku pomiędzy przewidywanymi działaniami a zdefiniowanymi wskaźnikami produktu, rezultatu i wpływu,
  - właściwość założonych wartości docelowych wskaźników.
- Ewaluacja systemu wdrażania Programu:
  - przejrzystość, wydajność i adekwatność zaproponowanego systemu wdrażania Programu,
  - antycypacja obszarów zagrożenia w systemie wdrażania Programu,
  - adekwatność i wydolność systemu monitorowania Programu.
- Ewaluacja wykorzystania wniosków płynących z poprzedniego okresu programowania
  - stopień wykorzystania wiedzy dostarczonej przez ewaluacje programów z lat 2004–2006.

Rezultaty tej części ewaluacji zostały włączone bezpośrednio do pierwszej wersji raportu końcowego. Ewaluatorzy zwrócili uwagę na braki w zaproponowanym systemie monitorowania, wynikające głównie z niedoskonałości wskaźników. Jednakże w związku z tym, iż do opracowania wskaźników został powołany odrębny zespół ekspertów, temat ten był tylko w ograniczony sposób dyskutowany z Zamawiającym.

Szczegółowy zestaw metod zaplanowanych do wykorzystania przy realizacji poszczególnych komponentów ewaluacji *ex ante* został zaproponowany przez wykonawców. Opierały się one wyłącznie na analizie danych zastanych, zwłaszcza danych pochodzących ze źródeł statystyki publicznej, dokumentów programowych i strategii zarówno krajowych, jak i wspólnotowych, a także, jedna w ograniczonym zakresie, z wyników ewaluacji przeprowadzonych w obszarze EFS w okresie 2004–2006.

4. REALIZACJA BADANIA

W odpowiedzi na ogłoszenie przetargowe zamieszczone, w grudniu 2005 roku, oferty badawcze zgłosiło tylko dwóch wykonawców. Jedną z przyczyn, którymi pracownicy Ministerstwa tłumaczą nikłe zainteresowanie tym konkretnym przetargiem, była specyfika badania, odbiegająca zakresem od typowych ewaluacji programów. Brak pracy w terenie i wyłączność pracy z danymi zastanymi powodowały, iż badanie mogło się wydawać nieco abstrakcyjne, a przez to mało atrakcyjne dla potencjalnych wykonawców. Należy zauważyć, że budżet badania nie stanowił czynnika wpływającego na decyzję o przedłożeniu oferty, gdyż zarówno jego wartość, jak i relatywne odniesienie do zakresu badania były porównywalne z innymi ewaluacjami. Ze względu na obowiązujące w Czechach prawo zamówień publicznych, które nakazuje anulowanie przetargu w przypadku wpłynięcia mniej niż trzech ofert, Zamawiający musiał zwrócić się do Sekretarza Stanu w Ministerstwie Pracy i Polityki Społecznej o wydanie specjalnej zgody na kontynuowanie wyboru wykonawcy. Po jej uzyskaniu, wybrana została oferta czeskiej firmy konsultingowej DHV CR, współpracującej z zespołem ekspertów akademickich.

Niewątpliwie jednym z ciekawszych elementów badania, było to, iż projektowanie ewaluacji, wybór wykonawcy, oraz późniejsza współpraca z wykonawcą zarządzane były wspólnie przez pracownika odpowiedzialnego za ewaluację EFS oraz osobę z zespołu przygotowującego Program. Podejście to umożliwiło efektywną komunikację oraz systematyczną wymianę pomiędzy Wykonawcą a bezpośrednim odbiorcą wyników ewaluacji, czyli zespołem programującym PO ZLZ. Dodatkowo, mimo że nie powołano odrębnej grupy sterującej ewaluacją dla tego kontraktu, założenia do badania, a następnie raporty, były dyskutowane na spotkaniach Grupy Sterującej Ewaluacją EFS. Forum to gromadzi 20 osób, w tym przedstawicieli instytucji publicznych wdrażających EFS oraz akademików i praktyków ewaluacji.

Ewaluacja *ex ante* rozpoczęła się w pół roku po rozpoczęciu prac nad PO ZLZ, dzięki czemu zaczynając swoje analizy, Wykonawca posługiwał się już wstępnym dokumentem programowym. Dalsze etapy pracy nad oceną zostały dostosowane do powstających równolegle elementów programu. Ewaluacja miała w pełni interaktywny i interakcyjny charakter

dzięki bliskiej współpracy między Wykonawcą a Zamawiającym. Interaktywność procesu polegała na tym, iż wykonawcy realizowali poszczególne komponenty ewaluacji w miarę postępu prac nad Programem. Jak pokazała praktyka, osoby odpowiedzialne za programowanie były zainteresowane bliską współpracą z ewaluatorami. Poza regularnymi spotkaniami, wykonawcy informowani byli drogą elektroniczną o zasadniczych zmianach wprowadzanych do Programu, zwłaszcza tych dotyczących modyfikacji zakresu priorytetów oraz alokacji środków finansowych, co umożliwiło im w sposób systematyczny prowadzenie analizy dokumentu. Interakcyjność procesu badawczego przejawiała się tym, iż treść Programu była wielokrotnie rewidowana pod kątem uwag zgłaszanych przez Wykonawcę oraz rekomendacji formułowanych w raportach cząstkowych i wstępnych wersjach raportu końcowego.

Podstawową trudnością tego badania, wpisaną niejako w charakter samego procesu ewaluacji *ex ante*, była konieczność radzenia sobie przez Wykonawcę z nieustannie zmieniającym się przedmiotem badania, czyli Programem. Zamawiający wybrał rekomendowany przez KE sposób prowadzenia ewaluacji, równolegle do procesu programowania, a tym samym wprowadził konieczność ciągłej aktualizacji wyników badania oraz dostosowywania wniosków z prowadzonej analizy i uaktualniania rekomendacji przez Wykonawcę. Miało to również praktyczne przełożenie na termin ukończenia ewaluacji, ponieważ ze względu na przedłużający się proces przygotowywania OP ZLZ konieczne okazało się przesunięcie terminu przedłożenia raportu końcowego oraz przedłużenie kontraktu z DHV CR.

Dodatkowym ograniczeniem, które zaistniało w trakcie procesu ewaluacji była konieczność zmiany zakresu badawczego, będąca konsekwencją formalnych zmian w systemie zarządzania EFS w Czechach. W poprzednim okresie programowania, Instytucją Zarządzającą dla całego Funduszu było Ministerstwo Pracy i Polityki Społecznej i to ono podjęło się przeprowadzenia wspólnej ewaluacji *ex ante* dla dwóch programów finansowanych z EFS: PO *Zasoby Ludzkie i Zatrudnienie* oraz PO *Praga – Adaptacyjność*. W połowie 2006 roku podjęta jednak została decyzja o przekazaniu zarządzania PO *Praga – Adaptacyjność* Urzędowi Miejskiemu w Pradze. Konsekwencją podjęcia tej politycznej decyzji była konieczność podziału kontraktu na dwie części, wprowadzenie zmian zakresu badania oraz zawarcie nowych umów z Wykonawcą.

5. REZULTATY BADANIA

Ze względu na charakter ewaluacji *ex ante*, głównym rezultatem badania był sam proces jego prowadzenia; mniejsze znaczenie miał zaś raport końcowy. Dużym sukcesem ewaluacji *ex ante* PO ZLZ było zapewnienie interaktywności procesu badawczego. Jednocześnie, nie miała

miejsca sytuacja, jak w przypadku ewaluacji *ex ante* czeskiego NSRF 2007–2013, gdzie wskutek bardzo bliskiej współpracy doszło do zatarcia naturalnej granicy pomiędzy Zamawiającym i Wykonawcą, i ten pierwszy wręcz zaproponował, aby ewaluator przygotował fragmenty NSRO, do których miał najbardziej szczegółowe uwagi.

Treść raportu końcowego z tego badania, a zwłaszcza jego pozytywny wydźwięk mogą być w pierwszej chwili zaskakujące dla czytelnika. Ocena Programu, jako całości oraz poszczególnych komponentów, które podlegały analizie, wydaje się praktycznie bezkrytyczna, pozostawiając niewiele rekomendacji dla dalszych zmian w dokumencie. Na tej podstawie wydawać by się mogło, iż ewaluacja przeprowadzona była wyłącznie ze względu na konieczność wypełnienia formalnych wymagań Komisji Europejskiej. Tymczasem, kluczem dla zrozumienia wyników badania jest interaktywność procesu badawczego. Raport końcowy zawiera podsumowanie procesu badawczego, głównych wniosków i rekomendacji. Celem raportu końcowego nie była bowiem próba wpłynięcia na treść PO ZLZ, a raczej przedstawienie przez Wykonawcę jego opinii na temat rodzaju, sposobu i zasadności wyborów dokonywanych w trakcie przygotowywania Programu. Warto przypomnieć, iż w trakcie procesu badawczego Wykonawca przedstawił dodatkowo dwa raporty cząstkowe oraz dwa wstępne raporty końcowe. Każdy z tych raportów odzwierciedlał zmiany wprowadzane do Programu oraz formułował konkretne rekomendacje. Proces przygotowywania tych raportów był otwarty na zmiany dokonywane w PO ZLZ.

Zgodnie z warunkami postawionymi przez Zamawiającego raport końcowy z ewaluacji został przygotowany w językach czeskim oraz angielskim i opublikowany na stronach Ministerstwa. Dokument został również przekazany do Komisji Europejskiej; wypełniono więc tym samym wymogi Rozporządzenia 1081/2006. Główne wnioski z raportu zostały włączone w formie oddzielnego rozdziału w PO ZLZ, gdzie poza oceną programu i rekomendacjami sformułowanymi przez Wykonawcę znalazło się uzasadnienie dla przyjęcia bądź odrzucenia części tych rekomendacji.

6. WNIOSKI ZE STUDIUM PRZYPADKU

O ile zaletą ewaluacji *ex ante* PO ZLZ była interaktywność i interacyjność procesu badawczego, o tyle słabą stroną badania okazał się brak wykorzystania przez Wykonawcę wyników przeprowadzonych ewaluacji programów finansowanych ze środków EFS w okresie 2004–2006. Pomimo braku ewaluacji *mid-term* i uaktualnienia tej ewaluacji prowadzonej przez kraje UE15 w 2005 roku, Czechy zrealizowały wiele badań tematycznych dotyczących wdrażania EFS, których wyniki nie zostały w dostateczny sposób wykorzystane przez wykonawców.

W ocenie zarówno Wykonawcy, jak i Zamawiającego zrealizowana ewaluacja *ex ante* wpłynęła w sposób zasadniczy na ostateczną strukturę i treść Programu Operacyjnego *Zasoby Ludzkie i Zatrudnienie 2007–2013*. Wydaje się, że ewaluacja przysłużyła się do lepszego zdefiniowana celów Programu, poprawy jakości analitycznej strony dokumentu, zwłaszcza w obszarze identyfikacji potrzeb. Przysłużyła się również do zwiększenia spójności dokumentu, ulepszyła jego logikę oraz wpłynęła na bardziej strategiczne podejście do programowania wsparcia ze środków EFS. Co więcej, porównując obecną ewaluację *ex ante*, z podobnym ćwiczeniem przeprowadzonym w poprzednim okresie programowania dla PO Rozwoju Zasobów Ludzkich 2004–2006, widać zasadniczą zmianę podejścia do procesu badawczego. Przeważająca część ewaluacji 2004–2006 przeprowadzona była wewnętrznie, przez pracowników Ministerstwa. W obszarach, gdzie dodatkowe analizy realizowane były przez ekspertów zewnętrznych, zabrakło zarówno strategicznego podejścia, jak i interaktywności z Zamawiającym.

Trudno jednak ocenić długoterminowe znaczenie tej ewaluacji. Wydaje się, że mając wynegocjowany Program z Komisją Europejską, zabraknie motywacji, aby w przyszłości sięgnąć po wyniki badania. Tymczasem jedną z rekomendacji formułowanych przez KE dla procesu ewaluacji *ex ante* było wykorzystanie jej jako źródła informacji na temat obszarów, gdzie zidentyfikowano potrzebę przeprowadzenia dalszych, pogłębionych analiz, wskazując tym samym tematy potencjalnych ewaluacji *on going*, formułując pytania badawcze oraz wskaźniki. Można powątpiewać, czy ta ewaluacja *ex ante* stanie się punktem odniesienia dla kolejnych badań, prowadzonych już w ramach ewaluacji *on-going* Programu Operacyjnego *Zasoby Ludzkie i Zatrudnienie 2007–2013*.

CZĘŚĆ CZWARTA

BADANIE EWALUACYJNE – POMOC W ZARZĄDZANIU I BUDOWANIU PARTNERSTWA

ROZDZIAŁ 4.1.

Bartosz Ledzion

WPROWADZENIE

Niniejsza część podręcznika zawiera studia przypadków badań ewaluacyjnych, które wspierały w mniej lub bardziej udany sposób:

- zarządzanie programami, to znaczy pomagały w procesach wdrażania, pomiarze bieżącej jakości, skuteczności i efektywności programu, usuwaniu blokad;
- budowanie partnerstwa wśród interesariuszy programów, czyli wzmocnianie powiązań oraz interakcji między uczestnikami badanych projektów/programów.

Pierwsze studium przypadku dotyczy ewaluacji okresowej Programu Środki Przejściowe (*Transition Facility*). Przykład ten został wybrany ze względu swój zdecydowanie odmienny – od innych badań tego typu prowadzonych w polskich instytucjach – charakter. Po pierwsze, badania prowadzone w ramach projektu, miały wymiar *stricte* eksperckie – w każdą ewaluację zaangażowani byli eksperci, bez których wiedzy przygotowanie badań byłoby prawdopodobnie niemożliwe. Po drugie, proces ewaluacji, procedury, wyniki, a nawet struktura raportów z ewaluacji, zostały określone w *Podręczniku ewaluacji okresowej Programu Phare i Transition Facility (Środki Przejściowe) w Polsce*.

Drugi artykuł poświęcony jest omówieniu badania ewaluacyjnego *Identyfikacja barier w realizacji działań współfinansowanych ze środków Europejskiego Funduszu Społecznego w ramach Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego*. Opisany przypadek dobrze ilustruje przebieg jednej z pierwszych w Polsce ewaluacji (nastawionej na usuwanie blokad wdrażania) – trudności i wyzwania związane ze zdobywaniem nowych doświadczeń zarówno przez zamawiającego, jak i wykonawcę.

Przedmiotem trzeciego studium przypadku jest ewaluacja bieżąca sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego. Opisany projekt badawczy stanowi interesujące studium przypadku oceny funkcjonowania sieci, prezentuje nowatorskie metody badawcze,

opisuje słabe i mocne strony procesu badawczego, wskazując jednocześnie przyjęte rozwiązania i działania naprawcze.

W ostatnim artykule przedstawiamy studium przypadku ewaluacji ciągłej czterech programów współpracy ponadgranicznej INTERREG IIIA dla Austrii i krajów ościennych – Czech, Węgier, Słowacji i Słowenii. Analizowany przypadek jest jednym z nielicznych europejskich przykładów badania ciągłego, tj. ewaluacji bieżącej (prowadzonej w trakcie wdrażania programu), ale składającej się z badań i konsultacji połączonych w jeden, spójny i wielomiesięczny kontrakt.

ROZDZIAŁ 4.2.

Wojciech Pander
Elżbieta Kozłowska

EWALUACJA BIEŻĄCA PROGRAMU ŚRODKI PRZEJŚCIOWE (*TRANSITION FACILITY*) W POLSCE

Niniejszy artykuł dotyczy ewaluacji okresowej Programu Środki Przejściowe (*Transition Facility*) w Polsce. Była to ewaluacja specyficzna, co wynikało z kilku faktów. Po pierwsze, ewaluacja prowadzona przez Wykonawcę trwała szesnaście miesięcy, co, biorąc pod uwagę praktykę i zakres dotychczas prowadzonych badań ewaluacyjnych w Polsce, jest praktyką nieczęstą[1]. W trakcie jej trwania przygotowywano raporty sektorowe dla poszczególnych obszarów tematycznych. W ramach każdego takiego obszaru (sektora) analizowane były projekty. Każdy projekt musiał być poddany ocenie co najmniej raz w toku jego realizacji, przy czym zwykle oceniany był częściej. Po drugie, prowadzone w ramach projektu badania, miały wymiar *stricte* ekspercki – w każdą ewaluację zaangażowane było grono ekspertów tematycznych (specjalistów), bez których wiedzy przygotowanie każdej ewaluacji byłoby prawdopodobnie niemożliwe. Po trzecie, proces ewaluacji, procedury, wyniki, a nawet struktura

[1] Ewaluacja Programu *Transition Facility* (Środki Przejściowe) prowadzona była *de facto* od początku realizacji Programu, w ramach innego niż opisywany kontraktu, który dotyczył przede wszystkim oceny projektów realizowanych ze środków Programu Phare.

raportów z ewaluacji zostały określone w *Podręczniku ewaluacji okresowej Programu Phare i Transition Facility (Środki Przejściowe) w Polsce*. To powoduje, że omawiana ewaluacja bieżąca (*on-going*) ma charakter zdecydowanie odmienny od innych badań tego typu prowadzonych w polskich instytucjach. W dalszej części artykułu opisany jest kontekst i przedmiot ewaluacji, w tym przedmiot badania, warunki operacyjne badania, potrzeby informacyjne Zamawiającego oraz zakres badania. W dalszej części przedstawione zostały szczegółowo metodologia badania, przebieg procesu badawczego, jak również najważniejsze wyzwania i ograniczenia ewaluacji okresowej. Kolejny podrozdział stanowi prezentację wyników badania, w tym opis sposobu ich rozpowszechniania. Całość zamykają wnioski, które kończy zestawienie mocnych i słabych stron procesu ewaluacji okresowej Programu Środki Przejściowe.

1. KONTEKST I PRZEDMIOT EWALUACJI

1.1. Opis przedmiotu badania

Program Środki Przejściowe (*Transition Facility*), stanowi kontynuację Programu Phare, wdrażanego w Polsce przed przystąpieniem naszego kraju do struktur Unii Europejskiej. Celem tego programu jest wsparcie procesów wzmacniania polskiej administracji, wymiaru sprawiedliwości oraz wdrażania prawodawstwa Unii Europejskiej. Środki programowe zostały przyznane Polsce na lata 2004–2006, na podstawie art. 34 Traktatu Akcesyjnego, jako dalsza pomoc w rozwoju możliwości administracyjnych związanych z wdrażaniem prawodawstwa wspólnotowego (*acquis communautaire*). Zasady realizacji projektów finansowanych w ramach Programu zostały określone w *Protokole Uzgodnień w sprawie wdrażania Środków Przejściowych* podpisanego pomiędzy Rządem Rzeczypospolitej Polskiej a Komisją Europejską. Program jest realizowany za pomocą następujących instrumentów:

- *twinning* i *twinning light* z partnerami z krajów członkowskich (procedura wyboru partnerów pozostała analogiczna do stosowanej w przypadku projektów Phare – rozwój instytucjonalny);
- pomoc techniczna;
- współfinansowanie „na małą skalę" zakupu sprzętu związanego z wdrażaniem *acquis communautaire*.

Programowanie Środków Przejściowych rozpoczęto jesienią 2003 roku. Opierając się na wstępnych raportach Komisji Europejskiej ogłoszonych w lutym i maju 2003 roku, strona polska przygotowała wstępne propozycje priorytetów wsparcia z nowego programu. Priorytety te zostały, w porozumieniu z Komisją Europejską, zaktualizowane i zweryfikowane w listopadzie 2003 roku. Propozycje projektów na rok 2004 oraz ogólny program na lata 2005 i 2006 zostały przedstawione Komisji

Europejskiej na początku 2004 roku. Natomiast ostateczne propozycje projektów na lata 2004–2006, po negocjacjach z KE, zostały zatwierdzone na podstawie decyzji finansowych Komisji Europejskiej wydanych kolejno: dla alokacji 2004 w dniu 29 listopada 2004 roku, dla alokacji 2005 w dniu 2 sierpnia 2005 roku oraz alokacji 2006 w dniu 27 lipca 2007 roku. Z wydaniem ostatniej decyzji oraz przeprowadzeniem procedury Wymiany Listów, w roku 2006 roku zakończono proces programowania Środków Przejściowych[2]. Termin realizacji projektów w ramach Programu Środków Przejściowych upływa z końcem 2009 roku.

Ewaluacja okresowa Programu Środki Przejściowe (*Transition Facility*) dotyczyła przeglądu projektów i działań będących w trakcie realizacji (tj. od daty podpisania decyzji przez KE do zakończenia okresu wydatkowania) na poziomie właściwego sektora monitorującego. Zamówienie dotyczyło projektów wdrażanych w ramach alokacji Środków Przejściowych 2004–2006. Oceniane projekty zostały przyporządkowane do pięciu sektorów: Rynek wewnętrzny, Sprawiedliwość i sprawy wewnętrzne, Administracja publiczna, Środowisko i transport oraz Rolnictwo. Większość sektorów monitorujących została dodatkowo podzielona na komponenty (np. sektor „Środowisko i transport" dzielił się na 3 komponenty: ochrona środowiska, bezpieczeństwo nuklearne, transport). Każdy projekt w okresie jego realizacji, musiał być przynajmniej raz poddany ewaluacji. O specyfice ewaluacji okresowej pisze się także w dalszej części artykułu.

1.2. Dlaczego podjęto badanie?

Ewaluacja okresowa została wpisana w proces zarządzania Programem Środki Przejściowe. Środki na ewaluację zostały zaalokowane w ramach jednego z projektów Programu, którego celem było wzmocnienie zdolności monitoringowych i ewaluacyjnych, a w konsekwencji poprawa wdrażania i zarządzania Programem.

Celem prowadzonej ewaluacji projektów w ramach programu Środki Przejściowe było zmniejszenie potencjalnych zagrożeń, które mogły powstać w fazie wdrażania projektów oraz zwiększenie efektów wdrażania projektów. Zgodnie z zapisami SIWZ przedmiot zamówienia obejmował:

- zapewnienie regularnej ewaluacji okresowej programu Środki Przejściowe (*Transition Facility*) przez niezależnego ewaluatora;
- opracowywanie na jej podstawie stosownych raportów i ekspertyz;
- współpracę ewaluatora z Jednostką Ewaluacyjną Narodowego Koordynatora Pomocy w opracowaniu Raportu o Postępach we Wdrażaniu Programu Phare i *Transition Facility* (Środki Przejściowe) w części dotyczącej oceny wdrażanych programów w ramach *Transition Facility*;

[2] Na podstawie strony: www.ukie.gov.pl

- wspieranie procesu monitorowania wdrażania rekomendacji przez ścisłą współpracę ewaluatora z Jednostką Ewaluacyjną Narodowego Koordynatora Pomocy;
- wparcie Grupy Sterującej w ramach działań związanych z koordynacją procesu ewaluacji okresowej.

1.3. Kim byli kluczowi interesariusze?

W związku z wdrażaniem Programu Środki Przejściowe (*Transition Facility*) Komisja Europejska przewidziała wykorzystanie struktur organizacyjnych, jakie zostały wypracowane w trakcie wdrażania programu Phare. W związku z tym do realizacji Programu zaangażowano: Urząd Komitetu Integracji Europejskiej, w tym Departament Obsługi Narodowego Koordynatora Pomocy oraz Departament Programów Rozwoju Instytucjonalnego, Narodowy Fundusz, którego rolę odgrywa Ministerstwo Finansów oraz Jednostkę Finansująco-Kontraktującą (Fundacja „Fundusz Współpracy"). Obok instytucji zaangażowanych w realizację Programu, kluczową rolę w procesie ewaluacji odgrywali beneficjenci poszczególnych projektów[3].

1.4. Jaki to był typ ewaluacji?

Ewaluacja Programu Środki Przejściowe miała charakter *on-going* (ewaluacji bieżącej). Specyfika tej ewaluacji polegała na ocenie projektów realizowanych w ramach Programu, co najmniej raz w cyklu ich realizacji, z położeniem nacisku na okres realizacji projektów i stan ich zaawansowania. Należy podkreślić, iż tego rodzaju badania ewaluacyjne, które cechuje powtarzalność oceny, stosowane są w Polsce wyjątkowo rzadko.

Zasady ewaluacji Programu Środki Przejściowe zostały określone w *Podręczniku ewaluacji okresowej Programu Phare i Transition Facility (Środki Przejściowe) w Polsce*[4]. Zgodnie z zapisami *Podręcznika*, ogólnym celem systemu ewaluacji okresowej jest pomoc przy poprawie ukierunkowania, skuteczności, oddziaływania i rozliczania programu *Transition Facility*. W ten sposób zapisy podręcznika jasno wskazały kierunek prowadzonych badań ewaluacyjnych. Warto podkreślić, iż podręcznik został opracowany na bazie wcześniejszych doświadczeń wdrażania Programu Phare.

Zamawiającym ewaluację okresową Programu *Transition Facility* (Środki Przejściowe) była Jednostka Finansująco-Kontraktująca (Fundacja „Fundusz Współpracy"), która działała w imieniu Jednostki

---

[3] Beneficjentami projektów były przede wszystkim instytucje i agendy rządowe, które realizują zadania wynikające z wdrażania *acquis communautaire*.

[4] Podręcznik opracowano na zlecenie Urzędu Komitetu Integracji Europejskiej.

Ewaluacyjnej Narodowego Koordynatora Pomocy (Departament Obsługi Narodowego Koordynatora Pomocy UKIE). Jednostka Ewaluacyjna NKP odpowiedzialna była za bezpośredni nadzór nad realizacją projektu.

2. WKŁAD

2.1. Jakie były potrzeby informacyjne i jakie pytania postawił Zamawiający?

Na potrzeby każdej z ewaluacji przygotowywano *Notatkę metodologiczną*, która zawierała katalog instytucji zaangażowanych w proces ewaluacji, katalog pytań badawczych wraz z propozycją metod badawczych, których wykorzystanie pozwalało na uzyskanie odpowiedzi na sformułowane pytania. Notatka wskazywała kierunek badań oraz najważniejsze kwestie problemowe w ujęciu kryteriów badawczych. Zakres badania i sformułowanych zagadnień problemowych najczęściej opierał się na wiedzy ewaluatora zdobytej na podstawie lektury fiszki projektowej oraz innych dostępnych materiałów, jak również na podstawie wiedzy eksperckiej z danego obszaru tematycznego. Na podstawie notatki metodologicznej prowadzono badania. W fazie badawczej nie dokonywano już zmian przyjętej metodologii, choć zdarzały się sytuacje, w których ewaluator zmuszony był do stosowania dodatkowych metod bądź narzędzi badawczych, gdyż te pierwotnie zaplanowane, nie dawały satysfakcjonującej wiedzy odnośnie do projektów lub jego efektów.

Podsumowując, można więc stwierdzić, iż potrzeby informacyjne Jednostki Ewaluacyjnej NKP uwzględniane były na etapie przygotowania każdego z badań. Do wykonawcy badania należało określenie szczegółowego zakresu prowadzonych badań, który następnie był akceptowany przez Jednostkę Ewaluacyjną NKP, która kierowała się potrzebami swoimi oraz innych interesariuszy. Zmiany mogły być również dokonane na wniosek Jednostki Ewaluacyjnej NKP.

2.2. Jak sprecyzowano zakres badania?

W odniesieniu do każdego z pięciu sektorów (Rynek wewnętrzny, Sprawiedliwość i sprawy wewnętrzne, Administracja publiczna, Środowisko i transport, Rolnictwo) zakres badania miał innych charakter czasowy i przedmiotowy. Dodatkowo w ramach ewaluacji realizowano ewaluacje tematyczne oraz Raporty krótkoterminowych ekspertów (STTS[5]), które miały charakter horyzontalny względem przedstawionych sektorów. Zróżnicowanie zakresu badań sektorowych było przede wszystkim wynikiem przedstawionych wymogów Komisji Europejskiej dotyczących

[5] STTS – ang. Short-term Technical Specialist.

konieczności ewaluacji projektów w cyklu rocznym. Tak jak już wcześniej wspomniano, dobór projektów i obszarów tematycznych do poszczególnych ewaluacji był planowany z dużym wyprzedzeniem (*Roczny plan ewaluacji*). W odniesieniu do oceny poszczególnych projektów zakres ewaluacji dotyczył zwykle podobnych obszarów, co z kolei wynika z możliwych do realizacji rodzajów projektów. Beneficjenci mogli bowiem realizować bądź projekty współpracy z partnerami z krajów członkowskich (*twinning* i *twinning light*), bądź to projekty pomocy technicznej. Trzecim możliwym rodzajem projektów były projekty zakładające zakup „na małą skalę" sprzętu związanego z wdrażaniem *acquis communautaire*. W ramach projektów twinningowych realizowano zwykle szkolenia dla kadry polskich instytucji bądź projekty doradcze mające na celu wymianę bądź zdobycie nowej wiedzy. Projekty pomocy technicznej obejmowały szeroki katalog działań, począwszy od projektów doradczych, badawczych i analitycznych, projektów wsparcia instytucjonalnego, po projekty zakładające rozpowszechnianie wiedzy bądź promocję określonych tematów i zagadnień.

Precyzowanie zakresu badania, które odbywało się na etapie przygotowania *Notatki metodologicznej*, uwzględniało więc zwykle charakter realizowanych projektów i dokonywane było wspólnie przez wykonawcę i Zamawiającego (w procesie konsultacji *Notatki metodologicznej*).

2.3. Jakie założenia metodologiczne postawiono w Opisie Przedmiotu Zamówienia?

Ze względu na złożoność przedsięwzięcia, na etapie wyboru wykonawcy niniejszego badania, nie sformułowano szczegółowych założeń metodologicznych badania, a jedynie poproszono o przedstawienie metodologii oceny 12 projektów reprezentujących poszczególne sektory. Wynikiem tego była duża dowolność w zakresie doboru metodologii oceny pozostałych projektów na etapie realizacji projektu. Konsekwencją tego była również możliwość przedstawienia odrębnej metodologii dla każdego z badań ewaluacyjnych, przy czym należy pamiętać, że projekty były ewaluowane zgodnie z właściwymi procedurami, standardami oraz w ramach struktur opisanych w *Podręczniku ewaluacji*, w odniesieniu do celów określonych w fiszkach projektowych oraz dokumentach programowych, co decydowało o spójności wyników poszczególnych badań.

2.4. Jakie warunki operacyjne postawiono przed badaniem?

Głównym celem kontraktu było zapewnienie regularnej ewaluacji okresowej programu Środki Przejściowe przez niezależnego ewaluatora. Założono, iż ewaluacja procesu wdrażania programu Środki Przejściowe powinna prowadzić do zmniejszenia ewentualnych zagrożeń oraz zwiększenia

efektów wdrażania projektów. Tak postawione zadanie determinowało warunki operacyjne zadania.

Ewaluacja prowadzona była w okresie od połowy sierpnia 2006 do połowy grudnia 2007 roku. Lista projektów poddanych ewaluacji ustalana była każdorazowo w *Rocznym planie ewaluacji* na dany rok. Przykładowo *Roczny plan ewaluacji na rok 2007* obejmował wykonanie 18 *Sektorowych raportów ewaluacyjnych* dla wszystkich pięciu sektorów oraz dwóch *Tematycznych raportów ewaluacy*jnych. Terminy wykonania raportów sektorowych zostały szczegółowo określone w Planie Ewaluacji, zaś terminy wykonania raportów tematycznych były każdorazowo ustalane przez JE NKP w porozumieniu z Dokonującym Oceny (ewaluatorem). Celem tak skonstruowanego systemu ewaluacji było umożliwienie Narodowemu Koordynatorowi Pomocy dokonywania oceny stanu realizacji Programu Środki Przejściowe (*Transition Facility*), a także przedstawienie zaleceń mających na celu usprawnienie jego zarządzania. Poszczególne oceny (ewaluacje) obejmowały także kwestie związane z postępami w realizacji danego programu oraz jego efektywnością i trwałością w odniesieniu do przyjętych celów, jak również uzyskaniem doświadczeń pozwalających na poprawę realizacji innych, podobnych programów.

Za przygotowanie rocznych planów ewaluacji odpowiedzialny był Wykonawca, który przy ich przygotowaniu musiał brać pod uwagę priorytety założone przez JE NAC w konsultacji z Grupą Sterującą, a następnie zatwierdzone przez Wspólny Komitet Monitorujący. Roczne plany ewaluacji musiały być także skoordynowane z innymi działaniami w zakresie monitorowania i ewaluacji podejmowanymi przez JE NAC i oparte na założeniu, iż każdy sektor powinien być poddany ewaluacji każdego roku, a każdy projekt przynajmniej raz w cyklu życia projektu.

Dobór projektów do ewaluacji uwzględniał także etap, na jakim znajdował się każdy projekt, kładąc nacisk na ocenę tych projektów, których realizacja była już zaawansowana. Jednakże nie zawsze możliwy był taki właśnie dobór projektów, co skutkowało tym, że w ramach ewaluacji oceniano zarówno projekty zaawansowane, jak i będące w bardzo wstępnej fazie realizacji (w której faktyczne działania w ramach projektów nie zdążyły się jeszcze rozpocząć). Tak więc zakres oceny projektów zaawansowanych był znacznie szerszy i różny od tego, jaki stosowano dla projektów w bardzo wstępnej fazie realizacji. Dodatkowo, biorąc pod uwagę sformułowany katalog sektorów, do których przyporządkowano projekty, nie zawsze obszar badawczy obejmował projekty o spójnym charakterze. Decydowało to o tym, że zakres potrzeb informacyjnych każdego badania ewaluacyjnego był inny, a w konsekwencji, wymagały indywidualnego podejścia oraz specjalnie zaprojektowanych pytań ewaluacyjnych, a także doboru specyficznej dla danego badania metodologii.

W odniesieniu do każdego rodzaju projektów przyjmowano w badaniu odmienną koncepcję realizacji badania. Inaczej bowiem oceniano

projekty, które nastawione były na zakup aparatury (tutaj skupiano się przede wszystkim na aspekcie efektywności kosztowej i adekwatności zakupów do potrzeb instytucji), a inaczej projekty zakładające przeszkolenie szerokiego grona osób pracowników Służby Cywilnej (tutaj z kolei prowadzono na szeroką skalę badania efektywności szkoleń i ich użyteczności dla pracowników). Każdy typ projektu miał więc swoje podejście metodyczne, natomiast ocena opierała się zawsze na przyjętych kryteriach oceny projektów.

W efekcie takiego podejścia badawczego skład zespołu badawczego w ramach każdej z ewaluacji zmieniał się. W zależności od złożoności zakresu projektów poddawanych ewaluacji, w skład zespołu wchodzili eksperci tematyczni, ekspert metodyczny (nadzorujący proces ewaluacji i odpowiedzialny za przygotowanie raportu końcowego) oraz asystenci.

3. PROCES BADAWCZY

3.1. Jakie podejście badawcze i metodologię przyjęli ewaluatorzy?

Konieczność dokonania oceny często bardzo specyficznych projektów realizowanych w ramach Programu, wiązała się z koniecznością pozyskiwania specjalistycznej wiedzy odnośnie do określonych zagadnień. Stąd też większość narzędzi i metod badawczych, stosowanych w omawianej ewaluacji okresowej miała charakter jakościowy i specjalistyczny (dlatego też w ewaluację okresową projektów Programu włączano wielu ekspertów tematycznych). W zasadzie tylko w odniesieniu do projektów twinningowych zakres stosowanych narzędzi badawczych wykraczał poza ramy jakościowe i obejmował również katalog narzędzi ilościowych. Do najczęściej stosowanych metod jakościowych należały: pogłębione wywiady indywidualne (IDI), wywiady grupowe (diady, triady), panele ekspertów oraz grupy fokusowe. W przypadku badań ilościowych podstawowym narzędziem było CATI (Computer Assisted Telephone Interview). Katalog stosowanych metod badawczych był szeroki, a zakres stosowanych metod wynikał ze specyfiki realizowanych projektów. Warto natomiast podkreślić w tym miejscu, że zdecydowanie jakościowy charakter prowadzonych badań (odpowiadający specyfice projektów i Programu Środki Przejściowe) skutkował głębszym i szerszym spojrzeniem ewaluatorów na analizowane projekty. W konsekwencji możliwe było z jednej strony dokonanie oceny procesu wdrażania projektów, identyfikowanie istniejących barier, czy chociażby neutralizowanie potencjalnych ryzyk projektowych (podejście operacyjne), z drugiej zaś, pozwalało to na szersze spojrzenie na analizowane zagadnienia (podejście strategiczne).

3.2. Jak przebiegał proces badawczy?

Proces badawczy prowadzony w odniesieniu do pięciu sektorów odbywał się w sposób określony w *Podręczniku ewaluacji okresowej Programu Phare i Transition Facility (Środki Przejściowe) w Polsce*. Podręcznik „wymuszał" niejako na Dokonującym Oceny procedurę przygotowania każdego raportu ewaluacyjnego, w tym terminy wykonania poszczególnych badań i przygotowania raportów, konieczność przygotowania w odpowiednim czasie rekomendacji (zaleceń) oraz terminy spotkań z interesariuszami procesu ewaluacji. Podręcznik determinował także schemat raportu końcowego z ewaluacji.

Konsekwencją proceduralnego charakteru każdej ewaluacji projektów realizowanych w ramach Programu Środki Przejściowe był fakt, iż każda ewaluacja miała identyczny, bądź prawie identyczny charakter i obejmowała maksymalnie pięciomiesięczny okres, od spotkania przedwstępnego (*pre-Kick-off-meeting*) do spotkania pokontrolnego (*debriefing meeting*). Proces ten nie uwzględniał także zróżnicowanej liczby projektów poddanych ewaluacji w ramach jednego badania, choć w przypadku większej liczby ocenianych projektów podejmowane kroki, które miały przyspieszyć proces ewaluacji, takie jak skracanie czasu przekazywania dokumentów przez beneficjentów do Wykonawcy. Analizując przyjęty w *Podręczniku* harmonogram realizacji ewaluacji, należy stwierdzić, że w opinii ewaluatorów, zdecydowanie lepszy skutek odniosłoby zastosowanie bardziej elastycznego podejścia, a zwłaszcza skrócenie czasu trwania procesu ewaluacji (lecz nie samego czasu przeznaczonego na badanie). Wydaje się, że bez uszczerbku dla jakości prac skróceniu mogły ulec kwestie organizacyjno-proceduralne, np: okres pomiędzy spotkaniem w sprawie *Tabeli zaleceń* a spotkaniem pokontrolnym, choć z drugiej strony jest to okres, w którym zalecenia Dokonującego Oceny są weryfikowane w praktyce, stąd skracanie tego okresu nie zawsze byłoby zasadne. Należy jednocześnie zauważyć, iż takie elastyczne podejście stosowano z powodzeniem w porozumieniu z Zamawiającym w końcowej fazie realizacji projektu.

Na podstawie wniosków z przeprowadzonego badania zespół badawczy opracowywał zestaw rekomendacji i zaleceń, których celem była poprawa jakości wdrażania każdego z ocenianych projektów. Opis proponowanych przez ewaluatora rekomendacji i zaleceń przedstawiany był w raporcie ewaluacyjnym przygotowanym dla danego komponentu. Szczegółowe dane na temat rekomendacji wraz ze wskazaniem zalecanego czasu ich wdrożenia oraz instytucji odpowiedzialnej za ich realizację przedstawiano w *Tabeli zaleceń*. Raport zawierający wyniki ewaluacji wraz z wnioskami oraz rekomendacjami każdorazowo przedstawiano interesariuszom badania (z reguły wszystkim podmiotom zaangażowanym w proces ewaluacji oraz wszystkim beneficjentom projektów).

W procesie oceny, każda z zaangażowanych instytucji mogła zgłosić swoje uwagi do produktów wypracowanych przez zespół ewaluacyjny. Uwagi te były uwzględniane lub – w przypadku rozbieżności – dyskutowane i omawiane ze zgłaszającym w celu wypracowania konsensusu. W przypadku, gdy nie udało się uzgodnić jednoznacznego stanowiska pomiędzy ewaluatorem a zgłaszającym uwagę, rozbieżności były wymieniane w *Tabeli rozbieżności,* stanowiącej jeden z elementów Raportu.

Na końcowym etapie procesu ewaluacji danego komponentu odbywało się spotkanie w sprawie *Tabeli zaleceń*. Na spotkaniu tym omawiano proponowane przez ewaluatora rekomendacje oraz dyskutowano o możliwościach i zasadności ich wdrożenia. Spotkanie to dawało możliwość bezpośredniej konfrontacji opinii ewaluatora z opiniami przedstawicieli instytucji poddanych ewaluacji. Wdrażanie przyjętych podczas tego spotkania rekomendacji było monitorowane przez JE NKP (na podstawie *Tabeli działań następczych*). Proces ewaluacji kończyło spotkanie pokontrolne, podczas którego przestawiano informacje na temat wdrożenia rekomendacji i efektów ich zastosowania.

3.3. Jakie były wyzwania i ograniczenia procesu ewaluacji?

Ze względu na złożoność realizacji projektów i liczbę zaangażowanych w ten proces instytucji, największym problemem (szczególnie w pierwszym okresie realizacji badań) był proces gromadzenia danych. Instytucje zaangażowane w proces ewaluacji, odpowiadające za poszczególne fragmenty Programu, zwłaszcza CFCU (*Central Financing and Contracting Unit*; Jednostka Kontraktujaco-Finansująca w Fundacji Fundusz Współpracy), odpowiadająca za bezpośrednie wdrażanie projektów Programu (organizowanie przetargów, bieżące rozliczenie projektów i kontrolę realizacji działań projektowych), posiadały specyficzny zasób informacji oraz różne procedury gromadzenia danych i informacji. W efekcie, w przygotowaniu pierwszych ewaluacji, dużą trudność sprawiło Dokonującemu Oceny zdobywanie ujednoliconych informacji. W początkowym etapie trwania projektu, ewaluatorzy na własną rękę podejmowali wysiłek zdobycia pewnej porcji istotnych z punktu widzenia celu badania informacji, bez świadomości, że dane te były łatwo dostępne w systemie sprawozdawczości zaangażowanych w proces ewaluacji instytucji. Jednocześnie informacja o dostępności tych danych nie została wyartykułowana na spotkaniach rozpoczynających ewaluację (*pre-Kick off-meeting* oraz *Kick-off-meeting*). Taka sytuacja niepotrzebnie utrudniła pracę zespołu badawczego.

Do wyzwań ewaluacji okresowej Programu Środki Przejściowe zaliczyć należy konieczność łączenia oceny projektów z różnych obszarów tematycznych, takich jak projekty z zakresu ochrony zdrowia z projektami dotyczącymi rynku telekomunikacyjnego czy też projektami dotyczącymi ochrony konsumentów. Usprawiedliwieniem jest fakt, iż ze

względu na ramy formalne tej ewaluacji (Plan ewaluacji) oraz wymogi (konieczność oceny projektu w cyklu rocznym) ułożenie ocenianych projektów w taki sposób, aby zawsze łączyły się tematycznie, byłoby prawdopodobnie niemożliwe.

Istotnym wyzwaniem procesu ewaluacji była także konieczność gromadzenia w krótkim czasie dużego zasobu wiedzy specjalistycznej. Wymagało to od Dokonującego Oceny dużej elastyczności, m.in. w zakresie doboru osób do paneli ekspertów, czy też w zakresie przygotowania raportów krótkoterminowych ekspertów (STTS). Zdarzały się również sytuacje, iż specjalistyczną wiedzę dotyczącą ocenianych projektów trzeba było zdobywać ze źródeł zagranicznych, ze względu na fakt, iż potencjał polskich instytucji nie był w tym obszarze wystarczający.

Zdecydowanie najtrudniejszym zadaniem w procesie ewaluacji Programu była ocena trwałości i oddziaływania poszczególnych projektów, zwłaszcza tych, które były we wczesnej fazie realizacji. W prowadzonych badaniach przyjęto założenie, iż ocena trwałości i oddziaływania odbywa się na podstawie planowanych i możliwych do osiągnięcia efektów, założonych na etapie przygotowania fiszki projektowej. Ocena pozostałych kryteriów oceny projektów nie nastręczała już tylu problemów. W przypadku analizy kryteriów skuteczności i efektywności, oceniano bieżącą realizację projektów. Opierając się na pytaniach badawczych, sformułowanych do danego kryterium, analizowano bieżącą realizację projektów, osiąganie założonych wskaźników, stopień wydatkowania środków oraz wiele innych aspektów realizacji projektów. W przypadku tych kryteriów wypracowano praktyczną i operacyjną wiedzę dotyczącą możliwości dokonania usprawnień w projektach. Natomiast w przypadku kryterium adekwatności sformułowane pytania ewaluacyjne dawały zwykle możliwość oceny projektów w szerszym, strategicznym kontekście.

3.4. Jaką rolę odegrali poszczególni interesariusze w procesie badawczym?

Jak już wspomniano, poszczególne raporty ewaluacyjne pięciu sektorów przygotowywane były na podstawie metodologii przyjętej na spotkaniu rozpoczynającym. Spotkanie rozpoczynające dawało również możliwość zgłoszenia przez beneficjentów uwag do przygotowanej przez Wykonawcę na zlecenie i zaakceptowanej przez JE NAC metodologii. Przyjęte podejście badawcze każdorazowo zakładało kontakt z wszystkimi interesariuszami procesu ewaluacji (PAO, JE NAC, CFCU, Beneficjenci).

Warto zwrócić także uwagę na fakt, że procedury ewaluacji określone w *Podręczniku ewaluacji okresowej Programu Phare i Transition Facility (Środki Przejściowe) w Polsce* nakazywały obowiązkowy udział w procesie oceny przedstawicieli wszystkich stron realizacji i wdrażania Programu Środki Przejściowe. Skutkiem tego był m.in. udział

przedstawicieli instytucji we wszystkich spotkaniach, które dotyczyły ewaluacji. W efekcie można powiedzieć, że ewaluacja ta miała partycypacyjny charakter (choć w przypadku mniej aktywnych i zaangażowanych beneficjentów udział ten miał charakter wyłącznie formalny).

Podsumowując, można więc stwierdzić, iż ewaluacja okresowa Programu Środki Przejściowe (*Transition Facility*) nastawiona była zdecydowanie na poprawę bieżącego zarządzania Programem, jednak nie brakowało w niej akcentów strategicznych. Proces gromadzenia wiedzy o realizowanych projektach, a następnie wykorzystanie zdobytej wiedzy, dotyczył wszystkich interesariuszy procesu ewaluacji, w tym: Służb PAO, Służb NAC, Jednostki Ewaluacyjnej NAC, CFCU oraz beneficjentów. Wyniki ewaluacji obejmowały co prawda najczęściej beneficjentów projektów, jednak zdarzały się również wnioski i rekomendacje, które dotyczyły szerokiego grona instytucji.

4. PRODUKTY BADANIA

4.1 Jaką wiedzę uzyskano w badaniu?

W ramach każdego z sektorów, w kolejnych latach projektu, przygotowywano raporty z ewaluacji. Na każdy z raportów składała się ocena kilku projektów (w zależności od sektora, od dwóch do sześciu projektów). Dodatkowo, w ramach projektu przygotowano również raporty tematyczne oraz raporty krótkoterminowych ekspertów (STTS), które cieszyły się dużym zainteresowaniem beneficjentów projektów (ze względu na fakt, że przedmiot raportu mógł być *de facto* ustalony wspólnie, instytucje miały możliwość wpływania na zakres przedmiotowy Raportu STTS). Warto zauważyć, iż ewaluacje tematyczne oraz raporty krótkoterminowych ekspertów wprowadzały elastyczność do procesu ewaluacji, a jednocześnie dotyczyły zagadnień szerszych niż ocena projektów według przyjętych procedur, co w konsekwencji nadawało ewaluacji charakter bardziej strategiczny (przykładowo, w ramach jednego z raportów STTS oceniono sytuację na rynku energii elektrycznej bezpośrednio po jego liberalizacji w odniesieniu do przygotowania przedsiębiorstw do świadczenia usług dla indywidualnych klientów).

4.2 Jak wyglądał proces rozpowszechniania wyników?

Specyfika ewaluacji bieżącej Programu Środki Przejściowe była m.in. efektem podejścia do rekomendacji, które w sposób sformalizowany wdrażane były przez odbiorców. Wykonanie rekomendacji było z kolei monitorowane przez Służby NKP, tak więc ich przyjęcie wiązało się nie tylko z wyrażeniem zgody przez beneficjenta na wdrażanie rekomendacji, ale także z procesem ciągłego monitorowania jego wykonania. Można więc

uznać, iż proces wdrażania rekomendacji w ramach Programu Środki Przejściowe jest jedynym, w pełni funkcjonującym, systemem monitorowania wdrażania rekomendacji badań ewaluacyjnych w Polsce. Wnioski z poszczególnych badań oraz przygotowane rekomendacje były prezentowane jednak nie tylko na spotkaniach w sprawie *Tabeli zaleceń* oraz spotkaniach pokontrolnych, ale także na spotkaniach Podsektorowych Komitetów Monitorujących dla poszczególnych sektorów. Spotkania te gromadziły wszystkich interesariuszy Programu. Na potrzeby realizacji projektu oraz w celu upowszechniania wyników prowadzonych badań, przygotowano także stronę internetową, na której prezentowano streszczenia wykonanych raportów wraz z rekomendacjami. Strona zawierała także opis projektu, konsorcjum firm, które prowadziło badanie, a także inne treści informacyjne.

W odniesieniu do efektów poszczególnych ewaluacji można stwierdzić, że charakter wypracowanych rekomendacji, ich jakość oraz wdrożenie wyników ewaluacji, zależały przede wszystkim od podmiotów realizujących projekty (doświadczenia i zaangażowania zespołu realizującego projekt) oraz kultury organizacyjnej instytucji beneficjenta. Instytucje o wysokiej kulturze organizacyjnej zdecydowanie chętniej współpracowały z Dokonującym Oceny w procesie ewaluacji, chętniej przyjmowały uwagi i wnioski (również te krytyczne) oraz chętniej przyjmowały i wdrażały sformułowane rekomendacje (zalecenia). W konsekwencji ewaluacja projektów realizowanych przez tych beneficjentów miała bardzo partnerski charakter i nastawiona była na szukanie wspólnych rozwiązań istniejących lub potencjalnych problemów. Warto także zauważyć, że w odniesieniu do tych instytucji występujących problemów i zagrożeń było zdecydowanie mniej, bądź też w ogóle nie występowały. Na sprawną realizację projektów (a w konsekwencji ewaluację) rzutował także jeszcze jeden aspekt realizacji projektów – zarządzanie. Projekty zarządzane przez osoby o zdolnościach menedżerskich zdecydowanie rzadziej borykały się z problemami realizacyjnymi oraz wewnętrznymi i zewnętrznymi zagrożeniami.

Ważnym aspektem realizacji projektów była tzw. trwałość instytucjonalna. Projekty realizowane od początku do końca przez jeden zespół były, w naszej ocenie, projektami lepszymi w sensie realizacji i organizacji prac projektowych. Duża rotacja pracowników oraz zmiany organizacyjne w instytucjach były najczęstszym powodem pojawiających się problemów. Problemów nie dało się uniknąć także w projektach, które były zarządzane w sposób poprawny. Przykładowo pojawiające się problemy z ogłaszaniem przetargów i ich rozstrzygnięciem, skutkowały przesunięciami planowanych harmonogramów realizacji projektów w znacznej liczbie projektów. Innego rodzaju problemy pojawiały się z kolei w projektach, które dotyczyły bardzo zaawansowanych obszarów tematycznych, takich jak rynek telekomunikacyjny czy rynek energetyczny. W tym przypadku pojawiające się problemy wynikały m.in. z braku

wiedzy potencjalnych wykonawców projektów na tematy będące przedmiotem ogłaszanych postępowań. Tak więc posiadanie sprawnej struktury zespołu realizującego projekt oraz aktywny lider projektu łącznie nie stanowili gwarancji bezproblemowego wdrażania projektów.

Podsumowując, należy podkreślić, że wyniki prowadzonych badań ewaluacyjnych były kierowane do szerokiego spektrum odbiorców instytucjonalnych, zarówno w Polsce, jak i w służbach Komisji Europejskiej. Do odbiorców i użytkowników raportów z ewaluacji okresowej należeli przede wszystkim: Beneficjenci Programu Środki Przejściowe, Narodowy Koordynator Pomocy, Krajowy Urzędnik Zatwierdzający, służby Komisji Europejskiej, Pełnomocnik ds. Realizacji Projektów Rozwoju Instytucjonalnego i Koordynatorzy Projektów objętych ewaluacją, Jednostka Finansująco-Kontraktująca.

5. WNIOSKI ZE STUDIUM PRZYPADKU

5.1. Jakie były efekty, rzeczywista użyteczność badania?

Ewaluacja okresowa Programu Środki Przejściowe (*Transition Facility*) miała charakter ewaluacji nastawionej na wspieranie procesu zarządzania i realizacji Programu. Jej zakres i końcowe efekty były wynikiem przyjętych procedur realizacji procesu ewaluacji. W efekcie, ocena poszczególnych projektów z reguły bardziej koncentrowała się na procesach wdrażania projektów, usuwaniu barier wdrażania, wspomaganiu procesu realizacji projektów, niż na określaniu kierunków interwencji, zasadności podejmowania określonych kierunków interwencji, czy poszukiwaniu szerszych wniosków dla każdego z sektorów. W opinii autorów, ewaluacja okresowa powinna być jednak częścią szerszego procesu oceny i nie powinna skupiać się tylko na operacyjnym wspieraniu realizacji projektów. Częściowo, co prawda, rolę tę odgrywały raporty krótkoterminowych ekspertów STTS oraz raporty tematyczne, które dawały szerszy pogląd na sektor i realizację projektów w kontekście zmian w sektorze, jednak liczba tych raportów była niewielka w porównaniu z raportami sektorowymi. Specyficzny charakter tej ewaluacji wynikał jednak nie tylko ze sztywnych ram procesu ewaluacji, ale również z konieczności planowania z dużym wyprzedzeniem działań ewaluacyjnych przez Jednostkę Ewaluacyjną Narodowego Koordynatora Pomocy. Trudno było bowiem dopasować Plan ewaluacji do etapu realizacji poszczególnych projektów.

W procesie ewaluacji Programu Środki Przejściowe (*Transition Facility*), ważne okazało się połączenie wielu elementów procesu ewaluacji (monitoring, współpraca z instytucjami, procedury). Zakres danych monitoringowych bardzo często nie dawał pełnego obrazu realizowanych projektów, stąd bardzo istotna była współpraca z poszczególnymi

instytucjami zaangażowanymi w realizację projektów. Stosowanie zróżnicowanych metod badawczych (jakościowych) pozwalało z kolei na właściwą triangulację wyników badania.

Ważną funkcją, którą spełniała ewaluacja okresowa Programu Środki Przejściowe było uczenie się sektora publicznego. Odnosi się to zwłaszcza do tych projektodawców, którzy nie mieli wcześniejszego doświadczenia w realizacji projektów. W odniesieniu do nich Dokonujący Oceny często formułował rekomendacje uwzględniające aktualny stan wiedzy beneficjentów na temat procedur wdrażania projektów, ich rozliczania, czy też specyfiki realizacji projektów. W tym aspekcie warto również podkreślić konieczność obustronnej edukacji, szczególnie w pierwszym okresie realizacji badań ewaluacyjnych, kiedy przedstawiciele Dokonującego Oceny, mieli niepełny obraz realizowanych projektów, a pozyskiwanie i gromadzenie wiedzy było procesem bardzo trudnym. W tym kontekście ważna okazała się obustronna wymiana informacji odnośnie do potrzeb i oczekiwań drugiej strony. Warto przy tym także zwrócić uwagę, że tę rolę odgrywały także inne instytucje (beneficjent Programu był wspomagany na etapie realizacji i rozliczania projektu przez wszystkie zaangażowane instytucje).

Podsumowując proces ewaluacji okresowej Programu Środki Przejściowe, warto podkreślić, że była to ewaluacja specyficzna. Specyfika ta wynikała przede wszystkim z obowiązujących procedur wynikających z *Podręcznika ewaluacji okresowej*. Bezpośrednim tego efektem był na pewno mało elastyczny charakter tej ewaluacji. Z drugiej jednak strony, procedury dostarczały Dokonującemu Oceny zarówno bardzo skutecznego narzędzia zbierania i pozyskiwania danych, jak i prezentowania, dyskutowania i wdrażania przedstawianych rekomendacji. Pod tym względem należy bardzo wysoko ocenić proces ewaluacji okresowej Programu Środki Przejściowe. Proces ewaluacji obejmował w tym przypadku cały cykl – od przygotowania założeń badania, przez realizację badań, dyskutowanie rekomendacji do ich operacyjnego wdrożenia, jak również nadzorowania wdrażania przyjętych rekomendacji.

5.2. Jakie były silne i słabe strony analizowanej ewaluacji?

Podsumowując zaprezentowane wnioski, do silnych stron procesu ewaluacji okresowej zaliczyć należy:

- udział w procesie ewaluacji wszystkich zaangażowanych stron (zarówno beneficjentów, jak i instytucji wdrażających i nadzorujących realizację Programu), co rzutuje na całościowy obraz realizowanych projektów oraz obiektywność wyników prowadzonych badań;
- partycypacyjny charakter ewaluacji – w ramach ewaluacji okresowej możliwe było podejmowanie dyskusji odnośnie do realizacji projektów, sformułowanych rekomendacji, zakresu wdrażania rekomendacji itd.;

- procedury ewaluacji pozwalające wdrożyć rekomendacje, a następnie – co ważniejsze – śledzić ich wykonanie;
- podejmowanie współpracy z Dokonującym Oceny, skutkujące „uczeniem się sektora publicznego", szczególnie w odniesieniu do beneficjentów, którzy nie mieli doświadczenia w realizacji projektów w ramach Programu Środki Przejściowe (był to proces wzajemnego uczenia się);
- duży wpływ ewaluacji okresowej na poprawę zarządzania projektami i wychwytywanie potencjalnych zagrożeń.

Za słabe strony procesu ewaluacji okresowej Programu Środki Przejściowe (*Transition Facility*) uznać można:

- niską elastyczność systemu ewaluacji okresowej Programu, wynikającą z przyjętych procedur ewaluacji;
- konieczność ustalania z dużym wyprzedzeniem Planu ewaluacji, określającego założenia procesu ewaluacji i kierunek ewaluacji (sektory, komponenty badania, projekty poddawane ewaluacji);
- konieczność ujednolicania zakresu tematycznego projektów poddawanych ewaluacji w ramach poszczególnych raportów sektorowych;
- w początkowym okresie prowadzenia ewaluacji okresowej trudne okazało się pozyskiwanie danych, w posiadaniu których były poszczególne instytucje uczestniczące we wdrażaniu Programu;
- operacyjny charakter ewaluacji okresowej Programu uniemożliwiający często (oczywiście z pewnymi wyjątkami) ocenę na poziomie strategicznym Programu;
- krótki okres realizacji badań w ramach ewaluacji okresowej (okres prowadzenia badań zgodnie z zapisami *Podręcznika* wynosi 4 tygodnie), uniemożliwiający często pogłębioną analizę zagadnień.

ROZDZIAŁ 4.3.

Anna Domaradzka

EWALUACJA *ON-GOING* II PRIORYTETU ZPORR – IDENTYFIKACJA BARIER W SYSTEMIE WDRAŻANIA

Poniższy rozdział poświęcony jest omówieniu badania ewaluacyjnego *Identyfikacja barier w realizacji działań współfinansowanych ze środków Europejskiego Funduszu Społecznego w ramach Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego*[1] jako przykładu ewaluacji *on-going*. Opisany przypadek dobrze ilustruje przebieg jednej z pierwszych ewaluacji z tego zakresu w Polsce, a co za tym idzie – trudności związane ze zdobywaniem nowych doświadczeń zarówno przez Zamawiającego, jak i Wykonawcę.

Badanie zostało zaplanowane i zrealizowane pod koniec 2005 roku, na zlecenie Krajowej Jednostki Oceny Departamentu Koordynacji Polityki Strukturalnej Ministerstwa Rozwoju Regionalnego. Głównymi odbiorcami i interesariuszami ewaluacji były Departament Wdrażania Programów Rozwoju Regionalnego Ministerstwa Rozwoju Regionalnego oraz instytucje regionalne zaangażowane we wdrażanie ZPORR. Badanie zrealizował zespół ewaluatorów reprezentujących firmę Kantor Polska. Projekt był współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w Ramach Programu Operacyjnego Pomoc Techniczna.

---

[1] Zob. Olejniczak K., Ledzion B., Domaradzka A. (2006), *Raport końcowy z ewaluacji pt. Identyfikacja barier w realizacji działań współfinansowanych ze środków Europejskiego Funduszu Społecznego w ramach Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego*, Warszawa, Wydawnictwo MRR.

1. KONTEKST I PRZEDMIOT EWALUACJI

1.2. Opis przedmiotu badania

Punktem wyjścia dla realizacji priorytetu II Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego były dane dotyczące poziomu bezrobocia w Polsce, które na etapie planowania programu było największe spośród wszystkich krajów Unii Europejskiej. Wysokie bezrobocie, niski wskaźnik zatrudnienia oraz aktywności zawodowej zostały zdefiniowane jako jedna z największych barier rozwoju gospodarczego w Polsce, a także przyczyna wykluczenia dużych grup społecznych. Za szczególny problem uznano duży odsetek ludzi długotrwale bezrobotnych, wskazujący na strukturalny charakter bezrobocia w Polsce. Analiza przyczyn istniejących problemów wykazała, że utrata miejsc pracy była w dużej mierze związana z procesami restrukturyzacyjnymi w rolnictwie oraz tradycyjnych gałęziach przemysłu. Na tej podstawie przygotowano wstępne założenia działań II priorytetu ZPORR, których głównym celem miał być rozwój zasobów ludzkich jako czynnika kluczowego dla rozwoju gospodarczego. W tym celu należało dążyć do podniesienia poziomu zatrudnienia przez zwiększenie elastyczności i mobilności przestrzennej i zawodowej pracowników oraz wzrost produktywności. Realizacja tak wyznaczonych celów, określonych w Narodowym Planie Rozwoju, była w dużej mierze uzależniona od efektywnego wykorzystania funduszy strukturalnych, w tym środków Europejskiego Funduszu Społecznego. Przyjęta strategia działań wymagała więc stworzenia z jednej strony wydajnej struktury zarządzania funduszami, z drugiej zaś budowania aktywnej polityki społecznej oraz polityki rynku pracy.

1.3. Stan wdrażania i kontekst instytucjonalny

W ramach realizowanej oceny analizie poddane zostały regionalne projekty służące wzmocnieniu rozwoju zasobów ludzkich. Zakres obejmował więc projekty realizowane w ramach wszystkich sześciu Działań priorytetu II ZPORR, czyli:

- Działanie 2.1. Rozwój umiejętności powiązany z potrzebami regionalnego rynku pracy i możliwości kształcenia ustawicznego w regionie.
- Działanie 2.2. Wyrównywanie szans edukacyjnych przez programy stypendialne.
- Działanie 2.3. Reorientacja zawodowa osób odchodzących z rolnictwa.
- Działanie 2.4. Reorientacja zawodowa osób zagrożonych procesami restrukturyzacyjnymi.
- Działanie 2.5. Promocja przedsiębiorczości.
- Działanie 2.6. Regionalne Strategie Innowacyjne i transfer wiedzy.

W SIWZ przedstawiono stan wdrażania II priorytetu ZPORR w okresie przed rozpoczęciem badania. Jak pokazuje tabela 1, odsetek realizacji

Tabela 1. Liczba i wartość podpisanych umów oraz wartość dokonanych płatności[1]

Działanie/priorytet	Liczba umów	Wartość podpisanych umów (wkład EFS)	% realizacji zobowiązań na lata 2004–2006	Dokonane płatności z kont programowych	% realizacji zobowiązań na lata 2004–2006
Priorytet 2	1 007 (1473)[2]	379 856 186	21,38	5 467 507	0,30
Działanie 2.1	249 (381)	109 554 908	27,60	0	0,00
Działanie 2.2	596 (773)	170 764 807	34,62	5 440 022	1,09
Działanie 2.3	36 (66)	22 024 951	9,99	0	0,00
Działanie 2.4	18 (40)	15 440 743	5,13	0	0,00
Działanie 2.5	30 (96)	16 669 335	9,01	0	0,00
Działanie 2.6	78 (117)	45 401 442	25,16	27 485	0,02

Źródło: informacja miesięczna, 2005 r. za: *Opis przedmiotu zamówienia: „Identyfikacja barier w realizacji działań współfinansowanych ze środków Europejskiego Funduszu Społecznego w ramach Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego”*, z uzupełnieniami i komentarzami z dnia 14.11.2005.

[2] Dane w nawiasach odnoszą się do końca września 2005 roku.

zobowiązań był w badanym okresie bardzo niski, co powodowało uzasadnione obawy dotyczące udanego zakończenia programu i stawiało pod znakiem zapytania dotychczasową efektywność wydawania środków II priorytetu ZPORR.

1.4. Dlaczego podjęto badanie?

W kontekście wprowadzanych zmian kluczowym zagadnieniem stała się ocena efektywności systemu wdrażania EFS w Polsce, w tym działań realizowanych w ramach priorytetu II ZPORR, oraz pogłębiona analiza problemów zidentyfikowanych w pierwszym okresie wdrażania. Ocena ta miała obejmować zarówno instytucje zaangażowane we wdrażanie, jak i beneficjentów ostatecznych. Przeprowadzone badanie miało z założenia pozwolić na zgromadzenie wiedzy niezbędnej dla rozpoznania i przeciwdziałania istotnym barierom we wdrażaniu działań finansowanych z EFS w ramach ZPORR. Głównym zadaniem ewaluatorów było więc wskazanie źródeł istniejących problemów oraz zaproponowanie rozwiązań racjonalizacyjnych możliwych do zastosowania w ówczesnym bądź kolejnych okresach programowania.

Główną przyczyną podjęcia badania były niepokojące wyniki monitoringu wdrażania II priorytetu ZPORR, które wskazywały na istnienie barier w absorbcji środków w regionach. Szczególnie działanie 2.3, skierowane do rolników, napotykało problemy z wdrażaniem. Zamawiający chciał się dowiedzieć, czy bariery te mają charakter systemowy czy też wynikają ze złego zaadresowania części działań, co skutkuje niskim poziomem zainteresowania potencjalnych beneficjentów. Badanie miało więc charakter interwencyjny i w założeniu miało służyć poprawie wdrażania oraz zmniejszeniu ryzyka związanego ze zbyt powolnym kontaktowaniem środków.

1.5. Jaki to był typ ewaluacji?

Zgodnie z zamówieniem ewaluacja miała mieć charakter oceny uzupełniającej i przyjęła formę ewaluacji *on-going* (ewaluacja bieżąca), której celem było lepsze zrozumienie zachodzących procesów oraz zaplanowanie ewentualnych działań naprawczych. Badanie przeprowadzone zostało w trakcie realizacji przedsięwzięcia i koncentrowało się na analizie efektywności systemu zarządzania i wdrażania programu, z naciskiem na identyfikację barier w jego realizacji. Celem badania było także przygotowanie opisu stanu realizacji przyjętych celów w stosunku do założeń, oraz wypracowanie rekomendacji, które pozwoliłyby uniknąć powielania w przyszłości błędów zidentyfikowanych na badanym etapie.

2. WKŁAD W EWALUACJĘ

2.1. Jakie były potrzeby informacyjne i jakie pytania postawił Zamawiający?

Przeprowadzona ocena miała pozwolić na zdobycie informacji, które umożliwiłyby rozpoznanie i pogłębioną analizę stanu wdrażania i realizacji działań finansowanych z EFS w ramach II priorytetu ZPORR oraz dawałyby podstawy do tworzenia możliwych do zastosowania rozwiązań racjonalizujących.

Głównym celem zlecanej ewaluacji była więc odpowiedź na następujące pytania ewaluacyjne[3]:

- Jaka jest struktura realizowanych projektów w ramach II priorytetu ZPORR i jakie z nich stanowią największy udział w podziale na:
  - rodzaj realizowanego działania w ramach II Priorytetu ZPORR,
  - rodzaj realizowanego projektu,
  - beneficjenci ostateczni projektu,
  - wielkość realizowanego projektu,
  - obszar, na którym projekt jest realizowany.
- Czy realizowane projekty są zgodne z zapisami dokumentów programowych. Na ile działania podejmowane w ramach II priorytetu realizują ogólne cele ZPORR?
  - Czy realizowane projekty wpływają pozytywnie na rozwój danego regionu? Czy projekty wdrażane przez beneficjentów odpowiadają na potrzeby rynku pracy na danym obszarze?
  - Czy zapisy programu w kontekście II priorytetu ZPORR są trafne i odpowiadają potrzebom lokalnych i regionalnych rynków pracy?
- Jakie są bariery i problemy organizacyjne w systemie, skutkujące zróżnicowanym zainteresowaniem beneficjentów poszczególnymi działaniami w ramach II priorytetu ZPORR? W jaki sposób można przezwyciężyć te bariery i problemy?
- Jakie są przyczyny zróżnicowanego zainteresowania beneficjentów realizacją poszczególnych działań ?
  - Czy przepływ informacji pomiędzy samorządem a beneficjentami ma wpływ na liczbę realizowanych projektów w regionie w ramach II priorytetu ZPORR?
  - Czy podział kompetencji między poszczególnymi instytucjami zaangażowanymi w realizację II Priorytetu ZPORR wpływa na efektywną i skuteczną realizację poszczególnych działań? Jeżeli tak, to w jakim stopniu?

---

[3] *Opis przedmiotu zamówienia: „Identyfikacja barier w realizacji działań współfinansowanych ze środków Europejskiego Funduszu Społecznego w ramach Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego”*, z uzupełnieniami i komentarzami z dnia 14.11.2005.

2.2. Jak Zamawiający sprecyzował zakres badania?

Na etapie uszczegóławiania zapisów SIWZ Zamawiający określił wstępne wymagania względem próby badawczej. I tak, w przypadku pytania 1 Zamawiający zaproponował, by analizę przeprowadzić na wszystkich projektach realizowanych w ramach II priorytetu ZPORR zebranych w bazie[4], która miała zostać przekazana Wykonawcy przez Zamawiającego. W przypadku pytania 2, po konsultacjach z Wykonawcą, Zamawiający zaproponował, by do badania wybrać próbę kilku (np. trzech) województw, tak by reprezentowane w niej było przynajmniej jedno województwo najmocniejsze, najsłabsze i pośrednie. Podobnie w przypadku pytania 3 Zamawiający zaproponował, by do odpowiedzi na to pytanie wybrać próbę województw (np. trzech, w tym jedno najmocniejsze, jedno najsłabsze i jedno pośrednie), przy czym projekty z działania „najsłabszego" czyli 2.4 miały stanowić 50% próby projektów do analizy.

Ponadto Zamawiający zastrzegł w SIWZ, że przedstawiona lista pytań nie jest zamknięta, a Wykonawca powinien zaproponować dodatkowe zagadnienia, które jego zdaniem należałoby przeanalizować w kontekście ogólnych celów badania. Ostatecznie, w wyniku konsultacji ze Zlecającym, za główny cel badania przyjęto zidentyfikowanie barier i problemów systemowych związanych ze wdrażaniem II priorytetu ZPORR (pytanie 3).

W zakres objęty badaniem weszły wszystkie projekty składane w ramach II priorytetu ZPORR z okresu programowania 2004–2006, czyli w praktyce wszystkie projekty, które zostały uwzględnione w bazie na koniec lipca 2005 roku. Dodatkowo część analiz statystycznych wykonano na podstawie danych z raportów monitoringowych na dzień 31 grudnia 2005.

2.2. Jakie założenia metodologiczne przedstawiono w Opisie Przedmiotu Zamówienia?

Zamawiający oczekiwał, że dla realizacji badania Wykonawca będzie korzystał z różnych narzędzi i metod badawczych, a do oceny zastosowuje zarówno dane ilościowe, jak i jakościowe. W praktyce oznaczało to, że w ewaluacji miały zostać użyte metody pogłębionego wywiadu indywidualnego (IDI) w instytucjach wdrażających oraz wywiadu telefonicznego (CATI) z beneficjentami.

2.3. Jakie postawiono warunki operacyjne?

Harmonogram realizacji oceny zakładał, że tydzień po podpisaniu umowy zostaną uzgodnione narzędzia badawcze, do dnia 5 grudnia 2005

[4] Baza zawierała dane według stanu na koniec lipca 2005.

roku Departamentowi Koordynacji Polityki Strukturalnej przekazana zostanie wstępna wersja raportu, a do dnia 15 grudnia 2005 roku nastąpi przekazanie ostatecznej wersji raportu. Jednak ostateczna wersja SIWZ została przekazana Wykonawcy dopiero w połowie listopada. W wyniku opóźnień związanych z przekazywaniem przez Zamawiającego danych niezbędnych do rozpoczęcia prac badawczych czas na realizację badania uległ skróceniu. W praktyce oznaczało to, że Wykonawca miał niewiele ponad miesiąc na realizację badania i przygotowanie raportu, a pierwotnie założone terminy oddania raportu uległy przesunięciu.

Badanie zostało zrealizowane przez trzyosobowy zespół ekspertów oraz dwóch asystentów. Budżet badania wynosił 24 tys. PLN brutto przyznanych z wolnej ręki. Biorąc pod uwagę niską atrakcyjność finansową przedsięwzięcia, główną przyczyną podjęcia badania przez Wykonawcę była chęć zdobycia doświadczenia.

3. PROCES BADAWCZY

3.1. Jakie podejście badawcze i metodologię przyjęli ewaluatorzy?

Po przekazaniu przez Zamawiającego ostatecznej wersji SIWZ, zespół ewaluacyjny rozpoczął pracę nad metodologią badania. Na potrzeby stworzenia narzędzi badawczych, opisane w SIWZ, trzy główne pytania badawcze zostały rozłożone na szczegółowe pytania ewaluacyjne. W przypadku pytania 3, dotyczącego barier w zarządzaniu, zagadnienia badawcze uszczegółowiono w taki sposób, aby uzyskane odpowiedzi umożliwiły identyfikację głównych problemów oraz barier na każdym z trzech etapów wdrażania wybranych działań ZPORR (były to: identyfikacja i informowanie potencjalnych beneficjentów, selekcja projektów oraz wdrażanie projektów). Ewaluatorzy identyfikowali problemy pojawiające się na każdym z tych etapów, ze szczególnym uwzględnieniem tych barier, które mogły pośrednio bądź bezpośrednio wpływać na realizację Działań ZPORR, tym samym przekładając się na wyniki całego Priorytetu. Takie podejście analityczne przedstawione zostało Zamawiającemu i przez niego zaakceptowane. Dodatkowo, po otrzymaniu danych (przede wszystkim bazy danych projektów II Priorytetu ZPORR oraz sprawozdań monitoringowych według stanu na dzień 31 grudnia 2005), ewaluatorzy zweryfikowali praktyczną możliwość odpowiedzi na pytania badawcze, szczególnie w zakresie pytań 1 i 2. Z uwagi na fragmentaryczność danych zastanych, zakres pytań badawczych musiał zostać doprecyzowany i nieco zawężony. Ostatecznie, badaniem zostały objęte trzy województwa: opolskie, mazowieckie oraz wielkopolskie, natomiast przedmiot analizy zawężono do Działań 2.3, 2.4 oraz 2.5. Zmienione podejście zostało przyjęte i zaakceptowane przez Zlecającego.

Tabela 2 przedstawia przegląd ogólnych i uszczegółowionych pytań ewaluacyjnych wraz z metodologią wykorzystaną do udzielenia na nie

Tabela 2. Pytania ewaluacyjne i metody zbierania danych

Główne kwestie	Szczegółowe pytania ewaluacyjne	Użyte metody
Analiza struktury projektów Priorytet II ZPORR	Jaka jest liczba i tematyka projektów realizowanych w ramach Priorytetu II ZPORR?	Analiza statystyczna danych MRR
	Jakie jest rozłożenie przestrzenne realizowanych działań w ramach Priorytetu II ZPORR?	Analiza statystyczna danych MRR Analiza przestrzenna – opracowanie map
	Jakie pojawiają się kategorie beneficjentów? Kto składa projekty?	Analiza statystyczna danych MRR
	Jaka jest wielkość realizowanych projektów?	Analiza statystyczna danych MRR
Analiza zgodności z zapisami dokumentów programowych	Czy realizowane projekty wpływają na rozwój regionu?	Wywiady z IW (WUP, IW, UM) Ankiety e-mailowe z beneficjentami, którzy podpisali umowy finansowania Wywiady telefoniczne (CATI) z wnioskodawcami odrzuconymi w procesie selekcji Wywiady z ekspertami Analiza ekspercka
	Czy realizowane projekty odpowiadają potrzebom rynku pracy na danym obszarze?	
	Czy zapisy znajdujące się w Priorytecie II ZPORR są trafne i odpowiadają potrzebom lokalnych i regionalnych rynków pracy?	

Analiza systemu zarządzania i identyfikacja barier	Jakie problemy pojawiają się na etapie identyfikacji i informowania beneficjentów? Jakie są ich przyczyny? Czy pojawiają się jakieś pomysły rozwiązania tych problemów?	Wywiady z IW (UM, WUP, IW, RIF) Wywiady z IZ ZPORR Ankiety mailowe z beneficjentami, którzy podpisali umowy finansowania Wywiady telefoniczne (CATI) z wnioskodawcami odrzuconymi w procesie selekcji *Desk research* (dokumenty programowe, sprawozdania monitoringowe) *Mindmapping*, schematy Wywiady z ekspertami
	Jakie problemy pojawiają się na etapie procesu selekcji? Jakie są ich przyczyny? Czy pojawiają się jakieś pomysły rozwiązania tych problemów?	
	Jakie problemy pojawiają się na etapie wdrażania projektów? Jakie są ich przyczyny? Czy pojawiają się jakieś pomysły rozwiązania tych problemów?	

Źródło: opracowanie własne za: Olejniczak K., Ledzion B., Domaradzka A. (2006), *Raport końcowy z ewaluacji pt. Identyfikacja barier w realizacji działań współfinansowanych ze środków Europejskiego Funduszu Społecznego w ramach Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego*, Warszawa 2006, Wydawnictwo MRR.

odpowiedzi. Jak widać, odpowiedzi na pytanie 1 udzielono głównie na podstawie danych statystycznych przekazanych przez Zamawiającego, natomiast w przypadku pozostałych dwóch pytań odpowiedzi udzielono na podstawie danych zebranych w trakcie przeprowadzonych badań ilościowych i jakościowych.

Po przeanalizowaniu pytań ewaluacyjnych oraz tematyki i obszaru badań, zdecydowano się na triangulację, czyli na wykorzystanie kilku komplementarnych metod badawczych. Miało to z jednej strony umożliwić zebranie możliwie szerokiej gamy informacji pozwalającej na ilościową obróbkę danych i formułowanie wniosków na podstawie jak największej liczby przypadków. Z drugiej strony, biorąc pod uwagę ograniczenia związane z małą grupą badanych organizacji, wynikające z tak, a nie inaczej postawionych pytań ewaluacyjnych (jak również krótki czas realizacji badania), postanowiono wykorzystać wszystkie zalety szczegółowej jakościowej analizy danych – przy czym dotyczyło to zarówno danych zastanych, jak i zebranych w trakcie ewaluacji.

Niezwykle ważną cechą łączenia różnych metod badawczych było podwyższenie wiarygodności zebranych w ten sposób danych. Każda z metod pozwalała badaczom na uzyskanie innej perspektywy na studiowane zagadnienie, a stosowane metody nawzajem się uzupełniały. Ponadto triangulacja metod pozwoliła zminimalizować słabe strony każdej z technik zbierania danych, a wykorzystać wszystkie ich zalety. W ten sposób możliwe było zgromadzenie szczegółowych i różnorodnych informacji dotyczących badanego zagadnienia, przy zachowaniu wiarygodności i rzetelności badawczej.

Wśród zastosowanych metod ilościowych znalazły się:

- ankiety e-mailowe z beneficjentami, którzy podpisali umowy finansowania („wygrani”),
- wywiady telefoniczne z wnioskodawcami odrzuconymi w procesie selekcji („przegrani”),
- analiza danych zastanych – statystyki IW i IŻ.

Oprócz tego zastosowano metody jakościowe, m.in.:

- pogłębione wywiady indywidualne,
- analizę danych zastanych – *desk research*,
- konsultacje ze specjalistami.

Wymienione metody stosowane były w odniesieniu do wszystkich etapów realizacji programu, tak by umożliwić diagnozę problemów powstających na różnych poziomach zarządzania i w różnych instytucjach, a także przeprowadzić analizę działań podejmowanych przez poszczególnych aktorów procesu związanego z wdrażaniem badanego programu. Na podstawie zebranych danych udało się zrekonstruować system przepływu informacji, zidentyfikować bariery powstające na poszczególnych etapach wdrażania, a także przyczyny problemów związanych z realizacją programu i zaproponować możliwe rozwiązania.

3.2. Jak przebiegał proces badawczy?

Proces zbierania danych realizowany był na trzech poziomach: Instytucji Zarządzającej, Instytucji Wdrażających oraz beneficjentów końcowych. Badanie w całości przeprowadzono na próbie celowej, zdefiniowanej pod kątem uzyskania jak najszerszego wachlarza informacji i umożliwiającej sformułowanie wyczerpujących odpowiedzi na pytania ewaluacyjne.

W pierwszym etapie wśród sześciu działań w ramach II priorytetu ZPORR, w porozumieniu ze Zleceniodawcą, wybrano trzy najbardziej problematyczne i wymagające analizy. Były to działania 2.3, 2.4 i 2.5 II priorytetu ZPORR. Następnie, na postawie statystyk gromadzonych przez IZ, dotyczących stopnia realizacji programu w poszczególnych województwach, wybrano próbę trzech województw. Województwa wybrano w taki sposób, by uzyskać możliwie najszersze spojrzenie na badane zagadnienia, a równocześnie spełnić rygorystyczne wymogi czasowe. Po dodatkowych konsultacjach z Zamawiającym zdecydowano się na realizację badania w województwach mazowieckim, wielkopolskim i opolskim. Opolskie zostało wybrane jako przykład województwa, w którym realizacja programu przebiegała zdecydowanie gorzej niż w innych województwach, wielkopolskie jako przykład sprawnej realizacji programu i źródło dobrych praktyk, zaś mazowieckie, jako mieszczące się w średniej i dające dodatkową możliwość przebadania większej liczby organizacji.

Tabela 3 zawiera informacje na temat zrealizowanej próby w odniesieniu do wielkości badanej populacji (określonej na podstawie danych przekazanych przez Zamawiającego). Jak łatwo zauważyć, początkowe założenia względem liczby respondentów nie zostały spełnione – przyczyny tego stanu rzeczy opisano w tabeli 3.

Na poziomie beneficjentów końcowych wykorzystano głównie badanie metodą ankietową, która pozwoliła na zebranie maksymalnej ilości szczegółowych informacji o wszystkich etapach uczestnictwa w programie od możliwie największej liczby organizacji z wybranych województw. Ci z beneficjentów, którzy podpisali umowy, otrzymali ankiety drogą mailową, a następnie byli monitowani telefonicznie. Natomiast ci wnioskodawcy, którzy odpadli na którymś z etapów i nie udało im się podpisać umowy, zostali zbadani przy wykorzystaniu wywiadu telefonicznego. Takie podejście badawcze wynikało z założenia, że „przegrani" wnioskodawcy będą mieli znacznie mniejszą motywację do wypełnienia i odesłania ankiet, niż ci, którzy uczestniczą w programie, więc bezpośredni kontakt telefoniczny stwarza większe szanse na otrzymanie odpowiedzi. Treść obu kwestionariuszy była prawie identyczna, z tym że ten skierowany do „przegranych" wnioskodawców był krótszy, bo nie obejmował kwestii dotyczących beneficjentów, którzy otrzymali dofinansowanie (związanych np. z podpisaniem umowy czy realizacją płatności). Dzięki zastosowaniu tych samych pytań w obu kwestionariuszach,

Tabela 3. Realizacja badania

Metoda i badana grupa	Próba założona	Próba zrealizowana	Wielkość badanej populacji	Uwagi
Ankiety mailowe z beneficjentami, którzy podpisali umowy finansowania	50	**32**	80	Mała liczba beneficjentów, którzy podpisali umowy finansowania, mała liczba kontaktów. Respondenci niechętnie odpowiadali na ankiety, szczególnie przy narzuconych krótkich terminach oddania wypełnionych kwestionariuszy. Respondenci byli trudno dostępni (wielu padło ofiarą grypy), co – biorąc pod uwagę krótki termin realizacji badania – uniemożliwiło dotarcie do wszystkich respondentów.
Wywiady telefoniczne (CATI) z wnioskodawcami odrzuconymi w procesie selekcji	25	**20**	50	Baza kontaktów przekazana przez Zamawiającego była niepełna i dostarczona bardzo późno. Część beneficjentów nie chciała odpowiadać na pytania ankiety. W większości przypadków przekazane nam dane adresowe dotyczyły beneficjentów odrzuconych na etapie oceny merytorycznej. Potencjalni beneficjenci odrzuceni na etapie oceny formalnej byli niedoreprezentowani. Respondenci byli trudno dostępni (wielu padło ofiarą grypy), co – biorąc pod uwagę krótki termin realizacji badania – uniemożliwiło dotarcie do wszystkich respondentów.
Wywiady indywidualne	IW = 7	7	–	–
	IZ = 4	4	–	–

Źródło: opracowanie własne za: Olejniczak K., Ledzion B., Domaradzka A. (2006), *Raport końcowy z ewaluacji pt. Identyfikacja barier w realizacji działań współfinansowanych ze środków Europejskiego Funduszu Społecznego w ramach Zintegrowanego Programu Operacyjnego Rozwoju* Regionalnego, Warszawa 2006, Wydawnictwo MRR.

możliwe było późniejsze połączenie zbiorów danych, tak by część analiz statystycznych realizowana była na próbie wszystkich respondentów. Ten zabieg miał na celu poprawę wiarygodności wyników, szczególnie w przypadku analiz w podziale na poszczególne działania.

Na poziomie Instytucji Zarządzającej i Instytucji Wdrażających wykorzystano przede wszystkim metody jakościowe (wykorzystanie metod ilościowych na tak małej próbie mijałoby się z celem), czyli wywiady indywidualne z przedstawicielami IZ i IW, a także analizę danych zastanych (*desk research*). Jakościowa analiza danych zastanych oparta była na raportach monitorujących (z dwóch poziomów – projektu i działania w danym województwie), protokołach z posiedzeń komitetów monitorujących, wcześniejszych raportów ewaluacyjnych, w tym także raportów ewaluacyjnych z badań o podobnej tematyce[5]. Analiza jakościowa uzupełniona została wnioskami z badań ilościowych – na podstawie danych dotyczących liczby aplikujących beneficjentów, podpisanych umów i realizowanych projektów, przygotowano zestawienia i analizy w formie map i tabel.

Dodatkowo wykorzystane zostały informacje zebrane na podstawie nieformalnych wywiadów z: pięcioma beneficjentami końcowymi; sześcioma osobami na stanowiskach kierowniczych w WUP; czterema pracownikami szeregowymi WUP oraz kierownikiem projektu szkoleń skierowanych do bezrobotnych i zagrożonych bezrobociem finansowanych z PHARE RZL 2001 (jako przykład dobrej praktyki). Przeprowadzono również dodatkowe konsultacje z psychologami, mające pomóc w odpowiedzi na pytanie o źródła problemów związanych z rekrutacją beneficjentów ostatecznych.

3.3. Jakie były wyzwania i ograniczenia procesu badawczego?

W trakcie badania zespół ewaluacyjny natrafił na wiele problemów, które w pewnym stopniu utrudniły realizację założeń. Biorąc pod uwagę zakres problematyki oraz pytań badawczych, termin realizacji badania był bardzo krótki. Dodatkowo wystąpiły trudności z uzyskaniem danych z monitoringu, a także listy kontaktów do beneficjentów, co spowodowało dodatkowe opóźnienia w realizacji badania.

Badanie w założeniu miało (z uwagi na przeznaczony czas i budżet) charakter jakościowy i punktowy – oparty na ograniczonej liczbie przypadków. Na wyraźne życzenie klienta ewaluatorzy postarali się nadać badaniu charakter ilościowy i zmaksymalizować liczbę zrealizowanych ankiet. Ostatecznie, rozmiar zrealizowanej próby był stosunkowo duży w porównaniu z rozmiarem populacji (zastosowano dodatkowe monity, by uzyskać jak najwyższą stopę zwrotu) i, zdaniem badaczy, wystarczający dla zbadania zagadnień objętych oceną. Najlepszym na to dowodem

[5] Ewaluacja *ex post* Programu Phare RZL 2000 i 2001.

był fakt, że wskazywane przez beneficjentów problemy i bariery w większości ankiet powtarzały się, co świadczy o tym, że udało się wyczerpać problematykę poruszaną w pytaniach.

W początkowej fazie realizacji badania badacze opierali się na bazie danych dostarczonej przez Ministerstwo Rozwoju Regionalnego – IZ PWW. Baza ta okazała się jednak niejednolita i niekompletna, a ponadto przedstawiała stan na dzień 31 lipca 2005 roku. W bazie tej zabrakło usystematyzowanych i szczegółowych danych dotyczących tematyki realizowanych projektów, a dane dotyczące typów beneficjentów i alokacji przestrzennej środków były niekompletne lub dzielone według niekonsekwentnie stosowanych kategorii. Z tych powodów, po uzgodnieniu z Zamawiającym, zadecydowano przeprowadzić badanie na podstawie danych ze sprawozdań monitoringowych z realizacji II priorytetu ZPORR w poszczególnych województwach według stanu na dzień 31 grudnia 2005 roku. Taka operacja ograniczyła zakres analizowanych danych tylko do danych na temat liczby podpisanych umów oraz wartości projektów; trzeba było natomiast zrezygnować z analizy jakościowej, uwzględniającej typy projektów czy rodzaj beneficjenta.

3.4. Jaką rolę odegrali interesariusze w procesie badawczym?

Poziom zaangażowania poszczególnych interesariuszy w proces badawczy był bardzo zróżnicowany, co nie pozostało bez wpływu na późniejszy odbiór badania. Zespół Krajowej Jednostki Oceny był na bieżąco zaangażowany w realizację oceny, wspomagając ewaluatorów swoją wiedzą i dając wskazówki co do jego optymalnej realizacji. Stała współpraca z KJO, zarówno przy określaniu metodologii, opracowywaniu narzędzi, jak i – w końcu – komentowaniu otrzymanych wyników, pozytywnie wpłyneła na adekwatność przeprowadzonej oceny oraz udane zakończenie projektu. Pozostali interesariusze, czyli Instytucja Zarządzająca ZPORR oraz Instytucje Wdrażające z regionów ograniczyły swoj udział w ewaluacji wyłącznie do uczestnictwa w wywiadach i dostarczania danych, natomiast na etapie realizacji nie interesowały się wynikami badania. Brak zaangażowania w proces badawczy, a co za tym idzie – niezrozumienie toku myślenia ewaluatorów, były jednymi z głównych przyczyn późniejszej negatywnej reakcji na przedstawione wyniki.

4. PRODUKTY BADANIA

4.1. Jaką wiedzę „wyprodukowano" w badaniu?

Głównym produktem badania był raport z ewaluacji, podzielony na trzy główne części, z których każda odnosiła się do jednego z zagadnień badawczych. Oprócz tego Wykonawcy przedstawili w raporcie (w formie

aneksów) dodatkowe materiały pomocnicze, m.in. mapy pokazujące zróżnicowanie regionalne w realizacji działań II priorytetu ZPORR oraz propozycję struktury strony internetowej programu. Przy wykorzystaniu schematów i map (mindmappingu) przygotowano przejrzyste podsumowania wyników, ilustrujące zidentyfikowane problemy, ich przyczyny oraz stopień zaawansowania prac nad ich rozwiązywaniem. Dodatkowo w raporcie zamieszczono zebrane od ekspertów (psychologów i socjologów) opinie na temat sytuacji i zachowań beneficjentów ostatecznych. Aby uatrakcyjnić formułę raportu, umieszczono w nim cytaty z wypowiedzi respondentów, ilustrujące opisywane problemy.

Jednym z głównych problemów zidentyfikowanych na podstawie analizy systemu wdrażania było niedostosowanie realizowanych działań do potrzeb rynku, co przekładało się na brak zainteresowania wnioskodawców. Grupy docelowe beneficjentów ostatecznych były w przypadku niektórych działań nieadekwatne do sytuacji w poszczególnych regionach (przykładem mogą być problemy z realizacją projektów skierowanych do rolników na terenie Śląska). To pociągnęło za sobą problemy z naborem projektów w ramach niektórych działań, a co za tym idzie – z realizacją założeń programu. Taka sytuacja była efektem braku analiz lokalnych rynków pracy, w tym potrzeb szkoleniowych.

Ważnym elementem analizy barier systemowych było również zaproponowanie metod przezwyciężania zidentyfikowanych barier i problemów. Ewaluatorzy starali się o jasne przedstawienie operacyjnego, kompleksowego pakietu użytecznych rekomendacji. Silną stroną przygotowanego raportu było to, że każda z rekomendacji odnosiła się do konkretnej przyczyny problemu. Ponadto propozycje rozwiązań zawarte w raporcie opierały się przede wszystkim na (zweryfikowanych przez ewaluatorów) sugestiach samych uczestników II priorytetu ZPORR – a więc osób kompetentnych i dobrze znających system wdrażania. Zaproponowane rekomendacje uwzględniały warunki instytucjonalne, a ich zawartość, mimo ograniczeń czasowych badania, została przedyskutowana z personelem zajmującym się poszczególnymi Działaniami II priorytetu ZPORR. Słabą stroną opisywanego badania były natomiast ograniczenia czasowe i finansowe, które wymusiły okrojenie zakresu badania oraz brak aktualnych szczegółowych danych umożliwiających pogłębiony opis działania systemu wdrażania.

Aby zmaksymalizować użyteczność ewaluacji, zastosowano zasadę partnerstwa i konsultacji ze zlecającym na każdym etapie realizacji badania. Wszystkie rekomendacje były poddawane dyskusji zarówno z Zamawiającym, jak i innymi uczestnikami procesu ewaluacji. Takie podejście miało na celu sformułowanie użytecznych zaleceń, mających realne szanse na wdrożenie. Innymi słowy, ewaluatorzy starali się swoim raportem stworzyć merytoryczną bazę do dyskusji nad podniesieniem efektywności systemu wdrażania projektów finansowanych z EFS.

4.2. Jak wyglądał proces upowszechniania wyników?

Raport został wydany przez Ministerstwo Rozwoju Regionalnego w 2006 roku i rozesłany do różnego rodzaju instytucji zajmujących się wdrażaniem EFS (np. Regionalnych Ośrodków EFS w całej Polsce). Jest również dostępny w internecie.

Wyniki raportu były publicznie prezentowane na spotkaniach Grupy Sterującej, w skład której wchodzili przedstawiciele departamentów MRR oraz regionów. Taka sytuacja dawała szansę na szerszą dyskusję nad wnioskami i rekomendacjami zawartymi w raporcie, szczególnie z przedstawicielami poszczególnych IW z regionów. Ze względu na obszerność raportu, nie wszyscy uczestnicy spotkań zdążyli się z nim zapoznać, nie można więc było mówić o ożywionej dyskusji. To doświadczenie pokazuje więc, jak ważnym elementem efektywnego rozpowszechniania wyników jest przygotowanie streszczenia raportu.

5. WNIOSKI ZE STUDIUM PRZYPADKU

Podsumowując, punktem wyjścia opisanej ewaluacji był stan wdrażania priorytetu II Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego. Jego realizacja miała na celu zmniejszenie poziomu bezrobocia w Polsce przez działania kierowane do osób zagrożonych bezrobociem, szczególnie w sektorze rolnictwa i przemysłu tradycyjnego. Głównym zadaniem ewaluatorów było wskazanie źródeł istniejących problemów we wdrażaniu działań finansowanych z EFS w ramach ZPORR oraz zaproponowanie rozwiązań racjonalizatorskich. Ewaluacja miała charakter oceny uzupełniającej i przyjęła formę ewaluacji *on-going*. Badanie w całości przeprowadzono na próbie celowej, zdefiniowanej pod kątem uzyskania jak najszerszego wachlarza informacji. Zastosowano zarówno jakościowe, jak i ilościowe metody zbierania danych, a więc: ankiety mailowe, wywiady telefoniczne, pogłębione wywiady indywidualne, analizę danych zastanych i konsultacje ze specjalistami. Zaproponowane rekomendacje miały charakter konkretnych zaleceń i uwzględniały warunki instytucjonalne, a ich treść została przedyskutowana z personelem zajmującym się poszczególnymi Działaniami II priorytetu ZPORR. Raport ewaluacyjny koncentrował się przede wszystkim na ocenie systemu jako całości, oraz propozycji działań naprawczych, mających na celu poprawę efektywności wdrażania programu.

Trudno ocenić rzeczywistą użyteczność opisanej tu ewaluacji, ze względu na duży opór interesariuszy badania przed przyjęciem wyników i wniosków z raportu. Już wstępne rozmowy związane z przeprowadzaną ewaluacją pozwoliły stwierdzić, że budzi ona duże kontrowersje

wśród instytucji odpowiedzialnych za wdrażanie i realizację programu. Taka sytuacja wynikała prawdopodobnie z faktu, że większość badanych po raz pierwszy uczestniczyła w procesie ewaluacji i postrzegała ją krytycznie. Dlatego od samego początku ewaluatorzy starali się podkreślić, że badanie nie ma charakteru audytu ani kontroli. Raport ewaluacyjny koncentrował się na ocenie systemu jako całości, a nie osób lub instytucji potencjalnie odpowiedzialnych za problemy (co zresztą wynika z zasad kodeksu etycznego, który dotyczy wszystkich badań, w tym badań ewaluacyjnych). Celem zawartych w raporcie spostrzeżeń miało być przede wszystkim podjęcie konkretnych działań naprawczych w ramach istniejącego systemu wdrażania. Opisane problemy nie obciążały żadnej konkretnej instytucji, mogły natomiast stanowić podstawę do występowania o dodatkowe środki czy etaty.

Choć w swoich założeniach i wymowie raport nie miał na celu formułowania zarzutów – ani instytucjonalnych, ani tym bardziej personalnych – spotkał się z dużą krytyką ze strony instytucji objętych oceną. Można odnieść wrażenie, że krytyczny wydźwięk raportu (wynikający z koncentracji na zagadnieniu problemów i barier) i związane z tym kontrowersje w pewnym stopniu przesłoniły jego praktyczne zastosowanie. Z perspektywy czasu można jednak stwierdzić, że obecnie świadomość istnienia problemów zidentyfikowanych w trakcie opisywanego tu badania jest znacznie wyższa i że w pewnym sensie przygotowywany raport przetarł ścieżki dla następnych działań ewaluacyjnych.

Najpoważniejszym ograniczeniem, któremu zespół ewaluacyjny musiał stawić czoła, był krótki czas realizacji badania i brak danych umożliwiających pełną realizację badań ilościowych. Dodatkowym obciążeniem był termin realizacji (koniec roku) – a więc moment dużego obciążenia rozliczeniami zarówno IW, jak i beneficjentów. Dzięki zastosowaniu triangulacji danych można jednak przyjąć, że zbadanie całej populacji nie zmieniłoby znacząco wyników ewaluacji. Kolejnym problemem była fragmentaryczność i niespójność danych z bazy projektów ZPORR, które spowodowały, że badacze musieli skoncentrować się głównie na analizie danych ze sprawozdań monitoringowych za IV kwartał 2005 roku. Na pewno uzyskanie bardziej wyczerpujących informacji o beneficjentach umożliwiłoby pełniejszą analizę badanych zagadnień.

W celu zminimalizowania zagrożeń związanych z napiętym harmonogramem, od początku projektu Wykonawca prowadził z Zamawiającym intensywne konsultacje, co pozwoliło między innymi na określenie kwestii priorytetowych dla Zamawiającego i skoncentrowanie się właśnie na nich. Takie podejście stanowi, zdaniem autorki, przykład dobrej praktyki, dowodzącej, że nawet w niesprzyjających warunkach możliwe jest zakończenie ewaluacji z sukcesem, jeżeli istnieje stała współpraca między Zamawiającym a Wykonawcą. Proces konsultacji umożliwił

Zamawiającemu przeprowadzenie pogłębionej analizy własnych potrzeb, co z pewnością poprawiło adekwatność badania, a w przyszłości mogło przełożyć się na lepsze planowanie kolejnych ewaluacji.

Podsumowując, mimo opisanych przeszkód, przeprowadzone badanie istotnie wpłynęło na poprawę wdrażania programu oraz usprawnienie procesów zarządzania wydatkowaniem środków II priorytetu ZPORR. Choć nie bez trudności, ewaluacja spełniła więc swoje zadanie.

ROZDZIAŁ 4.4.

Maciej Szałaj
Bartosz Ledzion

EWALUACJA BIEŻĄCA SIECI REGIONALNYCH OŚRODKÓW EUROPEJSKIEGO FUNDUSZU SPOŁECZNEGO

Rozdział poświęcony jest omówieniu badania ewaluacyjnego pt. *Ewaluacja bieżąca sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego* jako przykładu ewaluacji *on-going*. Opisany przypadek przedstawia przebieg jednej z pierwszych ewaluacji sieci organizacji w Polsce, a co za tym idzie – trudności związanych ze zdobywaniem danych, realizacją badań terenowych oraz nowymi doświadczeniami – zarówno po stronie Zamawiającego, jak i Wykonawcy. Opisany projekt badawczy stanowi interesujące studium przypadku, prezentuje nowatorskie metody badawcze, opisuje słabe i mocne strony procesu badawczego, jednocześnie wskazując przyjęte rozwiązania i działania naprawcze. Lektura poniższego tekstu może być szczególnie interesująca dla osób prowadzących badania związane z oceną organizacji działających formalnie i nieformalnie w ramach sieci współpracy, takich jak klastry, partnerstwa, konsorcja, związki.

1. KONTEKST I PRZEDMIOT EWALUACJI

1.1. Opis przedmiotu badania

Z uwagi na zakres tematyczny możemy mówić w tym przypadku o ewaluacji bieżącej projektu systemowego *Budowa sieci ROSZEFS finansowanego z pomocy technicznej SPO RZL (2004–2006)*, aczkolwiek pojęcie „projektu" należy tutaj rozumieć nieco szerzej, bowiem zakres projektu był bardzo szeroki – dotyczył oceny działalności sieci 49 Regionalnych Ośrodków EFS. Budżet projektu przekraczał kwotę 20 mln PLN.

Najważniejszą przesłanką stojącą za utworzeniem sieci były zaobserwowane problemy i ograniczenia w absorpcji Europejskiego Funduszu Społecznego (EFS), których przyczyn upatrywano w słabości instytucjonalnej i braku wystarczających kompetencji po stronie uprawnionych projektodawców. Przyjęto wówczas założenie, iż najbardziej adekwatną formą działania będzie dostarczenie usług szkoleniowych możliwie najszerszym kręgom instytucji uprawnionych do ubiegania się o dofinansowanie projektów ze środków EFS. Oferta ośrodków wchodzących w skład sieci była stopniowo wzbogacana o elementy doradztwa i działalności animacyjnej, ukierunkowanej na pobudzanie partnerskiej współpracy pomiędzy przedstawicielami różnorodnych środowisk lokalnych. W rezultacie, wraz z rozpoczęciem trzeciej edycji działalności, odstąpiono od pierwotnej nazwy sieci (Regionalne Ośrodki Szkoleniowe EFS), tak aby podkreślić zróżnicowany charakter oferty oraz funkcji pełnionych przez ośrodki.

Sieć Regionalnych Ośrodków EFS (ROEFS) stanowi obecnie istotny element otoczenia instytucjonalnego, zarówno szeroko rozumianej polityki społecznej, jak i polityki edukacyjnej czy kulturalnej. Znaczenie Regionalnych Ośrodków EFS wynika również z faktu, iż program, w ramach którego prowadzą one swoją działalność, jest jednym z największych polskich przedsięwzięć publicznych, ukierunkowanych na tworzenie trwałych podstaw dla rozwoju kapitału społecznego wśród społeczności lokalnych oraz dostarczanie niezbędnej pomocy merytorycznej kluczowym instytucjom społeczeństwa obywatelskiego (lokalne organizacje pozarządowe i placówki oświatowe).

Według danych Krajowego Ośrodka EFS, odgrywającego rolę koordynatora sieci, z usług doradczych oferowanych przez ROEFS skorzystało 53,5 tys. klientów, a ośrodki udzieliły wsparcia przy opracowywaniu lub realizacji 1711 projektów współfinansowanych z EFS.

Realizacja badania została zlecona przez Zamawiającego – Polską Agencję Rozwoju Przedsiębiorczości (PARP) – wykonawcy zewnętrznemu, wyłonionemu w postępowaniu przetargowym. Dzięki temu omawiana ewaluacja miała charakter zewnętrzny.

1.2. Dlaczego podjęto badanie?

Moment rozpoczęcia ewaluacji z punktu widzenia cyklu wdrażania programu należy umieścić w jego środkowej fazie, a realizację ewaluacji w fazie środkowej i końcowej. Ulokowanie badania w tak zakreślonym kontekście czasowym wynikało bezpośrednio z potrzeb powstałych w trakcie zarządzania siecią ROEFS przez PARP. W kontekście zaplanowanych na pierwszą połowę 2008 roku procedur wyłaniania instytucji odgrywających rolę ROEFS w ramach kolejnej edycji funkcjonowania sieci, zarysowała się potrzeba dokonania oceny bieżącego funkcjonowania ośrodków, by móc zidentyfikować przypadki, w których dane instytucje nie posiadały wystarczającego potencjału organizacyjnego do należytego wykonania powierzonego im zadania.

Dla tworzenia budżetu dla kolejnej edycji ROEFS istniała również potrzeba przeprowadzenia pogłębionych analiz wartości wskaźników produktów i rezultatu, osiąganych przez ośrodki. Projektując ewaluację, założono, iż tego rodzaju analizy pozwolą na oszacowanie realnego poziomu celów ilościowych, przypisywanych poszczególnym ROEFS, a przez to na bardziej racjonalne planowanie wielkości i struktury budżetów przyznawanych na prowadzenie ośrodków.

Trzecia edycja działalności sieci ROEFS była pierwszą, w ramach której sieć objęła tak dużą liczbę ośrodków. Dla większości instytucji prowadzących ośrodki były to pierwsze doświadczenia w prowadzeniu tego rodzaju projektów, dlatego też omawiana ewaluacja miała za zadanie dostarczyć możliwie pilnie także dobrych praktyk, które mogłyby znaleźć bezpośrednie zastosowanie w zarządzaniu ośrodkami oraz ewentualnym planowaniu działań ośrodka.

1.3. Kim byli kluczowi interesariusze?

Ewaluacja sieci ROEFS została zaprojektowania we współpracy dwóch komórek organizacyjnych PARP – Krajowego Ośrodka EFS (zarządzającego siecią ROEFS) oraz Sekcji Ewaluacji (wyspecjalizowanej komórki ewaluacyjnej, niezależnej od struktur odpowiadających z wdrażanie programów). Rola przedstawicieli KOEFS skoncentrowana była na wskazywaniu głównych potrzeb informacyjnych oraz formułowaniu oczekiwań praktycznych wobec projektowanej ewaluacji. Przedstawiciele Sekcji Ewaluacji odpowiadali natomiast za poprawność warsztatową projektu, przekładając potrzeby informacyjne instytucji na konkretne pytania badawcze, a następnie określając ramy metodologiczne badania.

Do pozostałych interesariuszy badania można zaliczyć wszystkie instytucje uczestniczące w systemie wdrażania Europejskiego Funduszu Społecznego, zaczynając od Instytucji Zarządzającej EFS, instytucji pośredniczących (Urzędy Marszałkowskie, Wojewódzkie Urzędy Pracy), po

organizacje odpowiedzialne za prowadzenie ośrodków. Udział tych podmiotów w badaniu ograniczał się głównie do roli respondentów w prowadzonych badaniach terenowych, jednakże należy podkreślić fakt, że instytucje te należą do kluczowych odbiorców rekomendacji. Ich zaangażowanie w proces badawczy było kluczowym czynnikiem powodzenia projektu.

1.4. Jakie założenia metodologiczne przedstawiono w Opisie Przedmiotu Zamówienia?

Głównym celem Ewaluacji sieci ROEFS była całościowa ocena funkcjonowania systemu Regionalnych Ośrodków Europejskiego Funduszu Społecznego. Ocena koncentrowała się głównie na trzeciej edycji projektu, jednakże eksperci w swoich pracach uwzględniali także wyniki poprzednich dwóch edycji.

Realizację projektu zaplanowano w dwóch fazach, obejmując zarówno badanie bieżącej działalności ośrodków EFS, jak i analizę oraz podsumowanie poprzednich edycji programu.

Założono, że w obu fazach realizacji projektu Wykonawca dokona oceny, opierając się na następujących kryteriach ewaluacyjnych:

- skuteczność działań prowadzonych przez Ośrodki, rozumiana jako osiąganie celów założonych w projektach,
- skuteczność działania sieci Ośrodków EFS, rozumiana jako przygotowanie potencjalnych beneficjentów do korzystania za środków EFS i zwiększenie absorpcji środków EFS w Polsce,
- adekwatność działań proponowanych przez Ośrodki do potrzeb klientów i potencjalnych projektodawców EFS oraz celów programowych,
- efektywność działań Ośrodków,
- wpływ działań prowadzonych przez Ośrodki na beneficjentów tych działań oraz na społeczności lokalne.

Założono, że procesie ewaluacji Wykonawca powinien wykorzystać zestaw narzędzi mających charakter ilościowy i jakościowy, obejmujący przede wszystkim:

- analizę danych zastanych
- pogłębione wywiady indywidualne,
- badania ankietowe (CATI),
- analizę SWOT sieci,
- analizę grup porównawczych,
- studia przypadków.

Na etapie procesu przetargowego Zamawiający określił zestaw kilkudziesięciu pytań badawczych, które pogrupowano według poniższych zagadnień ewaluacyjnych:

- W jakim stopniu funkcjonowanie ośrodków ROEFS wspomaga realizację celów EFS w Polsce?

- Jaki jest poziom i zakres oddziaływania ośrodków ROEFS?
- Na ile długotrwałe są efekty działalności ośrodków ROEFS?
- Jakie są ekonomiczne skutki działalności sieci ROEFS?

1.5. Jakie postawiono warunki operacyjne?

Zamawiający założył, że realizacja badania sieci ROEFS potrwa 8 miesięcy, obejmując okres od września 2007 roku do czerwca 2008 roku. Budżet badania, założony przez Zamawiającego wynosił około 300 tys. PLN.

W celu określenia budżetu i harmonogramu badania Zamawiający skorzystał również z niestandardowego rozwiązania w postaci ekspertyzy zewnętrznej, mającej na celu weryfikację wykonalności przyjętych przez Zamawiającego założeń badawczych oraz wypracowanie ostatecznej wersji projektu ewaluacji.

Odnośnie do wymogów formalnych na etapie postępowania przetargowego Zamawiający wymagał od oferentów referencji wskazujących na doświadczenie w zakresie realizacji ogólnopolskich badań ilościowych oraz badań ewaluacyjnych.

Ponadto każdy z oferentów był zobligowany przedstawić zespół badawczy złożony z personelu o dużym potencjale i doświadczeniu naukowo-badawczym, w tym: kierownika projektu, eksperta ds. badań ilościowych, eksperta ds. badań jakościowych oraz eksperta ds. organizacji pozarządowych.

2. PROCES BADAWCZY

2.1. Jakie podejście badawcze i metodologię przyjął ewaluator?

Głównymi celami ewaluacji sieci ROEFS były: 1) całościowa ocena funkcjonowania systemu ROEFS, 2) wypracowanie rekomendacji wskazujących zalecane kierunki rozwoju sieci ROEFS, 3) wypracowanie metodologii wyznaczania celów ilościowych na poziomie ROEFS.

Realizacja powyższych celów wymagała obrania określonych podejść badawczych. Dominującą rolę należy w omawianym przypadku przypisać podejściu zorientowanemu na cele oraz podejściu zorientowanemu na procesy. Pierwsze z wymienionych podejść odzwierciedlało bezpośrednio oceniane elementy ewaluacji sieci ROEFS. Na etapie realizacji badań terenowych wyrazem tego podejścia było zastosowanie badań ankietowych na próbie 4800 respondentów, prowadzonych wśród grup docelowych odbiorców usług ROEFS (jednostki samorządu terytorialnego, organizacje pozarządowe, instytucje szkoleniowe, instytucje rynku pracy). Badania te dostarczyły informacji pozwalających na ocenę użyteczności usług oferowanych przez ROEFS oraz wniosły informacje

kluczowe dla oceny wpływu ROEFS na korzystanie przez klientów z sieci z dofinansowania EFS i dla oceny skuteczności działań informacyjno-promocyjnych prowadzonych przez ośrodki.

Z uwagi na szeroki zakres oceny sieci ROEFS, obejmujący poza oceną efektów badanego programu szereg zagadnień związanych ze sprawnością organizacyjną sieci ROEFS oraz poszczególnych ośrodków, funkcje oceny realizowane były również w oparciu o podejście skoncentrowane na procesach. Na poziomie zastosowanych metod podejście to widoczne było w wywiadach indywidualnych, prowadzonych z przedstawicielami kluczowych instytucji (PARP, MRR, Komisja Europejska) oraz w studiach przypadku, realizowanych przez wizyty przedstawicieli zespołu ewaluacyjnego w wytypowanych ośrodkach, podczas których dokonywano analizy dokumentacji, przeglądu zaplecza infrastrukturalnego ośrodków, a także przeprowadzano wywiady z pracownikami ośrodków.

W tym miejscu warto zwrócić uwagę na interesujące podejście Wykonawcy (matrycy ośrodków) do wyboru ośrodków do pogłębionej analizy studium przypadku oraz wypracowania metodologii benchmarkingu i wskazania sposobów ustalania celów ilościowych. W pierwotnym założeniu narzędzie o nazwie *Matryca Ośrodków* miało powstać z zestawienia czynników o charakterze wewnętrznym, należących do cech jednostkowych ośrodków oraz czynników określających warunki otoczenia.

W grupie czynników wewnętrznych miały zostać zbadane takie cechy jak: zarządzanie personelem, jakość świadczonych usług, wielkość przychodów, struktura i poziom zatrudnienia.

Wśród czynników określających potencjał danego rynku, na którym działa ośrodek, przeanalizowane miały zostać m.in.: atrakcyjność inwestycyjna, stopa bezrobocia, wielkość i struktura nakładów inwestycyjnych oraz potencjał instytucjonalny.

Dodatkowo dla każdego wskaźnika miała zostać określona ocena punktowa w skali od 0 do 3. Określenie skali ocen planowano uszczegółowić w momencie zebrania i przetworzenia wszystkich danych dotyczących wskaźników, tak aby można było odnieść punktację do maksymalnej, średniej i najwyższej zarejestrowanej wartości. Ponadto dla każdego wskaźnika planowano przypisanie indywidualnych wag, świadczących o stopniu jego istotności z perspektywy działalności ośrodka.

Przetworzenie danych o wskaźnikach określających czynniki wewnętrzne i zewnętrzne miało stanowić podstawę do oszacowania potencjału ośrodka oraz danego rynku, wpisując tym samym ośrodek w konkretny obszar matrycy.

Podejście zorientowane na procesy wyrażało również drugi z głównych celów omawianej ewaluacji, czyli wypracowanie rekomendacji wskazujących kierunki rozwoju sieci ROEFS. Co istotne, podejście to uzupełniało się w tym aspekcie z podejściem zorientowanym na cele, albowiem proces wypracowywania rekomendacji wspierany był szeroką gamą danych zgromadzonych w badaniach terenowych.

Ostatni z głównych celów ewaluacji, ze względu na zakładany wyjściowo ilościowy charakter analiz, realizowany był w całości w ramach podejścia zorientowanego na cele.

Oprócz omówionych dwóch głównych podejść zastosowanych w ewaluacji sieci ROEFS, wskazać należy podejście responsywne, widoczne chociażby w zastosowanych w ramach studiów przypadków ankietach samooceny, wypełnianych przez koordynatorów badanych ośrodków. Zastosowanie podejścia responsywnego pozwoliło na pozyskanie cennych informacji, uzupełniających ocenę funkcjonowania sieci przez włączenie perspektywy kluczowych aktorów ocenianego systemu.

Pakiet metod i narzędzi badawczych, które zostały wykorzystanie w trakcie realizacji badania to:

- analiza danych zastanych z 3 edycji projektu;
- badania ankietowe (CATI) na łącznej próbie 4800 respondentów;
- pogłębione wywiady indywidualne z kluczowymi osobami na poziomie sieci ośrodków EFS (KOEFS, PARP, MRR, KE) – ok. 10 respondentów;
- badania ankietowe z przedstawicielami ROEFS – 49 respondentów;
- pogłębione wywiady indywidualne z pracownikami ROEFS (w tym samoocena organizacji) – ok. 250 respondentów;
- wywiady telefoniczne z przedstawicielami Instytucji Pośredniczących – 32 respondentów;
- analizy studium przypadku na 30 ośrodkach;
- analiza porównawcza;
- analiza SWOT.

2.2. Jak przebiegał proces badawczy?

Interakcje pomiędzy KOEFS a komórką ewaluacyjną PARP oraz zespołem ewaluacyjnym miały charakter regularny, przybierając formę konsultacji prowadzonych telefonicznie lub formę spotkań roboczych. W trakcie całego okresu realizacji badania miało miejsce kilkanaście spotkań, poświęconych ustalaniu oraz weryfikacji założeń kluczowych dla poszczególnych etapów. W spotkaniach po stronie Wykonawcy uczestniczył zazwyczaj lider zespołu ewaluacyjnego oraz członkowie zespołu, odpowiedzialni merytorycznie za dane zagadnienie (np. badania ilościowe, badania jakościowe, specyfika sektora organizacji pozarządowych).

Płaszczyzną komunikacji między interesariuszami były ponadto interakcje pomiędzy KOEFS a poszczególnymi ośrodkami sieci, w ramach których przekazywano informacje zapowiadające realizacje badań

terenowych przez zespół ewaluacyjny. Wartością dodaną interakcji przebiegających na wspomnianej płaszczyźnie było bez wątpienia zapewnienie komfortu pracy badaczy terenowych, którzy, odwiedzając ośrodki, spotykali się z większą gotowością udziału w badaniach oraz ze zrozumieniem dla podstawowych celów omawianej ewaluacji.

Odpowiednio zaplanowana komunikacja pomiędzy Zamawiającym (zarówno komórką ewaluacyjną PARP, jak i KOEFS) a Wykonawcą, stworzyła we wstępnej fazie badania podstawy dla sumiennych i dokładnych analiz danych zastanych oraz dla rzetelnej realizacji części terenowej. Cząstkowe efekty prac Wykonawcy były przekazywane Zamawiającemu, a następnie wspólnie konsultowane. W rezultacie Zamawiający miał bieżącą kontrolę merytoryczną nad badaniem, natomiast Wykonawca miał możliwość właściwego rozpoznawania oczekiwań klienta.

Oprócz wspomnianych mechanizmów zapewniania sumienności badawczej, które opierały się głównie na współpracy ewaluatorów z komórką ewaluacyjną PARP, wypracowane zostały również formy bieżącej komunikacji Wykonawcy z KOEFS, co przełożyło się nie tylko na dbałość o poprawność metodologiczną badania, ale pozwoliło także właściwie zadbać o dostosowanie ewaluacji do potrzeb zarządzania siecią ROEFS.

Na rzecz sumiennej i rzetelnej realizacji części terenowej przemawia natomiast wspólne wypracowywanie narzędzi badawczych (np. ankiet wywiadów telefonicznych) oraz poprzedzanie realizacji badań terenowych drobiazgowym pilotażem. Podczas części pilotażowych obecni byli zawsze przedstawiciele PARP oraz członek zespołu ewaluatorów, odpowiedzialny za dany etap procesu badawczego.

2.3. Jakie były wyzwania i ograniczenia procesu?

Jednym z kluczowych źródeł danych wykorzystywanych w ewaluacji sieci ROEFS była sprawozdawczość prowadzona na poziomie KOEFS oraz poszczególnych ośrodków sieci. Dane czerpane z tych źródeł stanowić miały podstawę dla przeprowadzenia benchmarkingu oraz wypracowania metod wyznaczania celów ilościowych dla ośrodków. W obszarze tym napotkano dość istotne ograniczenia, wśród których wymienić należy:

- brak spójnych, stosowanych w standardowy sposób przez wszystkie ośrodki definicji zmiennych wskaźnikowych,
- brak ujednoliconego i wystandaryzowanego zakresu danych sprawozdawczych,
- opóźnienie w przygotowaniu bazy danych, mające umożliwiać agregowanie danych sprawozdawczych na poziomie całej sieci ROEFS oraz pozyskanie danych kontaktowych klientów korzystających z usług ośrodków.

Efektem wskazanych ograniczeń było zawężenie analiz prowadzonych w ramach benchmarkingu do wąskiego zakresu danych z poziomu

produktu, które gromadzone były przez wszystkie ośrodki, według zbliżonego sposobu (spójna definicja zmiennych i sposób prezentacji danych). Nie udało się natomiast przenieść analiz na poziom rezultatów (np. liczba projektów przygotowanych przy wsparciu ośrodka, liczba partnerstw zawartych dzięki wsparciu ROEFS).

Kolejnym efektem ograniczeń dostępności danych był brak możliwości przeprowadzenia zaplanowanych pierwotnie badań ankietowych wśród klientów ośrodków ROEFS. W obliczu rozpoznanych ograniczeń podjęto decyzję o modyfikacji wyjściowych planów badawczych, w efekcie której przeprowadzone zostało badanie ankietowe na innej próbie (wnioskodawcy PO KL), którego zakres wpisywał się w tematykę omawianej ewaluacji oraz uwzględniał aktualne priorytety informacyjne Zamawiającego.

2.4. Jaką rolę odegrali poszczególni interesariusze w procesie badawczym?

Interesariuszem najsilniej zaangażowanym na wszystkich etapach ewaluacji był Krajowy Ośrodek EFS. W badania zostali także włączeni w różnych rolach wszyscy główni interesariusze ewaluowanego programu. Jak wspomniano, aktywny udział w badaniach terenowych przypadł pracownikom ośrodków wytypowanych do studiów przypadków. Zaangażowanie szerszych kręgów interesariuszy, w tym docelowych odbiorców usług ROEFS, uzyskano dzięki szeroko przeprowadzonym w ramach ewaluacji sieci ROEFS badaniom opinii docelowych odbiorców ewaluowanego programu. Mechanizmy włączania kluczowych interesariuszy zastosowano również na etapie dystrybucji wyników ewaluacji, które dostarczone zostały przedstawicielom wszystkich ośrodków sieci ROEFS oraz poddane konsultacjom w trakcie spotkań organizowanych przez PARP.

2.5. Produkty ewaluacji

Wymiernym produktem ewaluacji był raport otwarcia oraz 2 raporty z pierwszej i drugiej fazy badania. Raport otwarcia był poświęcony w dużej mierze prezentacji wniosków płynących z analizy danych zastanych dotyczących efektów działań ośrodków w ramach poprzednich dwóch edycji programu. Raport z pierwszej fazy badania dotyczył uwarunkowań III edycji działalności sieci ośrodków, prezentował analizę wyników badań terenowych oraz analiz porównawczych. Natomiast raport z drugiej fazy badania odnosił się głównie do analizy efektów funkcjonowania sieci Ośrodków EFS.

Proces ewaluacji przyniósł kilka rodzajów wiedzy: odkrywczą-spodziewaną, wiedzę koncepcyjno-techniczną, wiedzę diagnostyczną oraz preskryptywną.

2.6. Wiedza odkrywcza – spodziewana

Wyznaczenie granicy między wiedzą „odkrywczą" a „spodziewaną" jest zazwyczaj niezmiernie trudne. Uwidacznia się to szczególnie w przypadku badań biorących za przedmiot opinie oraz wyobrażenia określonych grup lub osób. Wiedza istnieje bowiem często w postaci nie do końca uświadomionej; wkład wnoszony przez ewaluację polega wówczas na usystematyzowaniu rozproszonych informacji. Podobną trudność zaobserwowano także w przypadku omawianej ewaluacji. Spostrzeżenie to odnosi się szczególnie do jakościowej części badania, zwłaszcza studiów przypadków. W trakcie prac nad raportem końcowym Zamawiający zwracał uwagę na nikłą wartość dostarczanych informacji. Można założyć, że w pewnym stopniu było to pochodną głębokiego zakorzenienia przedstawicieli Zamawiającego w kontekst badanego programu. Należy jednocześnie wskazać, iż analizując zgromadzone informacje, ewaluatorzy przyjęli zbyt mało krytycznie kategorie poznawcze właściwe dla badanych przez siebie pracowników ROEFS i innych instytucji. W efekcie zatarta została nieco perspektywa zewnętrznego obserwatora, co mogło rodzić wrażenie mało odkrywczego charakteru niektórych wyników ewaluacji.

2.7. Wiedza koncepcyjna – techniczna

Ewaluacja ROEFS, z uwagi na fakt, że miała służyć zarówno przekrojowej ocenie programu, jak i wypracowaniu konkretnych rozwiązań praktycznych, dostarczała wiedzy koncepcyjnej i technicznej. Wiedza koncepcyjna zgromadzona została na przykład w celu udzielenia odpowiedzi na pytań o dostosowanie ROEFS do sieciowej formy organizacji czy też w celu odtworzenia koncepcji funkcjonowania ROEFS typowej dla samych ośrodków oraz dla instytucji różnego szczebla. Wiedza techniczna znalazła natomiast zastosowanie w przypadku projektowania metod wyznaczania celów ilościowych oraz tworzenia konkretnych rekomendacji i zalecanych rozwiązań (np. narzędzie diagnostyczne opracowane, jako element rekomendacji dostarczonych w raporcie z I fazy badania).

2.8. Wiedza diagnostyczna

Część diagnostyczna omawianej ewaluacji skoncentrowana była na dostarczeniu wiedzy dwojakiego rodzaju – po pierwsze wiedzy kontekstowej, dającej przekrojowe rozpoznanie środowiska wdrażanego programu, po drugie – wiedzy dotyczącej mechanizmów funkcjonujących wewnątrz sieci ROEFS.

Diagnostyka otoczenia pozwoliła na rozpoznanie stopnia znajomości sieci ROEFS wśród grup potencjalnych klientów ROEFS, a także na

rozpoznanie zagadnień kluczowych z punktu wiedzenia problemów, którym w założeniu ROEFS miał przeciwdziałać (np. przyczyny braku korzystania z dofinansowania EFS).

Wiedza diagnostyczna w obszarze mechanizmów wewnętrznych stanowiła z kolei podstawę do przeprowadzenia oceny użyteczności usług ROEFS. Na podstawie zgromadzonych w badaniach terenowych informacji wysnuto również wnioski pozwalające na zweryfikowanie innych zgodności różnych aspektów funkcjonowania sieci ROEFS (np. odnośnie do sposobu rekrutacji klientów, strategii działalności obieranych na poziomie ośrodków, a także aktywności klientów ROEFS w pozyskiwaniu dofinansowania).

2.9. Wiedza preskryptywna

Cechy wiedzy preskryptywnej spełniają rekomendacje dostarczane w obu fazach ewaluacji sieci ROEFS. Zaproponowane rozwiązania zawierały w sobie zarówno wykaz zalecanych działań, jak i przypisanych do poszczególnych zaleceń adresatów. Z uwagi na brak wiedzy dotyczącej szczegółów przyszłej organizacji sieci oraz dokładnego harmonogramu dalszych działań, nie precyzowano wskazań czasowych dla przedstawionych rekomendacji.

2.10. Rozpowszechnianie wyników

Wyniki omawianej ewaluacji rozpowszechnianie były w dwóch etapach, zgodnie z dwufazowym harmonogramem realizacji badania. W obu przypadkach dystrybucję wyników rozpoczynały spotkania, w trakcie których przedstawiciele zespołu ewaluatorów prezentowali efekty swoich prac. Do udziału w spotkaniach zapraszani byli przedstawiciele Zarządu PARP oraz innych kluczowych komórek organizacyjnych instytucji zamawiającej ewaluację, reprezentacji Regionalnych Ośrodków EFS, samorządów wojewódzkich oraz komórek ewaluacyjnych zlokalizowanych w różnych instytucjach administracji publicznej[1].

Spotkania były także okazją do konsultacji wyników, a zgłaszane w ich trakcie uwagi uwzględniano przy opracowywaniu ostatecznych wersji raportów. Na użytek osób, które nie miały możliwości uczestniczyć w spotkaniach, na stronach internetowych PARP zamieszczone zostały prezentacje.

Na stronach PARP umieszczone zostały również ostateczne wersje raportów, ze wskazaniem dotyczącym cytowania wyników i kontaktami do osób odpowiedzialnych za merytoryczną stronę ewaluacji.

---

[1] Prezentacja przedstawiona podczas spotkań dostępne są na stronach internetowych PARP http://www.parp.gov.pl/index/more/3138

3. WYKORZYSTANIE EWALUACJI

3.1. Użycie ewaluacji

Kwestia wykorzystania wyników ewaluacji pozostaje otwarta z uwagi na krótki okres, jaki upłynął od momentu dostarczenia ostatnich rekomendacji. Wiadomo jednak, że zarządzający siecią (KOEFS) jest żywo zainteresowany wykorzystaniem przynajmniej niektórych rekomendacji oraz znacznej części wyników badań. Ze szczególnym zainteresowaniem zarządzającego spotyka się narzędzie diagnozy potrzeb klientów, opracowane jako element rekomendacji w ramach I fazy ewaluacji oraz wynikami obrazującymi przyczyny ograniczonej aktywności w pozyskiwaniu dofinansowania i różnorodne aspekty korzystania z usług ROEFS.

3.2. Wykorzystanie ewaluacji

Badanie przeprowadzone w ramach ewaluacji sieci ROEFS dostarczyło wyniki diagnozujące szereg zagadnień związanych z ubieganiem się o dofinansowanie ze środków EFS, zwłaszcza Programu Operacyjnego *Kapitał Ludzki*. Uzyskane wyniki pozwalają wyciągnąć wnioski odnośnie do specyfiki poszczególnych województw oraz grup potencjalnych projektodawców (np. organizacje pozarządowe i instytucje samorządowe z terenów wiejskich). Odbiorcą wyników ewaluacji mogą być zatem bardzo zróżnicowane grupy instytucji i osób indywidualnych zainteresowanych lub zaangażowanych w realizację przedsięwzięć wpisujących się w unijną politykę strukturalną.

3.3. Ograniczenia wykorzystania ewaluacji

Na obecnym etapie trudno ocenić, czy dojdzie do istotnych zakłóceń wykorzystania wyników omawianej ewaluacji. Możemy jednak wskazać na potencjalne źródła zagrożeń. Wykorzystanie wyników ewaluacji sieci REOFS może zostać zablokowane przez zmianę systemową oraz brak pewności co do przyszłej formuły funkcjonowania sieci. W momencie zamykania badania nadal nie sprecyzowano przyszłych kompetencji ani ulokowania instytucjonalnego KOEFS. Nieokreślona była również rola ośrodków w odniesieniu do Instytucji Pośredniczących we wdrażaniu EFS na szczeblu województw. Zakres oczekiwanych zmian będzie jednak najprawdopodobniej na tyle istotny, że nie pozostanie bez wpływu na wykorzystanie wyników ewaluacji.

ROZDZIAŁ 4.5.

Karol Olejniczak

EWALUACJA BIEŻĄCA PROGRAMÓW INTERREG IIIA AUSTRIA – NOWE KRAJE CZŁONKOWSKIE[1]

1. WSTĘP

W niniejszym rozdziale przedstawiono studium przypadku ewaluacji ciągłej czterech programów współpracy ponadgranicznej INTERREG IIIA dla Austrii i krajów ościennych – Czech, Węgier, Słowacji i Słowenii. Analizowany przypadek jest jednym z nielicznych europejskich przykładów badania ciągłego – ewaluacji bieżącej (prowadzonej w trakcie wdrażania programu), ale składającej się z wielu badań i konsultacji połączonych w jeden, spójny kontrakt. W rozdziale prezentowany jest kontekst decyzji o podjęciu takiego badania, warunki wyjściowe samej ewaluacji, przebieg badania, a także jego wyniki i wiedza przekazana interesariuszom. We wnioskach przedstawiono funkcje, które spełniło to badanie, a także zalety i ograniczenia ewaluacji ciągłej.

[1] Rozdział jest adaptowanym do potrzeb niniejszej publikacji przedrukiem z książki Olejniczak K. (2008), *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INTERREG III*, Warszawa, Wydawnictwo Naukowe Scholar, s. 140–152.

2. KONTEKST I PRZEDMIOT EWALUACJI

Programy INTERREG od początku lat 90. były jedną z Inicjatyw Wspólnotowych, nakierowanych na wspieranie współpracy między regionami Unii Europejskiej i przyczynianie się do harmonijnego, zrównoważonego i trwałego rozwoju. W omawianym tu okresie 2000–2006 funkcjonowały trzy komponenty INTERREG:

- Komponent A, poświęcony wspieraniu współpracy transgranicznej (ang. *cross-border cooperation*) pomiędzy sąsiadującymi ze sobą regionami UE, położonymi przy granicach wewnętrznych i zewnętrznych (zarówno lądowych, jak i morskich);
- Programy komponentu B, nakierowane na wzmacnianie współpracy transnarodowej (ang. *transnational cooperation*) w ramach ponadnarodowych stref współpracy obejmujących duże, europejskie obszary funkcjonalne (np. region Morza Bałtyckiego, region Alp);
- Komponent C, nakierowany na wspieranie współpracy międzyregionalnej (ang. *interregional cooperation*), czyli wymianę doświadczeń w ramach sieci współpracy zawiązywanych pomiędzy dowolnymi regionami Unii Europejskiej

W latach 2000–2006 w całej UE były aż 72 takie programy (zaś po rozszerzeniu UE aż 83), z czego najliczniejszy był komponent A (53)[2]. Budżet przeznaczony na programy INTERREG był stosunkowo niewielkim ułamkiem całości budżetu polityki spójności[3]. Niemniej jednak, w najnowszym okresie programowania (lata 2007–2013) ranga współpracy terytorialnej wzrosła i Inicjatywy Wspólnotowe INTERREG znalazły swoją kontynuację jako trzeci cel europejskiej polityki spójności.

Austria w ówczesnym okresie (lata 2000–06) uczestniczyła w aż sześciu programach współpracy ponadgranicznej (INTERREG komponent A). Program INTERREG był więc dla niej ważny – stanowił narzędzie budowania dobrych stosunków sąsiedzkich. W latach 2000–2003 sytuacja była o tyle wyjątkowa, że cztery z tych programów funkcjonowały na granicach zewnętrznych z państwami kandydującymi do Unii Europejskiej. W 2004 roku regiony sąsiedzkie z Czech, Słowacji, Słowenii, Węgier (dotychczas finansowane z programu PHARE-CBC) miały zostać włączone w trwający już (od 2000 roku) program INTERREG. Kwestia płynnego włączenia nowych krajów członkowskich (ang. NMS – *New Member States*) w działający już system, była dla Austrii na tyle wysokim priorytetem, że koordynację nad całym procesem rozszerzania INTERREG-ów przejął rząd federalny.

---

[2] Szerzej na temat genezy, struktury i funkcjonowania programów INTERREG zob. Bachtler et al. 2005.

[3] W latach 2000–2006 wszystkie Inicjatywy Wspólnotowe (prócz INTERREG-u, także LEADER+, Urban i Equal) miały do dyspozycji 5% budżetu polityki spójności (5,3 mld EUR ze 129 mld EUR).

2.1. Jaka była charakterystyka badanych programów?

Według formalnych wytycznych UE (o czym szerzej w kolejnym podpunkcie) badaniem ewaluacyjnym miały zostać objęte cztery programy INTERREG III A: Austria-Słowacja (AT-SK), Austria-Czechy (AT-CZ), Austria-Węgry (AT-HU), Austria-Słowenia (AT-SI). Co ciekawe, badanie, z przyczyn formalnych, miało w założeniu dotyczyć strategii, procesów i instytucji tylko po austriackiej stronie granicy (INTERREG dotyczył tylko krajów członkowskich – zob. przypis 6).

Wszystkie analizowane cztery programy austriackie były w 2003 roku średniozaawansowane[4]. Paradoksalnie okazało się to problemem, gdyż w ten zaawansowany system i w ponad połowie przypisane pieniądze (a więc w trwające już projekty) należało włączyć partnerów z nowych krajów członkowskich.

Struktura celów analizowanych 4 programów była do siebie zbliżona[5]. Należy jednak zwrócić uwagę, że w momencie podejmowania ewaluacji średniookresowych były to cele wyłącznie regionów austriackich. Partnerzy z państw kandydackich (Czech, Węgier, Słowacji, Słowenii) mieli swoje strategie i dużo skromniejsze źródło finansowania – program PHARE-CBC. Strategie PHARE-CBC były zwykle nakierowane na inwestycje infrastrukturalne, gdyż słaba infrastruktura była główną bolączką regionów przygranicznych państw kandydackich. Oczywiście koordynowano wizje i strategie (INTERREG vs PHARE-CBC), prowadząc wspólne rozmowy i konsultując kształt dokumentów. Niemniej jednak wciąż były to różne systemy[6].

2.2. Dlaczego podjęto badanie?

Sama decyzja o podjęciu badania była wymuszona zapisami prawnymi. Zgodnie z Rozporządzeniem UE dotyczącym funduszy strukturalnych

---

[4] Dla AT-CZ rozdysponowano 50% środków, dla AT-SI 57%, dla AT-SK 41%, a dla AT-HU 34,6%. Rzeczywiste wydatki wynosiły odpowiednio: 13,4%, 8,4%, 6,8%, 7,9% (źródło: dane z programów i baza EPRC).

[5] Pierwszym priorytetem we wszystkich 4 programach był rozwój gospodarczy, drugim pod względem wielkości finansowania – rozwijanie infrastruktury, a trzecim – zasoby ludzkie (ÖIR-Managementdisenste GMBH, 2001a, b, c, d, 2002, 2004a, b).

[6] To rozdwojenie środków finansowania wynikało z tego, że programy INTERREG były częścią funduszy strukturalnych – mogły więc współfinansować inwestycje tylko w krajach członkowskich UE. Austria miała więc programy INTERREG na każdej ze swoich granic UE, ale poza programem niemieckim (gdzie środki były wspólne), pozostałe programy finansowały tylko partnerów austriackich. Natomiast każde z państw stowarzyszonych z UE miało do dyspozycji programy z funduszu PHARE CBC (*Cross-Border Cooperation*), które w założeniu miały przystawać do programów INTERREG i finansować współpracę dla partnerów spoza UE. W praktyce trudno było prowadzić projekty przystające do siebie, gdyż INTERREG był planowany w perspektywie kilku lat, zaś PHARE-CBC rocznie; odmienne były też procedury, system zarządzania, nadzór różnych dyrekcji Komisji Europejskiej etc.

(Council Of The European Union 1999: art. 42), wszystkie państwa członkowskie były zobowiązane do przeprowadzenia – dla programów współfinansowanych ze środków UE – ewaluacji w połowie wdrażania interwencji (tzw. ewaluacje *mid-term* – średniookresowe). Ten półmetek wypadał w 2003 roku. Co więcej, już wtedy Komisja dawała sygnały, że badania zostaną prawdopodobnie powtórzone w 2005 roku. Urząd kanclerski wiedział, że w ciągu dwóch najbliższych lat będzie zmuszony zlecić przeprowadzenie czterech ewaluacji średniookresowych w 2003 roku (dla każdego z programów INTERREG) oraz powtórzyć te badania, jako uaktualnienie ewaluacji średniookresowych, w 2005 roku. Jednocześnie w latach 2003–2005 austriacka administracja programów INTERREG miała sprostać wyzwaniu rozszerzenia UE. Pojawił się więc pomysł, by formalny wymóg UE nieco zmodyfikować i wykorzystać jako pomoc w procesie rozszerzania programów INTERREG.

2.3. Kim byli interesariusze?

Głównymi interesariuszami badania były oczywiście austriackie instytucje zaangażowane w INTERREG. Struktura organizacyjna była standardowa dla wszystkich programów czterech austriackich INTERREG-ów (i zgodna z rozporządzeniami UE). Instytucją Zarządzającą (IZ) był Federalny Urząd Kanclerski w Austrii. Jak już wspomniano, instytucja ta była bardzo zainteresowana płynnym wdrażaniem INTERREG (jako głównego narzędzia stosunków dobrosąsiedzkich). Działały też Wspólne Sekretariaty Techniczne (WST) wspomagające na co dzień działania IZ, Wspólne Komitety Monitorujące (WKM), włączające wszystkie instytucje zaangażowane w programy oraz Wspólne Komitety Sterujące (WKS). Bardzo silną pozycję miały w tych instytucjach austriackie władze regionalne (Instytucje Pośredniczące) na co dzień zajmujące się wdrażaniem poszczególnych programów. Dla nich także INTERREG był głównym narzędziem rozwijania współpracy przygranicznej. Wszystkie te instytucje były ulokowane w Austrii i odnosiły się tylko do strony austriackiej.

Potencjalni interesariusze badania po stronie czeskiej, słowackiej, węgierskiej i słoweńskiej mieli oddzielne struktury w ramach programów PHARE-CBC. Austriacy starali się integrować te różne systemy instytucjonalne, przez włączanie w prace Komitetów i Sekretariatu przedstawicieli instytucji zarządzających PHARE-CBC w Czechach, Słowacji, Słowenii i Węgier. Wyzwaniem planowanego rozszerzenia INTERREG było więc przejście od lustrzanych struktur narodowych, na rzecz jednego systemu, wspólnego dla dwóch stron granicy.

W tym miejscu należy podkreślić różnicę w hierarchii priorytetów Austrii i krajów kandydackich oraz wynikające z tego późniejsze niskie zaangażowanie aktorów z nowych krajów UE w proces ewaluacji (wbrew oczekiwaniom partnerów austriackich). Dla nowych krajów program

INTERREG był w tamtym okresie kwestią drugoplanową, gdyż przygotowując się do członkostwa w UE, koncentrowały się na dużo bardziej znaczących dla nich sprawach – głównych funduszach strukturalnych czy polityce rolnej.

Analizując warunki kontekstowe, należy także zwrócić uwagę na fakt, że pomiędzy interesariuszami we wszystkich czterech omawianych tu programach (Austria–Czechy, Austria–Słowacja, Austria–Słowenia, Austria–Węgry) następowało zderzenie kultury administracyjnej Europy Zachodniej (nowoczesnego zarządzania opartego na wynikach) z administracjami postkomunistycznymi – a więc kierującymi się inną logiką, i mimo długiego procesu dostosowań do wymogów UE, wciąż zachowujących wiele z klasycznej administracji publicznej (sztywne procedury, hierarchiczność procesu decyzyjnego, niska elastyczność i innowacyjność). Brak kultury ewaluacyjnej w tych państwach był jednym z elementów tych różnic. Dodatkowym utrudnieniem – związanym również z kulturą administracji – były różne podejścia ze strony Komisji Europejskiej. W przypadku zarządzania programami europejskimi biurokracje krajów kandydujących były przyzwyczajane przez ostatnie 15 lat do procedur zarządzania Dyrekcji Generalnej – Pomoc Zewnętrzna (DG-External Aid) charakteryzującej się, w dużym uproszczeniu, ścisłymi regułami, dużym stopniem centralizacji i kontroli na poziomie Komisji. Tymczasem Austria i wdrażany w niej INTERREG były nadzorowane przez Dyrekcję Generalną ds. Polityki Regionalnej (DG-Regio), dającą dużo większą niezależność i samodzielność administracjom krajowym. Badanie ewaluacyjne mogło być więc postrzegane w państwach sąsiedzkich jako kolejny element kontroli.

Podsumowując kwestie kontekstowe analizowanych ewaluacji, można zidentyfikować dwa kluczowe elementy. Pierwszym było rozczłonkowanie (procedur, procesów, funduszy i interesów) między Austrią i jej regionami (wdrażającymi INTERREG-i) a państwami sąsiedzkimi współfinansowanymi z programów PHARE-CBC. Elementem drugim była presja reformy i wyzwanie integracji – po rozszerzeniu UE zarówno te odmienne systemy zasady, jak i pieniądze miały zostać połączone – od 2004 roku INTERREG automatycznie zaczynał obejmować nowe kraje członkowskie.

3. WKŁAD W BADANIE

3.1. Jakie były oczekiwania głównych interesariuszy?

Oczekiwania Zlecającego – Federalnego Urzędu Kanclerskiego – były jasne. Po pierwsze, ewaluacja miała pomóc operatorom programu w podniesieniu ich efektywności. Po drugie, oczekiwano, że pomoże we w miarę płynnym procesie integrowania nowych krajów w działające już

programy INTERREG. Szły za tym konkretne oczekiwania co do formy i podejścia badawczego. Urząd Kanclerski chciał, by był to proces interaktywny, by obserwacje ewaluatorów były na bieżąco używane w programie, by mieć ciągłość śledzenia rekomendacji, a wreszcie – by rozbudzić tym badaniem zainteresowanie partnerów z nowych krajów UE, pomóc im włączyć się w bieżącą dyskusję.

3.2. Jak sformułowano zakres badania?

Zlecający zdecydował się na ewaluację *on-going* – w tym wypadku najtrafniej tłumaczoną jako „ewaluacja ciągła". Po pierwsze, zintegrował badania w przekroju czasowym. Połączył w jeden pakiet planowane i wymagane w kolejnych latach ewaluacje – średniookresową (*mid-term*), uaktualnienie ewaluacji średniookresowej (*mid-term update*), a także dodatkowe ewaluacje bieżące (*on-going*). Po drugie, Zlecający zintegrował badanie pod kątem przestrzennym. Badania na wszystkie cztery INTERREG-i połączył w jeden pakiet badawczy. Dzięki temu – wykonywane jednocześnie i przy analogicznej metodologii – miały być zarówno elementem integrowania interesariuszy, jak i elementem wymiany doświadczeń ze wszystkich ośmiu programów (4 programów INTERREG i 4 programów PHARE-CBC).

Kontrakt na badanie ewaluacyjne zawierał więc wykonanie ewaluacji średniookresowych, bieżącą asystę instytucji programowych w kolejnych 3 latach oraz uaktualnienie badania średniookresowych (w sumie 3,5 roku badań i działań) dla czterech programów INTERREG IIIA (Austria–Czechy, Austria–Słowacja, Austria–Słowenia, Austria–Węgry). W zamówieniu (ang. *Terms of Reference*) umieszczono zapis, by w skład zespołu ewaluacyjnego wchodzili eksperci z nowych krajów członkowskich.

Taka strategia Zlecającego wynikała z kilku kwestii. Po pierwsze, Urząd Kanclerski miał negatywne doświadczenia z poprzedniego okresu programowania odnośnie do tradycyjnej, punktowej ewaluacji średniookresowej – ówczesne badania okazały się, z perspektywy operatorów programu, bezużyteczne. Były tylko dodatkowym narzędziem kontroli przez Komisję. Z drugiej strony, Zlecający miał już doświadczenia pozytywne, płynące z zastosowania interaktywnej, partnerskiej ewaluacji ciągłej[7]. Wreszcie, Zlecający był świadomy, że z powodu opóźnień programu, ewaluacja średniookresowa będzie powtórzona za dwa lata. Doszedł więc do wniosku, że skoro i tak będzie musiał zaangażować środki w spełnienie tego wymogu, to warto użyć go do własnych celów.

Ważnym elementem był też swoisty lobbing prowadzony przez jednego z głównych ewaluatorów w Austrii. Od początku lat 90. rozwijał

[7] Podejście „ciągłe" zastosowano z sukcesem przy ewaluacjach Celu 2 w latach 1995–1999, zaś podejście interaktywne w ewaluacji *ex ante* INTERREG-ów austriackich.

on zarówno naukowo, jak i w praktyce swoich kontraktów dla rządu austriackiego, podejście partycypatywne. Jego głównymi inspiracjami było myślenie systemowe, w tym systemowe podejście do procesów uczenia się i poprawy zarządzania programów (Hummelbrunner 2000) oraz podejście skoncentrowane na wartości użytkowej badania (*utilization-focused approach* M. Pattona). Do swoich pomysłów przekonał osoby z Federalnego Urzędu Kanclerskiego. Pomysł został przetestowany, a po tym, jak sprawdził się w ewaluacjach programów Celu 2 oraz w badaniu Inicjatywy URBAN, zyskał zwolenników w administracji austriackiej. Tę sytuację można określić jako lobbing bądź też uprzedni proces edukacyjny administracji (przygotowanie gruntu pod badanie).

3.3. Kim byli główni uczestnicy procesu ewaluacji?

W Instytucji Zlecającej (Urząd Kanclerski) był wyraźny lider i promotor pomysłu – jeden z urzędników wysokiej rangi, który przez lata był koordynatorem wszystkich INTERREG-ów, a w badanym okresie był silnie zaangażowany (zarówno na poziomie Austrii, jak i na poziomie europejskim) w prace nad nowym kształtem polityki współpracy terytorialnej po 2007 roku. Drugą zaangażowaną instytucją był Wspólny Sekretariat Techniczny (JTS – ang. *Joint Technical Secretariat*). Na potrzeby ewaluacji stworzono dwie grupy. Pierwszą była Grupa Sterująca Ewaluacją (ang. *Evaluation Steering Group*) złożona z 3–4 osób z kluczowych instytucji austriackich. Drugą, szerszą, była Międzyprogramowa Grupa Sterująca (*Cross-Programme Steering Group – CPSG*). W jej skład wchodzili przedstawiciele instytucji zarządzających i głównych interesariuszy ze wszystkich 4 programów (w tym przedstawiciele nowych krajów członkowskich). Grupa ta miała być platformą wymiany poglądów, poszukiwania wspólnych cech i rozwiązań dla wszystkich 4 programów.

Wykonawcą kontraktu został austriacki instytut studiów regionalnych (ÖIR – Österreichisches Institut für Raumplanung). Zespół ewaluatorów składał się 8–10 osób (po 2–3 osoby do programu). Prowadzącym dla każdego z badań był ewaluator austriacki, ale współpracował z nim ewaluator z danego nowego państwa członkowskiego. Kierownikiem zespołu i głównym koordynatorem był, mający wieloletnie doświadczenie, ewaluator austriacki.

4. PROCES EWALUACJI

Ze względu na obraną przez Zlecającego formę kontraktu (ewaluacja ciągła), trzeba spojrzeć całościowo na cały trzyipółroczny proces ewaluacji złożony z wielu badań i konsultacji, raportów formalnych (z ewaluacji średniookresowych i jej uaktualnień) oraz roboczych. Kalendarium tej

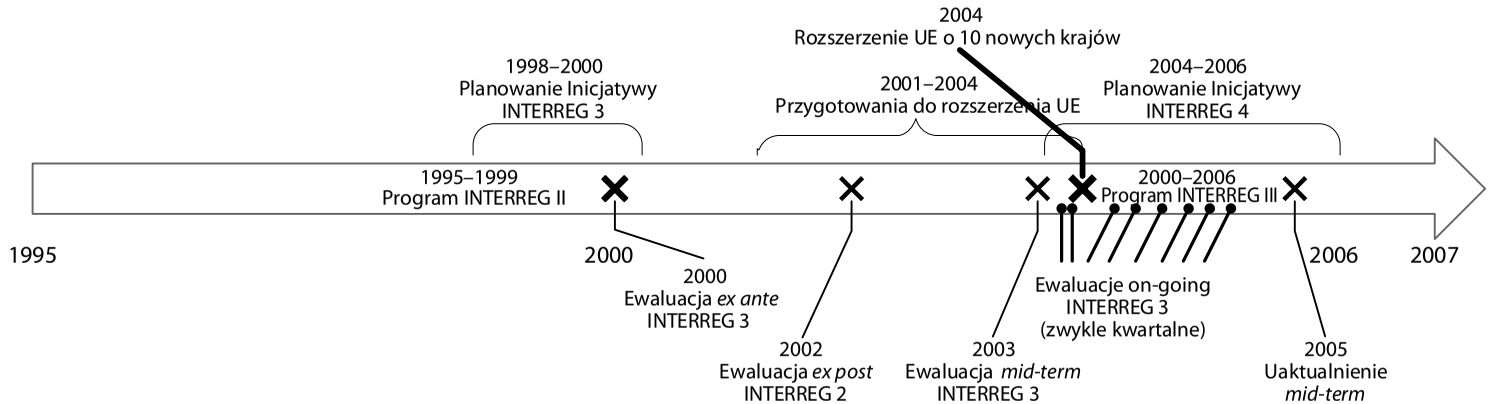
ewaluacji w kontekście całego planowania i zarządzania programami INTERREG dla Austrii przedstawiono na rysunku 1.

Z uwagi na długi czas badania i jego interaktywność, trudno było w sposób jednoznacznie wydzielić fazy strukturalizacji, obserwacji i analizy, oceny. Można było jednak zidentyfikować kilka etapów całego procesu na dwóch płaszczyznach. Pierwszą płaszczyzną była orientacja merytoryczna prac. I w tym wypadku można zauważyć, jak prace przesuwały się od identyfikowania problemów i wspólnych kwestii, przez wykonanie badania średniookresowego na potrzeby Komisji (analiza głównie procesów na poziomie programu), etap integrowania aktorów z nowych krajów (przygotowanie do rozszerzenia) – łączenie procedur, analizę problemów na poziomie projektów, i wreszcie – w trakcie oraz po uaktualnieniu badania – wsparcie prac nad nowymi dokumentami programowymi. Natomiast na płaszczyźnie organizacji procesu badawczego wyróżniały się trzy etapy (Hummelbrunner 2004). Pierwszym było wypracowanie pytań, wspólnych definicji i kontaktów. W drugim etapie prowadzono analizy dla każdego z 4 programów. Trzecim etapem były studia przypadku projektów. Dalsza analiza będzie przebiegać właśnie według tych trzech etapów.

Celem **etapu pierwszego** prac ewaluatorów było po pierwsze – zidentyfikowanie obszarów zainteresowań do dalszych badań, a po drugie – wypracowanie wśród głównych interesariuszy 4 różnych programów wspólnego rozumienia i wspólnych definicji. Etap ten miał więc wyraźne charakter fazy strukturalizacji.

Głównym narzędziem były spotkania ewaluatorów z interesariuszami na forum Grupy Międzyprogramowej (GM). Podczas dyskusji animowanych przez ewaluatorów wypracowano listę kluczowych kwestii, które miały być następnie badane. Była to wspólna lista dla wszystkich 4 programów, co dawało możliwość do przyszłych porównań. Ewaluatorzy nie tylko animowali dyskusję, ale także sugerowali nowe pytania (wychodzące poza perspektywę programu, na której koncentrowali się uczestnicy spotkań).

Jednocześnie w tym etapie przygotowywany był raport z ewaluacji średniookresowej. Z uwagi na krótki czas i inaczej postawione priorytety, zarówno ewaluatorom, jak i zlecającym chodziło o jak najszybsze wykonanie raportu (głównie pod kątem Komisji Europejskiej). Ewaluatorzy wykorzystali ewaluację średniookresową do wsparcia procesu identyfikowania problemów i wyzwań. Zastosowanym narzędziem były warsztaty bilateralne z udziałem IZ, KM i instytucji pośredniczących (w praktyce wdrażających) każdego z 4 programów. Zrobiono też jeden wspólny warsztat przedstawicieli 4 programów. Na nim ewaluatorzy zbierali pomysły od interesariuszy (w tym pomysły wstępnych rekomendacji). Pomysły rekomendacji zweryfikowano w wywiadach z IZ. Wnioski z badania zasiliły dyskusję na forum Grupy Międzyprogramowej o liście kluczowych kwestii badawczych.



Rysunek 1. Kalendarium planowania i ewaluacji programów INTERREG dla Austrii[8]

[8] Austria przystąpiła do UE w 1995 roku, a więc już w trakcie trwania programu INTERREG II (1994–99). Program dla granicy austriacko-czeskiej uruchomiono w 1995 roku. Finansowanie z INTERREG objęło tylko stronę austriacką (pieniądze z funduszy strukturalnych nie mogą finansować państw spoza UE) – regiony czeskie, słowackie, słoweńskie, węgierskie były finansowane znacznie mniejszymi pieniędzmi z PHARECBC.

Nowy program – INTERREG 3 przygotowano, zgodnie z okresem planowania polityki regionalnej UE, na lata 2000–2006. W 2000 roku po jego zaplanowaniu, a przed jego uruchomieniem zrobiono wymaganą ewaluację *ex ante*. Na jej podstawie poprawiono dokumenty programowe i uruchomiono Program. Już w trakcie trwania Programu INTERREG 3 wykonano ewaluację *ex post* INTERREG 2 (zgodnie z wytycznymi w dwa lata po zamknięciu programu – czyli 1999+2). W 2003 roku wykonano – zgodnie z wymaganiami UE – ewaluację *mid-term*. Ponieważ jednak programy INTERREG były w praktyce we wczesnym stadium działania Komisja wprowadziła wymóg powtórzenia badania w 2005 (*mid-term update*).

Od 2000 roku Austriacy przygotowywali się do włączenia krajów sąsiedzkich do programu INTERREG (po rozszerzeniu UE, pieniądze przydzielone w 2000 roku na program miały współfinansować również projekty czeskie, słowackie, słoweńskie, węgierskie). Aby to włączanie sąsiadów odbywało się płynnie, strona austriacka zlecała od 2003 do 2005 roku szereg szczegółowych badań (np. kwestii poprawy promocji programu w Czechach, kwestii blokad finansowania, poprawienia spójności procedur).

Od samego początku wybrano podejście partycypatywne, zorientowane na użytkowników – operatorów programu, a co z tego wynika – zorientowane na zarządzanie (poprawę procedur i procesów, wyjaśnienie problemów, efekty projektów). Stosowane techniki były wysoce interaktywne (warsztaty, dyskusje, same warsztaty dla badania miały jednak charakter dość formalny). Za interaktywne należy też uznać relacje w tej fazie między ewaluatorem a Zlecającym. Warto podkreślić, że w proces włączona została szersza rzesza interesariuszy, choć zainteresowanie i udział niektórych regionów były ograniczone[9]. Specyfiką tej ewaluacji było również to, że rozmowy dla wszystkich 4 programów były jednoczesne i wspólne. Dawało to możliwość nie tylko dodatkowych interakcji, lecz także odkrywania synergii między czterema programami.

W etapie drugim pytania z etapu pierwszego zostały szczegółowo zbadane dla każdego programu. Ewaluatorzy odgrywali głównie rolę animatorów dyskusji – organizowali i prowadzili spotkania ułatwiające dialog między partnerami programu. Szczególny akcent był położony na włączanie uczestników z nowych krajów członkowskich. Prowadzono np. kwestie wskaźników i monitoringu, analizę przystawalności projektów, analizy szybkości procedur (od ogłoszenia konkursu na projekty, przez aplikacje, wybór, pierwsze wypłaty).

Prezentowanie wyników analiz pod koniec prac było ciekawą, wspólną (ewaluatorów i operatorów programów) formą oceny. Ponieważ analiza dla wszystkich 4 programów była prowadzona w ten sam sposób, według tych samych pytań, dawało to możliwość porównań. Wyniki badań danych programów pokazywano na tle innych, co automatycznie podnosiło pytania o efektywność procedur, przyczyny zróżnicowań. Tu ewaluatorzy weszli w rolę nie tylko edukatorów (pokazując szerszą perspektywę, zadając pytania o procesy), ale i doradców. Relacje nawet w tym momencie były interaktywne – operatorzy programów przedstawiali swoje wyjaśnienia sytuacji, a ewaluatorzy na bieżąco, wspierając to dalszymi danymi, komentowali. Fakt, że dyskusja na forum Grupy Międzyprogramowej przebiegała w wewnętrznym gronie, mimo poruszenia wielu kwestii krytycznych, umożliwiła dialog, a sytuacje konfliktowe (nawet przy omawianiu istotnych problemów) były rzadkie.

Etapem trzecim były studia przypadków. Jedynym kluczem wyboru była decyzja operatorów programów. Ewaluatorzy przyglądali się w terenie tym projektom, które zostały wyznaczone przez instytucje zarządzające i wdrażające. Z reguły były to projekty problemowe, bo chciano się dowiedzieć o rzeczywistej sytuacji na poziomie beneficjentów. Dla każdego programu wykonano ok. 15–18 studiów (po dwóch stronach granicy). Wyniki były prezentowane Instytucjom Zarządzającym każdego

---

[9] Głównymi aktorami procesu były Federalny Urząd Kanclerski i Instytucje Zarządzające poszczególnych programów. Udział partnerów z nowych krajów członkowskich był zróżnicowany. Najmocniejszy był udział Słowacji, najsłabszy – Węgier.

z programów, ale projekty łączono też w grupy tematyczne i dyskutowano je na forum międzyprogramowym. Na tym etapie ewaluacji pojawiła się dodatkowa inicjatywa Zlecającego. Z uwagi na obłożenie kwestiami rozszerzenia Grupa Międzyprogramowa przestała funkcjonować, wyczerpały się też tematy do wielostronnej dyskusji. W związku z tym uruchomiono bilateralne platformy nauki (gdzie interesariusze z dwóch stron granicy, spotykali się i dyskutowali, w asyście Urzędu Kanclerskiego, i pomocy ewaluatorów).

Warto zwrócić uwagę, że działania z zakresu obserwacji, analizy i oceny były procesem ciągłym i opierały się na sprzężeniu zwrotnym (analiza – prezentacja wyników – reakcja interesariuszy – wnioski do dalszego badania – pogłębienie kwestii, które pojawiły się w reakcji na raport). Ewaluator cały czas odgrywał rolę animatora. Był też nie tyle konsultantem, ile doradcą (gdyż wychodził daleko poza minimum swoich działań). Stopniowo, w momencie uzyskiwania szerszej perspektywy (informacji, zestawień i porównań z 4 programów), wchodził w rolę edukatora.

5. PRODUKTY I WYKORZYSTANIE EWALUACJI

5.1. Jakie Produkty przyniosło badanie?

W przypadku tego badania trudno analizować jakość konkretnego raportu ewaluacyjnego (raportów formalnych – prócz ewaluacji średniookresowej i jej uaktualnienia – było kilka, dodatkowo opatrzonych mniej formalnymi prezentacjami i ekspertyzami). Tak więc ten punkt z zaprezentowanego w rozdziale 1.1. modelu-mapy odnosi się nie tylko do wiedzy zawartej w raporcie średniookresowym, ale wiedzy z całego procesu ewaluacji ciągłej.

Formalne raporty z ewaluacji średniookresowej, a potem jej uaktualnienia, nie przynosiły żadnych niespodzianek, a zawarta w nich wiedza dotyczyła głównie procedur, procesów i problemów z nimi związanych (*know what*). Poszukiwanie przyczyn tego stanu rzeczy (*know why*) oraz sposobów adresowania problemów (*know how & when*) było krystalizowane dopiero w późniejszych badaniach i rozmowach. Te rodzaje wiedzy zostały silnie rozbudowane w interaktywnym procesie badawczym, na przestrzeni kolejnych miesięcy. Dotyczyły zarówno procedur, jak i instytucji, procesów. Ciekawe jest jednak, że kwestie oceny funkcjonowania instytucji centralnych – głównego Zlecającego – zostały zupełnie pominięte.

Badanie ciągłe (*on-going*), mimo że dotyczyło kwestii technicznych – z zakresu zarządzania programami, dało uczestnikom nową wiedzę również o szerszym kontekście procesów czy o źródłach inspiracji. Wiedza ta płynęła z porównań i zestawień wyników czterech programów, będących przedmiotem ewaluacji.

Silny był również moduł wiedzy o tym, kto ma działać (*know who*) – i to nie tylko w znaczeniu, kogo włączać w programy, ale również w jaki sposób (jakie metody aktywizują poszczególnych uczestników – szczególnie tych z nowych krajów; jak z nimi dyskutować).

Warto zaznaczyć, że w wyniku ewaluacji pojawiły się jeszcze dwie kwestie, które można uznać za wiedzę koncepcyjną. Pierwszą była wiedza zlecających na temat użyteczności podejścia interaktywnego (była ona budowana przez ewaluatora jeszcze przed kontraktem). Drugim rodzajem wiedzy było zidentyfikowanie różnic między aktorami w postrzeganiu tych samych terminów, a następnie wypracowanie wspólnego rozumienia kwestii kluczowych z perspektywy programów (np. kwestii rozumienia „projektu ponadgranicznego"). Wydaje się to bardzo ważnym osiągnięciem w perspektywie koordynacji i wspólnego zarządzania 4 programami.

Ważnym produktem badania były także kontakty pomiędzy interesariuszami. Niewątpliwie ewaluator stworzył silne, stałe kontakty ze zlecającym Federalnym Urzędem Kanclerskim oraz Sekretariatem koordynującym zlecenie. Federalny Urząd Kanclerski był też promotorem wyników (to on wyszedł z propozycją stworzenia platform wiedzy).

Omawiając zbudowane kontakty z interesariuszami, trzeba podkreślić różnice między instytucjami austriackimi a partnerami z nowych krajów członkowskich. W przypadku tych pierwszych były to kontakty stałe i rozległe – objęły wszystkich aktorów. W przypadku nowych krajów członkowskich problemem był brak ciągłości instytucjonalnej i silna rotacja kard. Uczestnicy procesu i negocjacji często zmieniali się, zmieniały się również instytucje odpowiedzialne za rozmowy i współpracę w programie. To wywoływało pewną frustrację zarówno u zlecających, jak i ewaluatorów, gdyż musieli odbudowywać kontakty po każdej zmianie. Zlecający podkreślał, że poczucie współwłasności, współuczestnictwo i kontakt z ewaluatorem byłyby dużo silniejsze, gdyby wszyscy aktorzy procesu wnieśli wkład finansowy w kontrakt (był on finansowany wyłącznie przez stronę austriacką, centralną instytucję). Warto zwrócić uwagę, że w budowaniu kontaktów z szerszą grupą interesariuszy pomagała ewaluatorom instytucja zlecająca kontrakt. Zarówno w tworzeniu relacji, jak i całym badaniu ważną rolę odegrał fakt, że zespoły ewaluacyjne były mieszane (Austriak i osoba z kraju sąsiedzkiego). Zasadniczo pomagało to w rozeznaniu kontekstu i badaniu w nowych krajach członkowskich (choć w przypadku węgierskim okazało się problematyczne – konflikty między krajowym ewaluatorem a instytucją węgierską zostały przeniesione z innych projektów i zablokowały na jakiś czas działania w analizowanej ewaluacji).

Ocena rozległości kontaktów nawiązanych poza programem zależy też od tego, co przyjmiemy jako program. Jeśli uznamy, że program to INTERREG, natomiast PHARE-CBC było czymś zupełnie zewnętrznym,

wtedy niewątpliwie tak rozumiane kontakty zewnętrzne były bardzo szerokie i głębokie. Jeśli jednak będziemy poszukiwać kontaktów nawiązanych na poziomie europejskim, poza Phare-CBC i INTERREG austriackim, to należy stwierdzić, że takowe się nie pojawiły.

5.2. Jak rozpowszechniano wyniki?

Formalne raporty z ewaluacji średniookresowej, a potem i jej uaktualnienia, były rozpowszechniane w wąskim gronie głównych interesariuszy (Zlecający, Komitety Monitorujące, Instytucje Zarządzające i pośredniczące, sekretariaty techniczne), a następnie umieszczone na stronie internetowej. Były one raczej przyczynkiem do dalszych prac.

Sytuacja wyglądała zupełnie inaczej w odniesieniu do dokumentów roboczych, prezentacji, raportów okresowych (spoza formalnego wymogu UE – a więc niewysyłanych do Komisji Europejskiej). Zlecający miał tu jasną strategię rozpowszechniania wyników. Wychodził z założenia, że fragmenty badań prezentowane na bieżąco i jako wkład w określone dyskusje, są dużo lepiej przyswajalne przez uczestników programu. Dlatego też wyniki ewaluacji były rozpowszechniane na bieżąco – w ramach prac i spotkań instytucji programu – głównie Komitetów Monitorujących i Sterujących (zwykle w formie krótkich not, prezentacji, podsumowań – były one potem zbierane w raporty, ale tylko w celu kontraktowym – jako podstawa wypłat). Zawsze towarzyszyła im dyskusja (animowana przez Sekretariat Techniczny i ewaluatorów). Od 2006 roku kluczowe dla rozpowszechniania wyników były platformy wiedzy (*learning platforms*), na których spotykali się operatorzy programów z dwóch stron granicy, wspomagani przez ewaluatorów i WST. Ewaluatorzy odegrali aktywną rolę w tym procesie.

5.3. Czy wykorzystano ewaluację?

Niewątpliwie odbiorcy ewaluacji przyswoili nową wiedzę, którą przyniosły badania. Co prawda nie zawsze postrzegali, w sposób uświadomiony, że jest to wiedza nowa, gdyż jako osoby włączone od początku w jej tworzenie, szybko ją interalizowali. W przypadku nowych krajów członkowskich największy przyrost wiedzy był wśród aktorów słowackich (według oceny Federalnego Urzędu Kanclerskiego). Niemniej jednak duża część wiedzy została w tych krajach stracona z powodu niestabilności instytucjonalnej, braku ciągłości personelu.

Odbiorcy ewaluacji dowiedzieli się nowych rzeczy przy okazji zderzenia wniosków z czterech programów. Pokazanie jednego z programów na tle pozostałych dało nową perspektywę. Ewaluatorzy austriaccy pogłębili swoją wiedzę na temat skutecznego stosowania ewaluacji partycypatywnej. Niestety ewaluatorzy z krajów ościennych, znów z powodu braku ciągłości personalnej, skorzystali mniej.

Wiedza budowana w procesie ewaluacji ciągłej została wykorzystana przez interesariuszy programów także w planowaniu kolejnych inicjatyw. Używano jej na bieżąco w pracach Grupy Zadaniowej ds. Rozszerzenia (przy przekształcaniu programów z lat 2000–2004), a także odwoływano się do niej (zarówno Urząd Kanclerski, jak i nowe kraje członkowskie) w dyskusjach nad programami na lata 2007–2013.

Ciągłość procesu ewaluacji i jej wewnętrzny, nieformalny charakter uniemożliwiły ilościowe podejście do rekomendacji (ile zatwierdzono, ile wdrożono). Wdrożone rekomendacje czy raczej sugestie (bo wnioski z badań nie zawsze były przedstawiane w formie rekomendacji – te tworzyli dopiero uczestnicy dyskusji nad wynikami) dotyczyły strategii programu (dostosowanie strategii do potrzeb nowych krajów) oraz procesów (procedury na poziomie programów i projektów). Rekomendacje z ewaluacji średniookresowej zostały też użyte jako punkt odniesienia dla jej uaktualnienia (Bauer-Wolf 2005; Hummelbrunner, Omersel-Weeks 2005; Kojan, Navratilova 2005; Reiner, Karasz 2005). Szersze wykorzystanie formalnych raportów można też prześledzić w publikacjach zewnętrznych. Samego raportu użyto jako źródła informacji w studiach zewnętrznych prowadzonych przez INTERACT (Bachtler et al. 2005; INTERACT 2005a, b, c, d).

Badanie w znaczący sposób poszerzyło współpracę i kontakty zarówno interesariuszy austriackich z nowymi krajami członkowskimi, jak również operatorów austriackich pomiędzy programami. Kontakty zostały podtrzymane i są obecnie wykorzystywane.

Trudno było w sposób jednoznaczny oszacować koncepcyjne wykorzystanie raportu w debatach dotyczących nowej polityki współpracy terytorialnej na lata 2007–2013, jednak pewne ślady wniosków z ewaluacji średniookresowej można było dostrzec w dokumentach Federalnego Urzędu Kanclerskiego z 2004 roku, będących głosem w ówczesnej ogólnounijnej debacie nad przyszłym kształtem INTERREG IV. Ponadto spostrzeżenia na temat nowoczesnego systemu monitoringu dla programów współpracy terytorialnej rozwinięto w formie dalszych prac i studiów już na poziomie europejskim (Hummelbrunner 2006). Wpłynęły one na sposób myślenia i praktyki monitorowania INTERREG IV. Ten efekt w dużej mierze wynika z aktywności naukowej samego ewaluatora.

6. WNIOSKI ZE STUDIUM PRZYPADKU

6.1. Jakie funkcje spełniło badanie ewaluacyjne?

Odnosząc opisany przypadek do szerokich funkcji ewaluacji, można stwierdzić, że badanie wypełniło **funkcję wzmocnienia partnerstwa i poczucia współwłasności uczestników**. Zgodnie z założeniem zlecających, pomogło nawiązać, a następnie wzmocnić relacje z instytucjami

z nowych krajów członkowskich. Interaktywność procesu i jego sprzężenie dla czterech programów umożliwiły zbudowanie sieci współpracy i kontaktów także między programami. Badanie aktywizowało interesariuszy i włączyło w dyskusję nad programami instytucje i osoby dotychczas słabo obecne (NMS, niektóre regiony). Najsilniej jest to widoczne w zwiększeniu aktywności aktorów słowackich.

Badanie ewaluacyjne czterech austriackich INTERREG-ów **wyraźnie spełniło też funkcję poprawy wdrażania** we wszystkich analizowanych programach. Obserwacje z ewaluacji i idące za tym wnioski praktyczne umożliwiły tzw. *streamlining* (włączanie do głównego nurtu) projektów, poprawę wielu procedur na poziomie programów, a przede wszystkim usprawnienie przepływów informacji między instytucjami zaangażowanymi we wdrażanie (po stronie austriackiej i nowych państw członkowskich).

Funkcja poprawy jakości planowania również była obecna w analizowanym przypadku. Ewaluacja pomogła przystosować programy do potrzeb nowych krajów członkowskich. Co więcej, dała też znaczący merytoryczny wkład do planowania kolejnych interwencji – w nowym okresie programowym.

Kolejna funkcja – poznawcza (wsparcia procesów uczenia się w sektorze publicznym) była bardzo silna, ale tylko w niektórych – bardziej bieżących i operacyjnych – aspektach. Sam raport średniookresowy pomógł zidentyfikować mapę problemów – kolejne działania przynosiły zaś na bieżąco wyjaśnienia czynników i procesów warunkujących działanie interwencji. Ważnym elementem nabytej przez uczestników wiedzy było wypracowanie wspólnego rozumienia procesów, definiowania ich i myślenia o projektach współpracy terytorialnej. Badanie spełniło tę funkcję ewaluacyjną głównie w zakresie budowania zdolności ewaluacyjnych – wśród aktorów austriackich było to oswojenie ich i przekonanie do ewaluacji partnerskiej, interaktywnej. Uczestnicy z nowych krajów członkowskich nauczyli się też współpracy i dyskusji. Funkcja uczenia się wyszła ponadto poza badane programy i może mieć wpływ na całą politykę. Obserwacja z badań posłużyła do wypracowywania nowego podejścia do monitoringu pomocy terytorialnej. Badanie dawało wiedzę na temat dobrych praktyk, ciekawych rozwiązań, ale czerpiąc głównie z czworokąta badanych programów, a więc w znikomym stopniu sięgając poza badaną rzeczywistość. Koncentrując się na bieżących potrzebach informacyjnych uczestników, nie dało jednak znaczącego wkładu w rozumienie szerszych teorii, podstaw polityki, a przede wszystkim wyjaśnienia relacji przyczynowo-skutkowych między interwencją a jej efektami.

Wreszcie **realizację funkcji *wzmocnienia odpowiedzialności w sektorze publicznym* (ang. *accountability*) należy uznać za bardzo ograniczoną**. Pokazanie skuteczności programów było adresowane głównie do Komisji Europejskiej, a i tak było ograniczone (stopień

analizy osiągnięcia celów został praktycznie pominięty). Pokazaniu stanu programów służyły przede wszystkim raporty średniookresowe (przygotowywane, gdy programy były wciąż w swoim początkowym stadium – nie dotykały więc nawet kwestii efektów projektów). Wynikało to po części ze słabego stopnia zaawansowania programów (brak efektów, które można by odnieść do celów), a po części z przyjętej optyki ewaluacji (skoncentrowanie się na procedurach i potrzebach informacyjnych operatorów programów). Kwestia efektów i przystawalności projektów po dwóch stronach granicy była badana pod koniec kontraktu ewaluacyjnego (gdy programy były już bardziej zaawansowane), ale wnioski z badań nie były prezentowane w formie oddzielnego raportu szerszej opinii publicznej czy Komisji.

6.2. Jakie czynniki wpłynęły na wykorzystanie ewaluacji?

Zanim zostaną przedstawione wnioski końcowe, warto rozważyć, które z czynników wpłynęły na **wykorzystanie ewaluacji**.

Niewątpliwie kluczowym czynnikiem był **kontekst geopolityczny** – perspektywa rozszerzenia Unii Europejskiej i związana z tym konieczność zreformowania programów pod kątem udziału nowych krajów. Związany tym był z kolei niezmiernie ważny czynnik wykorzystania synergii między 4 programami – jednoczesne ich badanie i współanalizowanie. Było to silnie eksploatowane zarówno przez zlecających, jak i ewaluatora.

Ważna okazała się też **postawa głównych zlecających** (w tym czynnik osobowościowy) – ich przekonanie o wartości nowego podejścia do ewaluacji, zaangażowanie oraz silna promocja tego pomysłu i pojawiających się efektów. Łączyło się to i wzajemnie wzmacniało z zaangażowaniem głównego ewaluatora, jego determinacją w użyciu partycypatywno-interaktywnego podejścia.

Przyjęta **interaktywna, partycypatywna i zorientowana na odbiorców filozofia badania**, a także aktywna rola ewaluatora jako animatora, doradcy, a z czasem i edukatora, w sposób jednoznacznie pozytywny wpłynęły na użycie badania. Wysoka aktywność Zlecającego i stworzenie instytucjonalnych platform wymiany informacji (Międzyprogramowa Grupa Sterująca, platformy wiedzy) były dodatkowym, znaczącym czynnikiem.

Również **sposób przekazywania wiedzy i forma relacji** okazały się ważne. Użycie ewaluacji zostało zintensyfikowane przez przyjętą formę raportowania – a właściwie formę przekazywania wiedzy – wprowadzanej na bieżąco, w sposób ciągły, jako wkład w toczącą się dyskusję i serwowanej nie w formie jednolitego raportu, a w małych pakietach – prezentacji, podsumowań dostosowanych do bieżącej grupy odbiorców. Forma relacji między zlecającymi, zaangażowanymi interesariuszami a ewaluatorem, a konkretnie – bliski bieżący kontakt, oparty na

zaufaniu, umożliwiła swobodniejszą dyskusję o problemach programu i projektów. Należy jednak podkreślić, że zaufanie udało się zbudować dzięki ciągłości procesu (horyzont czasowy badania to 3,5 roku) oraz temu, że był to proces wewnętrzny – dyskusja odbywała się w szerokim, ale stałym gronie operatorów programu.

Nie bez znaczenia wreszcie był **rodzaj wiedzy**. Była ona bardzo praktyczna, oparta w zasadzie na sumie doświadczeń uczestników (a więc dużo łatwiejsza do internalizacji), dostosowana do ich potrzeb informacyjnych.

6.3. Jakie siły i ograniczenia miało badanie?

Niewątpliwą siłą analizowanego badania była jego skuteczność w adresowaniu bieżących kwestii zarządzania oraz budowania partnerstwa i poczucia współwłasności wśród interesariuszy programów.

Wynikała ona z dwóch kwestii. Po pierwsze z odmiennej – w porównaniu do wymaganej wtedy przez Komisję Europejskiej – filozofii badania ewaluacyjnego. Opisują to najlepiej słowa austriackiego ewaluatora:

„W ewaluacji średniookresowej zarządzający mówią nam: «Znajdźcie i oceńcie, jak sprawy się mają». Wiedzą, że to rodzaj kontroli, że całość czytać będzie Komisja i starają się wszystko przedstawić w dobrych barwach. Przy ewaluacji bieżącej my mówimy: «nie marnujmy czasu, powiedzcie, jakie są wasze problemy, czego chcecie się wiedzieć i od razu będziemy nad nimi pracować». Pracujemy z nimi, ale Komisja nie ma dostępu do informacji roboczych – zobaczy tylko rezultat końcowy. To musi pozostać wewnętrzne, gdyż inaczej nikt nie odważy się szczerze rozmawiać".

Tak więc niższy stopień formalizacji i bliska współpraca między ewaluatorami i interesariuszami pozwoliły efektywnie rozwiązywać problemy i bieżące wyzwania.

Ponadto skuteczność badania wynikała z jego elastyczności i dostosowania do bieżących potrzeb informacyjnych. Zamiast jednego punktowego studium dotyczącego dużych kwestii, badano wiele drobnych pytań ważnych dla (bo wysuwanych przez) zalecających. Dzięki temu dużo łatwiej dostosowano badanie do tempa programu (tu najważniejszym momentem było rozszerzenie UE). Co więcej, użytkownikom łatwiej było na bieżąco przyswajać i stosować partie informacji, niż cały jeden raport.

Ta wysoka skuteczność w kwestiach bieżących jest też największym ograniczeniem tego badania, jeśli przyjmiemy perspektywę funkcji rozliczeniowej i poznawczej. Ewaluatorzy koncentrując się na „tu i teraz" programów, siłą rzeczy nie mogli podjąć obiektywnej, zdystansowanej i szerszej refleksji nad mechanizmami przyczynowo-skutkowymi i efektami programów (np. co już zaznaczyliśmy, pominięto rolę głównego interesariusza – Urzędu Kanclerskiego).

Analizowany przykład dobrze ilustruje więc omawiane już w pierwszych rozdziałach niniejszej książki dylematy wyboru między funkcjami ewaluacji, a także sprzęgnięte z tym wybory podejść ewaluacyjnych.

BIBLIOGRAFIA

Bachtler J., Taylor S., Olejniczak K. (2005), *A Study of the Mid Term Evaluations of INTERREG programmes for the programming period 2000 until 2006*, Vienna, INTERACT Programme Secretariat.

Bauer-Wolf S. (2005), *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria–Hungary*, Vienna, Budapest, Managing Authority of INTERREG IIIA Austria–Hungary Programme.

Council of the European Union (1999), *Council Regulation (EC) No 1260/1999 of 21 June 1999 laying down General Provisions on the Structural Funds*, 1260/1999, OJ L161, 26.6.1999.

Hummelbrunner R. (2000), *A systems approach to Evaluation. Applications of Systems theory and Systems thinking in Evaluations*, 4th European Evaluation Society Conference, Lausanne, 12–14.10.2000.

Hummelbrunner R. (2004), *On-going Evaluation INTERREG III A Programmes AT-CZ, AT-HU, AT-SI, AT-SK. Synthesis Cross-programme topics*, Graz, Federal Chencellery.

Hummelbrunner R. (2006), *Process Monitoring of Impacts. Proposal for a new approach to monitor the implementation of "Territorial Co-operation" programmes*, Graz, INTERACT – Managing Transition Point.

Hummelbrunner R., Omersel-Weeks B. (2005), *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria–Slovenia*, Graz, Ljubljana, Managing Authority of INTERREG IIIA Austria–Slovenia Programme.

INTERACT (2005a), *Programme Summary of the Austria–Czech Republic INTERREG IIIA Programme*, Maastricht, INTERACT Point Tool Box.

INTERACT (2005b), *Programme Summary of the Austria–Hungary INTERREG IIIA Programme*, INTERACT Point Tool Box, Maastricht.

INTERACT (2005c), *Programme Summary of the Austria–Slovakia INTERREG IIIA Programme*, Maastricht, INTERACT Point Tool Box.

INTERACT (2005d), *Programme Summary of the Austria–Slovenia INTERREG IIIA Programme*, Maastricht, INTERACT Point Tool Box.

Kojan M., Navratilova K. (2005), *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria–Czech Republic*, Vienna, Praha, Managing Authority of INTERREG IIIA Austria–Czech Republic Programme.

ÖIR-MANAGEMENTDISENSTE GMBH (2001a), *Austria–Czech Republic INTERREG IIIA. Community Initiative Programme*, Joint Secretariat – ÖIR, Vienna.

ÖIR-Managementdisenste GMBH (2001b), *Austria–Hungary INTERREG IIIA. Community Initiative Programme*, Vienna, Joint Secretariat – ÖIR.

ÖIR-Managementdisenste GMBH (2001c), *Austria–Slovakia INTERREG IIIA. Community Initiative Programme*, Vienna, Joint Secretariat – ÖIR.

ÖIR-Managementdisenste GMBH (2001d), *Austria–Slovenia INTERREG IIIA. Programme Complement*, Vienna, Joint Secretariat – ÖIR.

ÖIR-Managementdisenste GMBH (2002), *Austria–Slovakia INTERREG IIIA. Programme Complement*, Vienna, Joint Secretariat – ÖIR.

ÖIR-Managementdisenste GMBH (2004a), *Austria–Czech Republic INTERREG IIIA. Programme Complement*, Vienna, Joint Secretariat – ÖIR.

ÖIR-MANAGEMENTDISENSTE GMBH (2004b). *Austria–Hungary INTERREG IIIA. Programme Complement*, Vienna, Joint Secretariat – ÖIR.

Reiner K., Karasz P. (2005), *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria–Slovakia*, Vienna, Bratislava, Managing Authority of INTERREG IIIA Austria–Slovakia Programme.

CZĘŚĆ PIĄTA

BADANIE EWALUACYJNE – W POSZUKIWANIU EFEKTÓW I SZERSZYCH WNIOSKÓW

ROZDZIAŁ 5.1.

Bartosz Ledzion

WPROWADZENIE

Piąta część podręcznika zawiera studia przypadków ewaluacji efektów przeprowadzanych po zakończeniu realizacji interwencji. Badania te pokazują rzeczywiste efekty programów (w tym efekty uboczne), badają relacje przyczynowo-skutkowe między działaniami a zmianami na danym obszarze, w grupie docelowej, w środowisku.

Pierwsze studium przypadku dotyczy ewaluacji projektów szkoleniowych finansowanych w ramach Działania 2.4 Sektorowego Programu Operacyjnego *Rozwój Zasobów Ludzkich – Wzmocnienie zdolności administracyjnych*. Powyższe badanie jest wzorcowym przykładem oceny efektów zrealizowanych szkoleń oraz zbadania ich zgodności z wdrażaną polityką racjonalizacji zarządzania zasobami ludzkimi w służbie cywilnej.

Drugi artykuł dotyczy ewaluacji *ex post* Programu Współpracy Przygranicznej PHARE Polska–Niemcy. Zostało wybrane do studium przypadku, gdyż było jednym z pierwszych badań ewaluacji zrealizowanych w Polsce przez zespół polskich ekspertów. Jest to także, jak do tej pory, jedno z największych i najbardziej złożonych badań ewaluacyjnych przeprowadzonych w naszym kraju.

Przedmiotem trzeciego studium jest ewaluacja *ex post* regionalnego komponentu MSP Programu Phare 2003 SSG. Projekt ten zasługuje na uwagę przede wszystkim ze względu na zastosowanie nowatorskiego podejścia do pomiaru efektów programu (badanie efektu netto), wykorzystania metody PSM (Propensity Score Matching).

Celem czwartego studium przypadku jest zaprezentowanie przebiegu i wyników badania, które można rozpatrywać jako ewaluację z długą perspektywą czasową, pozwalając je uznać za ewaluację typu *ex post*. Impulsem podjęcia badań w gminach objętych Programem Inicjatyw Lokalnych była przede wszystkim naukowa ciekawość. Badanie zostało wymyślone, zaplanowane i przeprowadzone jako badanie naukowe

finansowane w ramach grantu KBN pn. *Regionalne i lokalne aspekty polityk sektorowych*.

W ostatnim artykule przedstawiamy studium przypadku z Wielkiej Brytanii: ewaluację *mid-term* programu przejściowego *Highlands & Islands 2000–2006*. Analizowany przypadek jest interesującym przykładem ewaluacji efektów z uwagi na zastosowanie różnorodnego warsztatu metod zarówno ilościowych, jak i jakościowych do oceny wpływu.

ROZDZIAŁ 5.2.

Magdalena Kuśmierz

EWALUACJA *EX POST* PROJEKTU SYSTEMOWEGO KPRM W ZAKRESIE SZKOLEŃ, REALIZOWANEGO W RAMACH DZIAŁANIA 2.4. SPO RZL

1. WSTĘP

W okresie poprzedzającym opisywaną w tym rozdziale ewaluację Kancelaria Prezesa Rady Ministrów (do października 2006 roku – Urząd Służby Cywilnej) zleciła wykonanie kilku ekspertyz z dziedziny zarządzania zasobami ludzkimi w służbie cywilnej, dotyczących przede wszystkiem opracowania standardów kompetencyjnych pracowników USC, czy zarządzania szkoleniami w służbie cywilnej (dalej: s.c.; obecnie trwa projekt, którego celem jest wartościowanie stanowisk). Przesłanką zleconej ewaluacji szkoleń, w okresie wprowadzania nowych standardów w obrębie polityki personalnej, było poszukiwanie odpowiedzi na pytanie o postępy we wdrażaniu zmian. Zlecona ocena była szczególnie istotnym narzędziem śledzenia zmian, bo w sytuacji nieistnienia innych wystandaryzowanych form kontroli procesu wdrażania standardów, dawała możliwość oceny efektów zrealizowanych szkoleń, a przy okazji (bo

nie było to głównym celem badania) zbadania ich zgodności z wdrażaną polityką racjonalizacji zarządzania zasobami ludzkimi w s.c. Dlatego wydaje się, że niniejszy przypadek jest dobrą ilustracją ewaluacji jako narzędzia poszukiwania efektów podejmowanych przez instytucję działań i pozwoli na formułowanie szerszych wniosków.

2. KONTEKST I PRZEDMIOT EWALUACJI

2.1. Dlaczego podjęto badanie?

Przed przystąpieniem do analizy procesu badawczego oceny *Ewaluacja projektów: szkolenia generalne i szkolenia językowe – I etap* zrealizowanych w ramach *Działania 2.4 Sektorowego Programu Operacyjnego Rozwój Zasobów Ludzkich – Wzmocnienie zdolności administracyjnych*, współfinansowanych z Europejskiego Funduszu Społecznego istotne jest przedstawienie specyfiki systemu wdrażania Działania 2.4. Priorytetu II SPO RZL. Kancelaria Prezesa Rady Ministrów uczestniczy we wdrażaniu Sektorowego Programu Operacyjnego *Rozwój Zasobów Ludzkich*, współfinansowanego ze środków Europejskiego Funduszu Społecznego jako Instytucja Wdrażająca, której rolę odgrywa Biuro Dyrektora Generalnego w KPRM, oraz jako Projektodawca (Beneficjent) – tę funkcję pełni Departament Służby Cywilnej w KPRM. Oznacza to *de facto*, że Kancelaria Prezesa Rady Ministrów zarówno składa wnioski o dofinansowanie projektów, jak i realizuje je i monitoruje. Wszystkie urzędy administracji rządowej są natomiast odbiorcami szkoleń (beneficjentami ostatecznymi), realizowanych w ramach Działania 2.4. KPRM realizuje również projekty w ramach Priorytetu 3 *Pomoc Techniczna*. System wdrażania Działania 2.4. SPO RZL wyglądał następująco:

- Instytucja Zarządzająca – Departament Zarządzania EFS w MGPiPS,
- Instytucja Pośrednicząca – Departament Zarządzania EFS w MGPiPS,
- Beneficjent końcowy – Departament Służby Cywilnej w Kancelarii Prezesa Rady Ministrów (uprzednio Departament Szkolenia i Rozwoju Służby Cywilnej w Urzędzie Służby Cywilnej),
- Ostateczni beneficjenci:
  - członkowie korpusu służby cywilnej zajmujący stanowiska wyższego i średniego szczebla zarządzania, zatrudnieni w ministerstwach, urzędach centralnych, urzędach wojewódzkich, urzędach administracji zespolonej i niezespolonej;
  - członkowie korpusu służby cywilnej pracujący w urzędach administracji rządowej, w komórkach odpowiedzialnych za realizację zadań wynikających z członkostwa w UE;
  - dyrektorzy generalni urzędów administracji rządowej.
- Ostateczni odbiorcy (projektodawcy): Departament Służby Cywilnej w Kancelarii Prezesa Rady Ministrów.

Mówiąc o warunkach kontekstowych, należy również zwrócić uwagę na fakt, że doświadczenie Zamawiającego w zakresie wdrażania i realizacji projektów finansowanych ze środków unijnych jest stosunkowo małe (wcześniej realizowane były projekty w ramach Phare[1]), również w odniesieniu do ewaluacji realizowanych projektów. Opisywane tutaj badanie było pierwszym zlecanym przez KPRM badaniem ewaluacyjnym. Dla ewaluatorów ocena projektów szkoleniowych dla członków s.c. – pomimo zgromadzonych licznych doświadczeń w ewaluacji projektów realizowanych w ramach SPO RZL – było przedsięwzięciem bezprecedensowym, którego głównym wyzwaniem była skomplikowana, składająca się z niejednorodnych komponentów, struktura organizacyjna s.c., jej specyfika oraz zasięg terytorialny.

Badanie objęło dwa projekty zrealizowane przez Departament Służby Cywilnej w KPRM (do 26 października 2006 roku Urząd Służby Cywilnej): *Szkolenia generalne* oraz *Szkolenia językowe – I etap* dla członków korpusu s.c., zrealizowanych przy wsparciu SPO RZL w ramach Działania 2.4 Sektorowego Programu Operacyjnego *Rozwój Zasobów Ludzkich – Wzmocnienie zdolności administracyjnych* współfinansowanych przez Europejski Fundusz Społeczny. Pierwszy projekt – *Szkolenia generalne* – był wdrażany w okresie od 1 listopada 2004 roku do 30 września 2006 roku. Celem ogólnym projektu było zapewnienie profesjonalnego i efektywnego wypełniania zadań administracji publicznej, wynikających m.in. z członkowstwa Polski w Unii Europejskiej. W ramach projektu przeszkolonych zostało ponad **1500** członków korpusu s.c. w ramach 8 bloków tematycznych, w 16 regionach kraju. Drugi projekt – *Szkolenia językowe – I etap* był wdrażany w okresie od 1 grudnia 2004 roku do 31 grudnia 2006 roku. Miał na celu intensywne szkolenie językowe ponad **400** członków korpusu s.c., zajmujących stanowiska samodzielne, koordynujące i kierownicze w s.c., ze szczególnym uwzględnieniem pracowników zaangażowanych w realizację zadań wynikających z członkowstwa Polski w Unii Europejskiej. Udział w projekcie rozpoczęło 413 pracowników. Celem ogólnym badania była ocena efektywności zrealizowanych projektów. Badanie rozpoczęło się 15 lutego 2008 roku, tak więc do zakończenia działań szkoleniowych od rozpoczęcia oceny upłynęło 14 miesięcy w przypadku szkoleń językowych, a w przypadku szkoleń generalnych – 17 miesięcy. Tak długi okres w kontekście

---

[1] Były Urząd Służby Cywilnej uczestniczył w realizacji czterech projektów współfinansowanych ze środków programu Phare:

- PHARE 1997 – *Ocena potrzeb w zakresie szkoleń związanych z integracją europejską*,
- PHARE 1999 – *Wzmocnienie administracji i zasobów ludzkich w celu przygotowania Polski do członkostwa w Unii Europejskiej* (PL99/IB/OT/1A),
- PHARE 2001 – *Podnoszenie świadomości członków korpusu służby cywilnej na temat dylematów etycznych* (PL0101.15),
- PHARE 2003 – *Wzmocnienie zdolności administracyjnych* (PL03/IB/OT-06).

wysokiej fluktuacji kadr w administracji oraz natężenia działań szkoleniowych skierowanych do tej samej grupy docelowej był brzemienny w skutkach dla rezultatów badania (szerzej w podrozdziale 4. *Proces badawczy*).

2.2. Rodzaj ewaluacji

Zlecona ocena była **ewaluacją szczegółową *ex post***. Cele ewaluacji wskazywały na jej charakter:

- **konkluzywny** – w zakresie efektów działań podejmowanych w ramach projektów w odniesieniu do wybranych beneficjentów.
- **formatywny** – w kwestii określenia skuteczności i wpływu wsparcia świadczonego w ramach projektu, które posłużą sformułowaniu rekomendacji dla przyszłych tego typu działań szkoleniowych.

2.3. Jaki był stan wdrażania i kontekst instytucjonalny badania?

Charakterystyka projektów szkoleniowych realizowanych przez DSC KPRM w ramach Działania 2.4. SPO RZL, zaprezentowana w tabeli 1, ilustruje skalę przedsięwzięć podejmowanych w ramach Działania 2.4. W tej skali wnioski z ewaluacji pojedynczego projektu szkoleniowego mogły być znaczącym źródłem wiedzy i sugestii na temat skutecznego i efektywnego planowania i zarządzania szkoleniami dla członków korpusu s.c. w przyszłym okresie. Świadomość ta zwiększała jeszcze doniosłość przedsięwzięcia oraz poczucie odpowiedzialności ewaluatorów za wyciągnięte wnioski i sformułowane rekomendacje.

Warto w tym miejscu wspomnieć o tym, że realizacja badanych projektów dokonywała się w czasie, kiedy zmieniła się ekipa rządząca i nastąpiła zmiana statusu s.c., a także struktura organizacyjna Urzędu Służby Cywilnej, zlikwidowanego w październiku 2006 roku (szerzej o tym w podrozdziale 4.3). Jego funkcje przejęła Kancelaria Prezesa Rady Ministrów. To zaburzenie stabilności instytucjonalnej, a w wielu przypadkach, ciągłości personalnej, było przyczyną trudności w odtworzeniu wiedzy na temat projektu i procesu jego realizacji, gdyż wraz z odejściem osób zaangażowanych w ten proces wiedza ta została bezpowrotnie utracona. W związku z tym ewaluatorzy występowali w trakcie badania w funkcji „archeologów", odtwarzających wiedzę o przeszłości ze sprawozdań oraz z przekazów słownych osób, które nie uczestniczyły od początku w projekcie.

2.4. Kim byli interesariusze badania?

Niewątpliwie opisywany kontekst stwarzał ogromne pole do popisu dla triangulacji metod i wielokrotnej wzajemnej weryfikacji informacji

Tabela 1. Charakterystyka projektów szkoleniowych realizowanych przez DSC KPRM w ramach Działania 2.4. SPO RZL

Nazwa Projektu	Wielkość Projektu	Rodzaj Beneficjenta Ostatecznego	Lokalizacja Projektu
Szkolenia dla dyrektorów generalnych	0,5–1 mln	dyrektorzy generalni urzędów administracji rządowej	Projekt ogólnopolski
Szkolenia językowe	1–10 mln	członkowie korpusu służby cywilnej zajmujący stanowiska samodzielne, koordynujące i kierownicze	Projekt ogólnopolski
Szkolenia generalne	1–10 mln	członkowie korpusu służby cywilnej zajmujący stanowiska samodzielne, koordynujące i kierownicze	Projekt ogólnopolski
Szkolenia na temat Unii Europejskiej	0,5–1 mln	członkowie korpusu służby cywilnej	Projekt ogólnopolski
Pomoc doradcza oraz szkoleniowa dla służby cywilnej	1–10 mln	członkowie korpusu służby cywilnej, w tym m.in.: dyrektorzy generalni, kierownicy komórek kadr i szkolenia, grupa trenerów	Projekt ogólnopolski

Źródło: *Charakterystyka projektów szkoleniowych realizowanych w ramach Priorytetu I i II Sektorowego Programu Operacyjnego Rozwój Zasobów Ludzkich*, raport dla Instytucji Zarządzającej Podstawami Wsparcia Wspólnoty w Ministerstwie Gospodarki i Pracy, Warszawa, sierpień 2005.

pochodzących z różnych źródeł i od różnych interesariuszy badania, którymi byli:

- pracownicy Instytucji Wdrażającej projekty w ramach Działania 2.4 SPO RZL, a także projektodawcy w ramach tego Działania;
- pracownicy urzędów administracji publicznej biorący udział w projektach szkoleniowych organizowanych w ramach Działania 2.4 SPO RZL;
- pracownicy komórek kadr i szkoleń w korpusie s.c.;
- członkowie korpusu s.c.;
- podmioty/osoby uczestniczące we wdrażaniu Priorytetu V PO KL *Dobre rządzenie* w zakresie szkoleń, także po stronie potencjalnych projektodawców.

3. WKŁAD

3.1. Na jakie potrzeby informacyjne miała odpowiedzieć ewaluacja?

Generalnym celem zamówienia była **ocena efektywności projektów** zrealizowanych przez Departament Służby Cywilnej w KPRM (do 26 października 2006 roku Urząd Służby Cywilnej) *Szkolenia generalne* oraz *Szkolenia językowe – I etap* dla członków korpusu służby cywilnej. KPRM zleciła ewaluację zrealizowanych w ramach Działania 2.4. SPO RZL działań szkoleniowych „w przeddzień" formułowania założeń do programu szkoleń w ramach kolejnego okresu programowania 2007–2013 w ramach Priorytetu V PO KL *Dobre rządzenie*. Ewaluatorzy mieli więc przed sobą perspektywę badania, które miało przynieść efekty w postaci wskazówek do opracowania kierunków interwencji w obszarze szkoleń w kolejnych latach. Wyniki ewaluacji miały posłużyć bardziej precyzyjnemu określeniu wymagań wobec potencjalnych wykonawców projektów szkoleniowych i doradczych dla korpusu służby cywilnej, oraz, w dalszej kolejności, pozwolić na lepszą identyfikację potrzeb szkoleniowych członków korpusu s.c. Istotne zatem, w kontekście tego badania, było wypracowanie odpowiednich wniosków w celu wyeliminowania rozwiązań, które nie sprawdziły się w ramach SPO RZL i wypracowanie rekomendacji, służących właściwemu i trafnemu programowaniu wsparcia sektora administracji publicznej z PO KL.

3.2. Jaki był zakres badania?

Zakres przedmiotowy oceny Programu, który sprecyzował Zamawiający, obejmował następujące zagadnienia badawcze:

- ocena zgodności (odpowiedniość) projektów szkoleniowych (bądź jej brak) z potrzebami uczestników szkoleń,

- określenie wpływu szkoleń na wzrost kwalifikacji zawodowych i sytuację zawodową uczestników szkoleń,
- odpowiedź na pytanie, na ile umiejętności zdobyte podczas szkoleń przydatne są ich uczestnikom w codziennej pracy,
- ocena poziomu satysfakcji ostatecznych beneficjentów z uczestnictwa w projekcie oraz określenie wpływu szkoleń na rozwój osobisty pracowników w nich uczestniczących,
- ocena trwałości (bądź nietrwałości) efektów przeprowadzonych szkoleń.

3.3. Jakie były założenia metodologiczne SOPZ?

Dostarczenie odpowiedzi na pytania badawcze sformułowane przez Zamawiającego i rozszerzone przez Wykonawcę, miało dokonać się przy użyciu metod wskazanych w *Opisie Przedmiotu Zamówienia* oraz innych zaproponowanych przez Wykonawcę. Zamawiający wymagał przeprowadzenia badania ilościowego z zastosowaniem badania sondażowego poprzedzonego analizą danych zastanych. Wymaganą przez Zamawiającego techniką przeprowadzenia badania sondażowego był wywiad kwestionariuszowy przeprowadzony z zastosowaniem narzędzia CATI (Computer Assisted Telephone Interview), czyli telefonicznego wywiadu kwestionariuszowego, lub CAWI (Computer Assisted Web Interview), czyli kwestionariusza internetowego. Ze względu na zakres danych dostępnych w bazie PEFS, ostateczna decyzja o wyborze narzędzia miała być podjęta po podpisaniu umowy i konsultacji Wykonawcy z Zamawiającym. Zamawiający wymagał przedstawienia przez Wykonawcę proponowanego doboru próby badawczej (metoda i wielkość), przy czym minimalna próba badawcza została określona na poziomach – 20% dla szkoleń generalnych i 30% dla szkoleń językowych populacji uczestniczącej w szkoleniach.

3.4. Jakie były warunki operacyjne badania?

Wymogiem Zamawiającego było, aby badanie prowadził co najmniej trzyosobowy zespół o udokumentowanym doświadczeniu w realizacji badań społecznych i ewaluacyjnych. Zamawiający dał Wykonawcy 84 dni na wykonanie całego badania oraz wyznaczył tzw. kamienie milowe badania:

Jak widać w tabeli 2, Zamawiający dał Wykonawcy najwięcej czasu na fazę strukturalizacji (cztery tygodnie), podczas gdy na fazę obserwacji (badania terenowe), fazę analizy i fazę przygotowania raportu końcowego, zaledwie po 2 tygodnie. Większość zamówień badań ewaluacyjnych składanych przez sektor publiczny cechuje pośpiech – opisywane badanie nie jest pod tym względem wyjątkowe. Możliwość szybkiego i sprawnego wykonania badania w wyznaczonym przez Zamawiającego

Tabela 2. Harmonogram badania przedstawiony przez Zamawiającego w *Szczegółowym Opisie Przedmiotu Zamówienia*

Produkty	Planowany termin zakończenia działania lub oddania produktu
Wstępna wersja raportu metodologicznego	W ciągu 14 dni od podpisania umowy
Ostateczna wersja raportu metodologicznego	W ciągu 28 dni od podpisania umowy
Przeprowadzone badania ilościowe oraz jakościowe	W ciągu 42 dni od podpisania umowy
Wstępne założenia do raportu ewaluacyjnego	W ciągu 56 dni od podpisania umowy
Projekt raportu końcowego	W ciągu 70 dni od podpisania umowy
Raport końcowy z badania (wersja ostateczna)	W ciągu 84 dni od podpisania umowy
Przekazanie całości efektów badania ewaluacyjnego	W ciągu 84 dni od podpisania umowy

Źródło: opracowanie własne.

terminie stanowi o przewadze konkurencyjnej Wykonawcy, jednak należy w tym miejscu zwrócić uwagę na ogromne ryzyko dla jakości wyników badania, jakie niesie za sobą takie napięcie harmonogramu. Praktycy badań społecznych wiedzą, że nawet „listonosz puka zawsze dwa razy", dlatego jako standard procedury badawczej przyjmują co najmniej trzykrotną próbę skontaktowania się z respondentem (w przypadku próby imiennej)[2]. Zasada ta stoi na straży rzetelności badań, jej przestrzeganie wymaga jednak czasu. W przypadku wywiadów CAWI, kiedy zwrot ankiet przedłuża się, konieczne są monity. W przypadku opisywanego badania, tak krótki czas przeznaczony na fazę gromadzenia danych wtórnych i zaskakująco (jak na tę grupę zawodową) niska skłonność do wzięcia udziału w badaniu – były przyczyną niezrealizowania próby w założonej przez Wykonawcę wielkości. Z kolei krótki okres na dostarczenie wstępnej wersji raportu metodologicznego, wyznaczony przez Zamawiającego, znacznie skracał czas poświęcony analizie bogatego materiału badawczego, zgromadzonego w fazie obserwacji.

Brak dostatecznej ilości czasu – wynikający z konieczności sprostania wymaganiom harmonogramu narzuconego przez Zamawiającego – groził tym, że zebrane dane nie zostaną gruntownie przeanalizowane, co w konsekwencji oznaczać mogło zaprzepaszczenie trudu badawczego.

[2] Zob. np. Sawiński, Sztabiński 2000.

Ostatecznie, tego ryzyka udało się Wykonawcy uniknąć przez zatrudnienie licznego i wykwalifikowanego personelu badawczego, co jednak wiązało się ze zwiększeniem kosztów badania po stronie Wykonawcy.

Wykonawca dysponował budżetem w wysokości 60 tysięcy PLN na zrealizowanie badania spełniającego wymagania postawione przez Zamawiającego oraz na wszystkie dodatkowe działania zaproponowane w złożonej przez siebie ofercie. Suma ta okazała się wystarczająca na pokrycie kosztów operacyjnych badania, które pochłonęły większość budżetu skalkulowanego przez Wykonawcę. Wynagrodzenie za wkład intelektualny ewaluatorów zajmowało w nim najmniejszą kwotowo pozycję. Warto jednak wspomnieć, że udział w projekcie realizowanym na zlecenie Kancelarii Prezesa Ministrów dawał badaczom satysfakcję pozafinansową, wynikającą z prestiżu projektu.

Na potrzeby realizacji badania, Zamawiający zobowiązał się do udostępnienia bazy danych PEFS oraz wszystkich niezbędnych dokumentów związanych z projektem, takich jak wnioski projektowe, sprawozdania oraz umowy z wykonawcami. Najwięcej problemów nastręczała jakość danych z bazy PEFS. Brakowało w niej adresów e-mailowych, co stawiało poważne ograniczenia dla zastosowania narzędzia CAWI (ankiety internetowej), ale było zrozumiałe ze względu na to, że mogło odwzorowywać rzeczywistość – nie wszyscy pracownicy urzędów mają dostęp do internetu. Trudniej jednak wytłumaczyć brak numerów telefonów kontaktowych. W wielu miejscach pojawiały się jedynie numery telefonów komórkowych, nierzadko prywatnych, w innych numery telefonów domowych. Nieujednolicenie zasad w tym względzie okazało się brzemienne w skutki dla badania sondażowego prowadzonego metodą CATI.

4. PROCES BADAWCZY

4.1. Jakie podejście badawcze przyjęli ewaluatorzy?

Oczekiwania Zamawiającego sprowadzały się do odpowiedzi na pytania o produkty i rezultaty zrealizowanych projektów oraz ich oddziaływanie na uczestników szkoleń. Wreszcie, sprowadzały się do odpowiedzi na pytanie, w jakim stopniu projekty te przyczyniły się do realizacji celów Działania – wzmocnienia zdolności administracyjnych s.c. Tak zakreślone potrzeby informacyjne Zamawiającego sugerowały wybór podejścia badawczego zorientowanego na cele. Ambicją zespołu ewaluatorów realizującego badanie z ramienia Wykonawcy – firmy CASE – Doradcy Sp. z o.o. – było, oprócz dostarczenia odpowiedzi na pytania o jakość i efektywność szkoleń zaoferowanych pracownikom s.c. w ramach badanych projektów, dostarczenie wiedzy o procesie zarządzania i wdrażania tych projektów. Podejście zorientowane na zarządzanie znalazło wyraz w propozycji Wykonawcy, dotyczącej dodania do wskazanych przez

Zamawiającego kryteriów (efektywność, skuteczność, użyteczność, wpływ, trwałość) jeszcze jednego: skuteczności zarządzania. Wraz z rozszerzeniem asortymentu kryteriów ewaluacyjnych i idącym za tym dodaniem zagadnień badawczych, Wykonawca zobowiązał się do oszacowania jakości dotychczasowych procesów, wskazania mocnych i słabych stron podjętych działań oraz do sformułowania wniosków i rekomendacji do wdrożenia w przyszłym okresie planowania, co w konsekwencji miało przyczynić się do sprawniejszego i skuteczniejszego zarządzania działalnością szkoleniową korpusu służby cywilnej.

Minimum metodologiczne przedstawione przez Zamawiającego w OPZ zostało przez zespół ewaluatorów mocno poszerzone już na etapie oferty – zdecydowano się zaproponować kombinację narzędzi przy realizacji badania sondażowego, stosując sekwencyjnie CAWI, a następnie CATI. Przyczyną takiego rozwiązania była próba przezwyciężenia niedoskonałości bazy danych PEFS przez dywersyfikację źródeł kanałów dostępu do respondentów. Spektrum metod badawczych również zostało mocno rozszerzone przez Wykonawcę, głównie o techniki jakościowe – Indywidualne Wywiady Pogłębione (IDI) oraz Grupowe Wywiady Zogniskowane (FGI) z przedstawicielami wszystkich kategorii interesariuszy (IŻ, IP, IŻ, Projektodawca, Beneficjenci Ostateczni, Pracownicy komórek kadr i szkoleń, Przedstawiciele Wykonawców szkoleń). Taki dobór informatorów miał ugruntowanie w teorii zarządzania zasobami ludzkimi. Wybrana grupa respondentów składała się z przedstawicieli „kwintetu szkoleniowego”[3]. Zwiększeniu uległa też, zaproponowana przez Zamawiającego, próba badawcza do badania sondażowego.

4.2. Jak przebiegał proces badawczy?

Plan realizacji badania można podzielić na 4 fazy:

- strukturalizację,
- obserwację,
- analizę i ocenę,
- fazę przygotowania raportu i rekomendacji

zgodne zarówno co do zakresu, jak i harmonogramu z wymaganiami sformułowanymi w SIWZ oraz ze standardami metodologii MEANS.

- Faza strukturalizacji miała na celu stworzenie szczegółowej metodologii badania ewaluacyjnego, zaprezentowanej w raporcie metodologicznym.

---

[3] Według jednego ze specjalistów zarządzania zasobami ludzkimi – Leslie Rae (2004) – w proces zarządzania i wdrażania szkoleń powinni być zaangażowani przedstawiciele pięciu grup posiadających największą wiedzę i doświadczenie w tym zakresie, czyli: uczestnicy szkoleń, kierownicy komórek ds. szkoleń, trenerzy, kierownicy komórek organizacyjnych, wyższa kadra kierownicza (w przypadku administracji np. dyrektorzy generalni urzędów) – tzw. kwintety szkoleniowe.

- Faza obserwacji przewidywała zebranie danych wtórnych; przeprowadzenie zaplanowanych badań, przygotowanie pierwszych raportów cząstkowych, a także ich przetworzenie.
- Faza analizy i oceny miała na celu interpretację danych zgromadzonych w trakcie badania oraz ich ocenę, wnioskowanie i formułowanie rekomendacji.
- Faza przygotowania raportu i prezentacji rekomendacji zakładała przeprowadzenie warsztatów/dyskusji z udziałem Zamawiającego o najważniejszych kwestiach wynikających z ewaluacji i o rekomendacjach. W efekcie konsultacje polegały na przygotowaniu i zaprezentowaniu Zamawiającemu wstępnej wersji raportu, i uwzględnieniu jego uwag i sugerowanych zmian w wersji ostatecznej. Prezentacja głównych wniosków i rekomendacji miała miejsce w dwa miesiące po przekazaniu ostatecznej wersji raportu, na spotkaniu z udziałem przedstawicieli beneficjenta systemowego w Działaniu 5.1 PO KL (projektodawca w Działaniu 2.4 SPO RZL) w KPRM, Departamentu Służby Cywilnej i Państwowego Zasobu Kadrowego (pracownicy wszystkich wydziałów realizujących projekty w nowej perspektywie) oraz IW.

4.3. Jakie były wyzwania i ograniczenia procesu?

Największym wyzwaniem podczas badania okazała się wspomniana nieciągłość personalna.

Po pierwsze, była ona spowodowana opisywaną już zmianą organizacyjną (przekształcenie USC w KPRM), która wpłynęła na odpływ lub wymianę kadr. Po wtóre, brała się ona z pierwotnej koncepcji Beneficjenta – powierzania prac związanych z wdrażaniem Projektu na podstawie umowy zlecenia osobom spoza własnego zasobu kadrowego. Wynikało to z pierwotnych niedoborów kadrowych Instytucji Beneficjenta. W początkowym okresie funkcjonowania Programu funkcje Beneficjenta i Instytucji Wdrażającej pełniły dwa Wydziały tego samego Departamentu. Ten „umowny" podział dobrze obrazuje, w jaki sposób rozwiązywano problemy będące wynikiem niedostatecznego, na tamte czasy, potencjału instytucjonalnego, koniecznego do wdrażania Działania 2.4. SPO RZL.

Brak ciągłości personalnej był również wynikiem koincydencji faktów – moment ewaluacji zbiegł się z czasem, w którym główne osoby zaangażowane w projekt po stronie Instytucji Wdrażającej i Projektodawcy były na urlopie macierzyńskim i wychowawczym. W związku z powyższym ci, którzy mieli odgrywać rolę najbardziej kompetentnych informatorów, byli niedostępni.

Ewaluatorzy realizujący badanie skonfrontowali się ponadto z problemem „dyspersji" wiedzy w urzędzie. Wąska specjalizacja pełnionych obowiązków – wynikająca zapewne z dużej ilości spraw danego typu

– spowodowała, że respondenci wywiadów indywidualnych mieli bardzo niekompletną wiedzę o projekcie. Zadaniem ewaluatorów stało się poskładanie tych pojedynczych elementów w spójną całość. Należy w tym miejscu wspomnieć, jak dużą rolę w uzyskaniu rzetelnych wyników i wiarygodnych wniosków w opisanej sytuacji odegrało zastosowanie zasad triangulacji (źródła danych zastanych oraz użycie różnych technik – ilościowych i jakościowych) oraz zaangażowanie w proces zbierania danych przedstawicieli wszystkich interesariuszy badania.

Drugim wyzwaniem podczas realizacji badania była skomplikowana struktura organizacyjna służby cywilnej. Badanie zakładało sformułowanie wniosków na temat zarządzania szkoleniami i efektywności zrealizowanych szkoleń relewantnych dla wszystkich urzędów wchodzących w skład systemu służby cywilnej. Ten natomiast składa się z tak zróżnicowanych i rozproszonych terytorialnie elementów jak: administracja rządowa centralna oraz terenowa (16 województw), która z kolei dzieli się na zespoloną i niezespoloną. Mówiąc obrazowo, badaną populację reprezentowały zarówno ministerstwa, izby skarbowe, jak i Urząd Okręgowego Inspektora Rybołówstwa Morskiego. Reprezentatywność wyników ze względu na zróżnicowanie terytorialne uczestników zwiększał zastosowany przez Wykonawcę zabieg warstwowania próby ze względu na województwo pochodzenia uczestników.

Zupełnie zaskakującym ograniczeniem w osiągnięciu założonych rezultatów badania okazała się niska motywacja uczestników szkoleń do wzięcia udziału w badaniu sondażowym. Wykonawca założył, że grupa docelowa badania jest zdyscyplinowana i skłonna do współpracy, dlatego przewidział wysoki zwrot z ankiet i niski poziom braku danych. Niestety rzeczywistość okazała się odmienna – kłopoty z dostępnością (spowodowane przez braki danych w bazie PEFS) powiększał jeszcze niski zwrot z ankiet CAWI i wysoki procent odmów udzielenia wywiadu CATI. Było to przyczyną niezrealizowania założonej przez Wykonawcę próby, choć liczba podjętych prób dalece przewyższała zadeklarowany przez Wykonawcę odsetek badanej populacji mający podlegać badaniu. Ostatecznie jednak zrealizowano próbę na poziomie wymaganym przez Zamawiającego w SIWZ.

Doświadczenie to legło u podstaw jednej z rekomendacji badania mówiącej o konieczności rozważenia przez Instytucję Wdrażającą wprowadzenia do umowy z beneficjentem końcowym zapisu zobowiązującego go do udzielania informacji zwrotnej na temat udzielonego wsparcia, przez okres 3 lat od zakończenia udziału w projekcie.

Zespół ewaluatorów w fazie obserwacji napotkał jeszcze jeden problem – respondenci badania – sami uczestnicy szkoleń, ale zwłaszcza przedstawiciele działów kadr i szkoleń – mieli kłopoty z ustosunkowaniem się do podlegających ewaluacji szkoleń. Nie potrafili zidentyfikować, o które szkolenia chodzi. Wynikało to z faktu, że badane szkolenia miały swoją kontynuację w kolejnych etapach. Projekty te były

analogiczne i nieznacznie przesunięte w czasie, co powodowało problemy z przyporządkowaniem wrażeń i doświadczeń do właściwego projektu szkoleń. Teoria badania efektywności szkoleń dopuszcza ewaluację *ex post* w okresie od 3 miesięcy do dwóch lat po ustaniu działań szkoleniowych (Bee, Bee 2004). Termin realizacji opisywanej oceny był więc całkowicie zgodny ze sztuką. Doświadczenie tego badania świetnie ilustruje fakt, że decyzji o czasie przeprowadzenia ewaluacji szkoleń *ex post* powinna towarzyszyć świadomość konsekwencji takiego, a nie innego wyboru momentu przeprowadzenia oceny. Badanie bliższe pierwszej wartości granicznej, mające zaletę „efektu świeżości" wiązać się jednak będzie z poważnymi ograniczeniami w ocenie oddziaływania projektu (oraz ocenie efektów szkolenia na wyższych poziomach czterostopniowej skali Kirkpatricka)[4]. Natomiast zwiększanie przedziału czasowego, zwłaszcza jeśli na jego przestrzeni będą miały miejsce inne inicjatywy szkoleniowe, ryzykuje „zatarciem" wspomnień i trudnością w odtworzeniu związków przyczynowo-skutkowych pomiędzy interwencją a jej rezultatami.

4.4. Jaką rolę odegrali poszczególni interesariusze w procesie badawczym?

Niestety, spodziewany i oczekiwany aktywny udział Zamawiającego w procesie badawczym był ograniczony. Natłok obowiązków wynikających z trwających nadal projektów finansowanych z Działania 2.4 SPO RZL, a także z planowania nowych działań w ramach Priorytetu V PO KL, niedawne zmiany organizacyjne i wymiana kadr na stanowiskach kierowniczych i specjalistycznych w Departamencie s.c. w KPRM sprawiły, że partycypacja Zamawiającego w bieżącym omawianiu wyników toczącego się badania była mała. Bardzo dobra, skądinąd, współpraca, sprowadzała się do udostępniania ewaluatorom wszystkich poszukiwanych przez nich dokumentów będących w posiadaniu Kancelarii oraz do wzajemnego przekazywania kolejnych wersji raportu via internet. Można nawet pokusić się o stwierdzenie, że w sytuacji, wspomnianej już kilkakrotnie, nieciągłości personalnej i zmiany struktury organizacyjnej, Zamawiający pokładał w ewaluatorach nadzieję na odtworzenie kontekstu i szczegółów procesu wdrażania badanych projektów szkoleniowych i wyciągnięcie z tych doświadczeń wniosków, na których będą mogły się oprzeć nowe osoby zaangażowane w proces.

Zamawiający wykazał się również dużą elastycznością, zezwalając na nieznaczne i uzasadnione modyfikacje terminów wykonywania

[4] Koncepcja Kirkpatricka (2001) opiera się na ocenie efektywności szkolenia na czterech poziomach: reakcji, uczenia się, zachowań i efektów. Poziomy te są rosnące zarówno w aspekcie perspektywy czasowej, w jakiej można badać ich mierniki (kolejny etap w coraz bardziej odległych od „interwencji" interwałach), jak i co do stopnia złożoności analiz koniecznych dla dokonania oceny.

poszczególnych etapów prac badawczych – miało to znaczenie zwłaszcza w przypadku fazy obserwacji – realizacji badań CAWI/CATI oraz wywiadów IDI.

5. PRODUKTY BADANIA

5.1. Jaka wiedza była efektem badania?

Wnioski z badania, w sferze dotyczącej zarządzania projektem, wskazywały na dysfunkcje systemu zarządzania szkoleniami, z których wiele zostało już wyeliminowanych w czasie pomiędzy zakończeniem badania a momentem ewaluacji lub ich instytucjonalne źródła były właśnie w trakcie reorganizacji (dokonane zmiany struktury były rezultatem pozytywnych i wychodzących naprzeciw nowym zadaniom przekształceń). Wiedza wygenerowana w procesie badawczym miała zatem głównie charakter operacyjny, odnoszący się przede wszystkim do tego, jak skutecznie zarządzać szkoleniami w służbie cywilnej. Wnioski z badania w tym aspekcie były zatem przydatne jako podbudowa do już przedsięwziętych działań naprawczych. Część wniosków dotyczących badania efektów szkoleń w urzędach (a raczej braku tego działania) i konieczności jego realizacji była powtórzeniem konkluzji z innych analiz dotyczących zarządzania szkoleniami w służbie cywilnej[5], co udowadniało, że wskazanie nań nie było przypadkowe, ale jest stałym źródłem problemów w efektywnym zarządzaniu szkoleniami w służbie cywilnej. Generalnie raport wskazywał elementy funkcjonujące nieprawidłowo i wymagające zmiany, źródła zdiagnozowanych problemów oraz propozycje rozwiązań, które mogłyby przyczynić się do uzyskiwania lepszych rezultatów w badanej sferze.

5.2. Jak wyglądał proces rozpowszechniania wyników?

Jednym z zapisów OPZ był wymóg – nałożony na Wykonawcę – opublikowania raportu z ewaluacji w formie książkowej w nakładzie 50 egzemplarzy (twarda okładka, papier kredowy). Punkt ten – druk niskonakładowej książki – był znaczącą, pod względem kosztów, pozycją w budżecie Wykonawcy. Należy jednak w tym miejscu zauważyć, że jest to interesujący przykład demonstrowania przejrzystości działań i dyscypliny finansowej administracji publicznej. Zabieg obciążenia Wykonawcy

[5] Zob. *Zarządzanie szkoleniami w służbie cywilnej – diagnoza i ocena efektywności. Raport opracowany w ramach projektu „Pomoc doradcza oraz szkoleniowa dla służby cywilnej”* (2006), Warszawa, USC; *Analiza potrzeb szkoleniowych w służbie cywilnej* (2005), Warszawa, USC. Obie publikacje można znaleźć na stronie Departamentu Służby Cywilnej w KPRM: www.dsc.kprm.gov.pl

obowiązkiem publikacji wyników ewaluacji realizuje zarówno wymóg racjonalnego gospodarowania środkami publicznymi, jak i transparentności – oto urząd administracji publicznej, pożytkujący publiczne pieniądze na inwestowanie w rozwój zawodowy urzędników, poddaje swoje działania obiektywnej ocenie i upublicznia jej wyniki.

Te i inne opracowania dotyczące szkoleń w służbie cywilnej, finansowane ze środków europejskich, dystrybuowane są i dostępne w Centrum Informacyjnym Służby Cywilnej w siedzibie KPRM przy ul. Szucha 2/4.

6. WNIOSKI ZE STUDIUM PRZYPADKU

6.1. Jakie czynniki wpłynęły na wykorzystanie ewaluacji?

Wyniki ewaluacji zostały zaprezentowane i poddane dyskusji na spotkaniu z przedstawicielami beneficjenta systemowego w Działaniu 5.1 PO KL (projektodawcą w Działaniu 2.4 SPO RZL) w KPRM – Departamentu Służby Cywilnej i Państwowego Zasobu Kadrowego, w tym z Zastępcą Dyrektora Departamentu, z pracownikami wszystkich wydziałów realizujących projekty w nowej perspektywie) oraz z reprezentantami Instytucji Wdrażającej SPO RZL (Biuro Dyrektora Generalnego KPRM). Rozmowy dotyczyły głównie możliwości wdrożenia rekomendacji przy projektowaniu szkoleń dla s.c. w nowym okresie programowania. Na spotkaniu wnioski z badania przedstawiał też wykonawca badania ewaluacyjnego – realizowanego symultanicznie z omawianym tutaj badaniem ewaluacyjnym – dwóch innych projektów szkoleniowych finansowanych ze środków SPO RZL: szkoleniowego oraz pomocy doradczej i szkoleniowej.

Rozdzielenie procesu badawczego tych inicjatyw wyniknęło z dopuszczonej przez Zamawiającego możliwości składania odrębnych ofert na realizację tych komponentów. W związku z tym, w wyniku wyboru najkorzystniejszych ofert, dwa komponenty badania realizowane były przez różne firmy badawcze. Pomimo różnych badanych populacji, odmiennej metodologii, wnioski z badań w dużej mierze pokrywały się, zwłaszcza tam, gdzie podkreślały konieczność dokonania rzetelnej i kompleksowej analizy potrzeb szkoleniowych w służbie cywilnej. Żywiołowość dyskusji, namysł nad operacjonalnością rekomendacji świadczą o tym, że wnioski i rekomendacje z badania były faktyczną podstawą do formułowania planów działania w nowym cyklu programowania.

Ogłoszone w czerwcu 2008 roku przez KPRM zamówienie na *Przeprowadzenie badania i analizę potrzeb szkoleniowych oraz opracowanie planu działania (strategii szkoleniowej) w służbie publicznej na lata 2008–2010* można traktować jako wprowadzenie w życie części sugestii zawartych w raporcie z opisywanej tutaj ewaluacji, a zatem jej pozytywne wykorzystanie.

6.2. Jakie funkcje spełniło badanie?

Można zdecydowanie stwierdzić, że ewaluacja przyczyniła się do wzmocnienia partnerstwa i poczucia współwłasności uczestników ewaluacji, chociażby przez fakt, że była jednym z pierwszych doświadczeń tego typu i dyskusja nad wnioskami stała się okazją do wspólnego namysłu nad przyszłością szkoleń w służbie cywilnej i efektywnością interwencji finansowanych ze środków europejskich w tym zakresie. Ewaluacja odegrała również rolę wspierania procesów uczenia się w sektorze publicznym przez zidentyfikowanie problemów, ich źródeł oraz wskazanie kierunków działań naprawczych. Pierwsze doświadczenia z ewaluacji projektów zrealizowanych przez KPRM były zaczątkiem budowania zdolności ewaluacyjnych instytucji, która w obecnym okresie programowania pełnić będzie funkcję beneficjenta systemowego Działania 5.1. PO KL. Niestety w momencie przygotowywania tego tekstu nie ma podstaw do formułowania wniosków na temat wykorzystania opisywanej ewaluacji dla celów lepszego planowania działań szkoleniowych oraz dla poprawy wdrażania projektów. Można mieć nadzieję, że wnioski i rekomendacje z ewaluacji będą podstawą budowania kryteriów wyboru ofert w ogłoszonym przetargu na przygotowanie strategii szkoleniowej dla s.c. na lata 2008–2010.

Wiele świadczy natomiast o tym, że ewaluacja spełniła w tym wypadku funkcję wzmocnienia odpowiedzialności w sektorze publicznym. Raport ewaluacyjny kładł bowiem silny nacisk na efekty badanych projektów szkoleniowych. Cele szczegółowe oceny, wskazane przez Zamawiającego w *Opisie Przedmiotu Zamówienia*, świadczą o tym, że znaczącym motywem do zlecenia ewaluacji było zbadanie skuteczności całego przedsięwzięcia szkoleniowego, w tym zwłaszcza określenie stopnia, w jakim zaspokojone zostały potrzeby uczestników w zakresie kształcenia oraz wymagania dotyczące efektów pracy i potrzeby instytucji. Jednak w związku z tym, że na szkolenie członków korpusu s.c. przeznacza się znaczące nakłady – których źródłem, jak zwykło się mówić, są pieniądze podatników (choćby i europejskich, jak w tym przypadku) – instytucje zarządzające tymi środkami przywiązują dużą wagę do uzasadnienia podejmowanych działań szkoleniowych przez wykazywanie, że korzyści płynące ze szkolenia przewyższają poniesione koszty.

6.3. Jakie siły i ograniczenia miało badanie?

Podsumowując, do mocnych stron badania ewaluacyjnego: *Ewaluacja Projektów: Szkolenia Generalne i Szkolenia Językowe – I Etap* zrealizowanych w ramach *Działania 2.4 Sektorowego Programu Operacyjnego Rozwój Zasobów Ludzkich – Wzmocnienie zdolności administracyjnych* współfinansowanych z Europejskiego Funduszu Społecznego, należy zaliczyć:

- synchronizację czasu zlecenia ewaluacji z momentem budowania planów działań w ramach nowego programu operacyjnego PO KL,

a w związku z tym możliwość wykorzystania w planowaniu interwencji w nowym okresie programowania wniosków i rekomendacji z ewaluacji projektów z poprzedniego okresu,

- przejrzystość procesu oceny oraz sposób upowszechniania jej wyników,
- wielość zastosowanych metod badawczych oraz włączonych w badanie grup interesariuszy (triangulacja),
- motywację instytucji do uczenia się.

Za słabe strony przedmiotowego badania można uznać:

- niestabilność instytucji w badanym okresie, zwłaszcza odpływ kadr zaangażowanych w badany projekt,
- niską motywację interesariuszy, zwłaszcza końcowych beneficjentów wsparcia do udziału w badaniu,
- niedoskonałości bazy danych uczestników szkoleń (brak numerów telefonów kontaktowych oraz adresów e-mail),
- krótki termin realizacji badania, zwłaszcza niedostateczną ilość czasu przeznaczonego na fazę obserwacji i analizy,
- niedoszacowanie kosztów badania, zwłaszcza publikacji raportu (niedopatrzenie Wykonawcy) – zmniejszające dochodowość przedsięwzięcia.

BIBLIOGRAFIA

Analiza potrzeb szkoleniowych w służbie cywilnej (2005), Warszawa, USC; www.dsc.kprm.gov.pl

Bee F., Bee R. (2004), *Narzędzia do oceny efektywności szkoleń*, Kraków, Oficyna Ekonomiczna.

Charakterystyka projektów szkoleniowych realizowanych w ramach Priorytetu I i II Sektorowego Programu Operacyjnego Rozwój Zasobów Ludzkich. Raport dla Instytucji Zarządzającej Podstawami Wsparcia Wspólnoty w Ministerstwie Gospodarki i Pracy (2005), Warszawa, Ministerstwo Gospodarki i Pracy.

Haber A. (red.) (2007), *Ewaluacja ex-post – Teoria i praktyka badawcza*, Warszawa, Polska Agencja Rozwoju Przedsiębiorczości.

Kirkpatrick D.L. (2001), *Ocena efektywności szkoleń*, Warszawa, Studio EMKA.

Olejniczak K. (2008), *Mechanizmy wykorzystania Ewaluacji – Studium Ewaluacji średniookresowych INTERREG III*, Warszawa, Wydawnictwo Naukowe Scholar.

Rae L. (2004), *Planowanie i projektowanie szkoleń*, Kraków, Oficyna Ekonomiczna.

Sawiński Z., Sztabiński P.B., Sztabiński F. (red.) (2000), *Podręcznik ankietera*, Warszawa, IFiS PAN oraz SMG/KRC Poland Media SA.

Zarządzanie szkoleniami w służbie cywilnej – diagnoza i ocena efektywności. Raport opracowany w ramach projektu „Pomoc doradcza oraz szkoleniowa dla służby cywilnej” (2006), Warszawa, USC, www.dsc.kprm.gov.pl

ROZDZIAŁ 5.3.

Martin Ferry
Karol Olejniczak

EWALUACJA *EX POST* PHARE-CBC POLSKA–NIEMCY 1999–2001[1]

1. WSTĘP

Analizowane w niniejszym rozdziale studium przypadku dotyczy ewaluacji *ex post* Programu Współpracy Przygranicznej PHARE Polska–Niemcy (*PHARE Cross-Border Cooperation programme between Poland and Germany* – PHARE CBC). Wybraliśmy je, gdyż było to jedno z pierwszych badań zrealizowanych w Polsce przez zespół polskich ekspertów. Jest to także, jak dotychczas, jedna z największych i najbardziej złożonych ewaluacacji efektów programów przeprowadzonych w naszym kraju.

2. KONTEKST I PRZEDMIOT EWALUACJI

2.1. Jaka była charakterystyka badanego programu?

Program Współpracy Przygranicznej PHARE Polska–Niemcy (Phare CBC) 1999–2001 w rzeczywistości składał się z trzech edycji programu, którego realizacja rozpoczęła się w 1994 roku (i był wdrażany rocznie do 2003 roku). Program był realizowany na terenach przygranicznych Polski i Niemiec (obejmując 8 podregionów w 3 polskich województwach: zachodniopomorskim, lubuskim i dolnośląskim)[2]. Z kilku przyczyn program

[1] Rozdział jest fragmentem szerszego raportu badawczego M. Ferry, K. Olejniczak *The use of evaluation in management of EU programmes – evidence from Poland*, Ernst & Young, Program "Sprawne Państwo", www.sprawnepanstwo.pl (XII 2008).

[2] W ciągu kolejnych edycji zasięg programu zmieniał się, koncentrując się na pasie powiatów przygranicznych.

został uznany zarówno przez Polskę, jak i przez Komisję Europejską za przedsięwzięcie szczególnie ważne. Po pierwsze, Komisja przykładała dużą wagę do programów współpracy przygranicznej, które uważała nie tylko za instrumenty wspierające rozwój gospodarczy, ale także za środek budowy powiązań ponad historycznymi granicami administracyjnymi, kulturowymi i społeczno-gospodarczymi. Granica polsko-niemiecka ma w tym kontekście szczególne znaczenie[3]. Po drugie, w porównaniu do innych programów Phare, realizowanych w Polsce w tym czasie, polsko-niemiecki program CBC 1999–2001 charakteryzował się znaczącym finansowaniem (około 150 mln EUR, w tym polskie współfinansowanie). Większość projektów obejmowała takie same działania jak „podstawowe" programy Phare, w tym transport i ochronę środowiska, ale program finansował także mniejsze projekty infrastrukturalne i rozwoju gospodarczego w celu zachęcenia do udziału lokalnych społeczności, euroregionów itp. W teorii wszystkie projekty miały zademonstrować element współpracy pomiędzy partnerami po obu stronach granicy.

2.2. Dlaczego podjęto badanie i kim byli interesariusze?

Ewaluacja została zamówiona przez instytucję zarządzającą programem – Władzę Wdrażającą Programy Europejskie – WWPE[4]. Główną przyczyną wdrożenia badania była ocena długoterminowych skutków interwencji. Ewaluacja edycji wdrażanych w latach 1999–2001 stanowiła kontynuację wcześniejszych badań, realizowanych w 2003 roku odnośnie do edycji PHARE-CBC, realizowanych w latach 1994–1999. Przyczyna nakładania się edycji z 1999 roku na zakres badań była bardzo prosta. Podczas badania ewaluacyjnego, niektóre z projektów były nadal w ostatniej fazie realizacji, stąd konieczność rewizyty w celu przeprowadzenia pełnej oceny ich wpływu.

Poza oceną wpływu oraz promocją wyników programu, Władza Wdrażająca miała jeszcze dwa powody do uruchomienia nowego badania. Pierwszym był proces uczenia się w ramach nowo uruchomionego Programu Sąsiedztwa na wschodniej granicy Polski (z Ukrainą i Białorusią). WWPE miała nadzieję, że dobre doświadczenia wyniesione z PHARE-CBC będą mogły zostać użyte w nowych programach INTERREG, które zostały uruchomione w Polsce po przystąpieniu do Unii Europejskiej. Drugą przyczyną był wymóg audytu zewnętrznego funduszu małych projektów. Władza Wdrażająca postanowiła połączyć te dwa zadania, by w praktyce audyt został uzupełniony ewaluacją. Ukrytym celem było wzmocnienie

[3] Programy PHARE-CBC zostały zaprojektowane jako przedsięwzięcia naśladujące programy INTERREG, realizowane po unijnej stronie granicy. Temat ten omówiliśmy bardziej szczegółowo w rozdziale 4.5 niniejszej książki.

[4] Poprzednio była to Władza Wdrażająca Program Współpracy Przygranicznej PHARE.

pozycji WWPE jako polskiego ośrodka know-how w dziedzinie zarządzania programów współpracy terytorialnej[5].

Stąd badanie posiada dwie różniące się między sobą zasadniczo grupy potencjalnych użytkowników – Władzę Kontraktującą oraz regiony i beneficjentów dawnych programów PHARE-CBC, a także przedstawicieli regionów wschodnich i potencjalnych użytkowników publicznych nowych programów terytorialnych.

Warto tutaj wspomnieć, że przejście od funduszy przedakcesyjnych do funduszy strukturalnych skutkowało podziałami terytorialnymi, które przynajmniej początkowo zagroziły ciągłości programowania. Przed ewaluacją *ex post* w latach 1999–2001 w WWPE działały oddzielne zespoły monitoringu i ewaluacji, odpowiedzialne za Phare CBC oraz przygotowania do programów INTERREG IIIA/Neighbourhood, które zastąpiły po akcesji programy CBC. Jednostka INTERREG/NP została obsadzona nowo przyjętymi pracownikami i była pozbawiona formalnych struktur operacyjnych w zakresie współpracy z jednostką PHARE. Stąd potencjał wykorzystania doświadczeń zdobytych w wyniku zaangażowania WWPE we wcześniejsze ewaluacje w ramach PHARE został poważnie ograniczony. Nie było to korzystne ze względu na nakładające się przepisy regulujące Phare CBC i INTERREG/NP.

Ewaluacja *ex post* Phare CBC zapewniła pierwszą możliwość wzmocnienia tych więzi organizacyjnych, w miarę angażowania członków zespołu INTERREG/NP w planowanie ewaluacji w celu wypracowania nauk *ex ante* dla celów przyszłego Programu Sąsiedztwa (*Neighbourhood Programme*) w Polsce wschodniej – jako „mini-ewaluacja" w ramach szerszej ewaluacji *ex post*.

3. WKŁAD EWALUACJI

Specyfikacja Istotnych Warunków Zamówienia analizowanego badania miała bardzo ogólny charakter. Określono następujące dwa główne cele:

- Oszacowanie zakresu, w jakim programy poprawiły sytuację społeczno-gospodarczą w regionach przygranicznych. Wyniki tego komponentu oczywiście byłyby istotne dla interesariuszy programu oraz szerzej pojętej społeczności tych regionów.

[5] WWPE (dawna WWPWP PHARE) była zaangażowana w realizację programów finansowanych przez UE od połowy lat 90. ubiegłego wieku. Nadzorowała realizację serii programów CBC oraz udzielała pomocy w realizacji programu Sprawiedliwości i Spraw Wewnętrznych Phare. W 2004 roku WWPE otrzymała status jednostki budżetowej, podlegającej Ministerstwu Spraw Wewnętrznych i Administracji. Działalność WWPE uległa dywersyfikacji, podczas gdy większość programów współpracy terytorialnej została przejęta przez Ministerstwo Rozwoju Regionalnego. Jednak w momencie rozpoczęcia badań, WWPE udzielała MRR pomocy w realizacji nowych programów INTERREG oraz wypracowywaniu nowych kompetencji.

- Określenie dobrych praktyk i doświadczeń wynikających z programu w celu informowania o rozwoju Programu INTERREG IIIA Polska-Białoruś-Ukraina. Wyniki tego komponentu były oczywiście przeznaczone dla programujących i interesariuszy, zaangażowanych w rozwój Programu Sąsiedztwa w Polsce wschodniej.

W ToR określono także następujące kwestie:

- minimalny poziom oczekiwanej metodologii wraz z wyraźnym wymogiem wizyt terenowych;
- ramy organizacyjne kontraktu – lista zadań takich jak: opis programu i kontekstu, ustanowienie kryteriów oceny, zbieranie danych, analiza i ocena, promocja wyników);
- wykaz oczekiwanych produktów – baza danych, główny raport z ewaluacji; wersja skrócona przedstawiająca główne wyniki ewaluacji; oraz krótszy raport określający dobre praktyki, skierowany do beneficjentów Programu Sąsiedztwa.

Poza powyższym określono bardzo niewiele szczegółowych wymogów dotyczących metodologii stosowanej przez ewaluatorów czy komponentów procesu badawczego. Decyzja co do stopnia szczegółowości i kierunku, uwzględnionych w SIWZ, wymaga osiągnięcia bardzo delikatnej równowagi. Podobnie jak dzieje się to w niniejszym wypadku, Istotne Warunki i Postanowienia ToR mogą przewidywać całkowitą swobodę wyboru metod lub szeroko określić takie metody (np. analiza biurowa danych z monitoringu czy bardziej szczegółowa analiza próby projektów) i pozostawić ewaluatorowi decyzję dotyczącą szczegółów metodologicznych (jak np. metody określenia próby czy potencjalne pokrycie ankiet). Pozostawienie niektórych z tych kwestii otwartych w Istotnych Warunkach i Postanowieniach ToR umożliwiło ewaluatorom wprowadzenie ich własnych pomysłów i zaproponowanie innowacji. Jednocześnie jednak nawet jeżeli aktorzy programu nie opisali swoich wymogów w ToR, to musieli je przemyśleć z wystarczającą przejrzystością i w takim stopniu uszczegółowienia, by umożliwić ocenę i porównanie składanych ofert. Szczegółowe, uprzednie przemyślenie tych kwestii byłoby także pomocne w uzmysłowieniu ewaluatorom wszelkich problemów związanych z danymi, które mogłyby ograniczać zastosowanie określonych opcji metodologicznych. Gdy ma się to na uwadze na etapie składania ofert oraz uzmysławia te kwestie potencjalnym ewaluatorom, oszczędza się czas, poświęcany na zapewnienie, że oferenci działają na podstawie znaczących założeń i proponują wyłącznie możliwe do zastosowania metody.

W tym miejscu należy podkreślić (co wspomnieliśmy w poprzednim punkcie), że WWPE podjęła także decyzję połączenia w jednym kontrakcie trzech różnych komponentów: ewaluacji *ex post*, audytu niewielkich funduszy projektowych oraz promocji wyników. Wymaga to od potencjalnych uczestników przetargów łączenia się w konsorcja, obejmujące zarówno doświadczenie w audytach, jak i w ewaluacji. Przetarg został

zrealizowany zgodnie z procedurą międzynarodową (procedura dwufazowa).

Kluczowym aspektem niniejszego studium przypadku jest czas i finansowanie, związane z ewaluacją oraz zakres możliwości tworzenia własnej metodologii pozostawiony ewaluatorom. Czas i finansowanie przeznaczone na realizację całego kontraktu wyniosły 500 tys. EUR oraz dziesięć miesięcy pracy (ze czego 1/3 została przeznaczona na ewaluację) postrzegano jako znaczące dla porównania z innymi ewaluacjami, podejmowanymi w kraju mniej więcej w tym samym czasie. WWPE wyznaczyła do procesu ewaluacji podstawowy zespół, składający się z trzech pracowników, aczkolwiek inne osoby odegrały określone role na różnych etapach realizacji procesu.

Badanie zostało zlecone konsorcjum, działającemu pod kierunkiem PricewaterhouseCoopers, aczkolwiek zadania ewaluacyjne były realizowane wyłącznie przez partnera konsorcjum – Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG) – Uniwersytet Warszawski. Podstawowy zespół ewaluacyjny składał się z sześciu ekspertów współpracujących z siecią dodatkowych współpracowników w poszczególnych sektorach.

4. JAK WYGLĄDAŁ PROCES EWALUACJI?

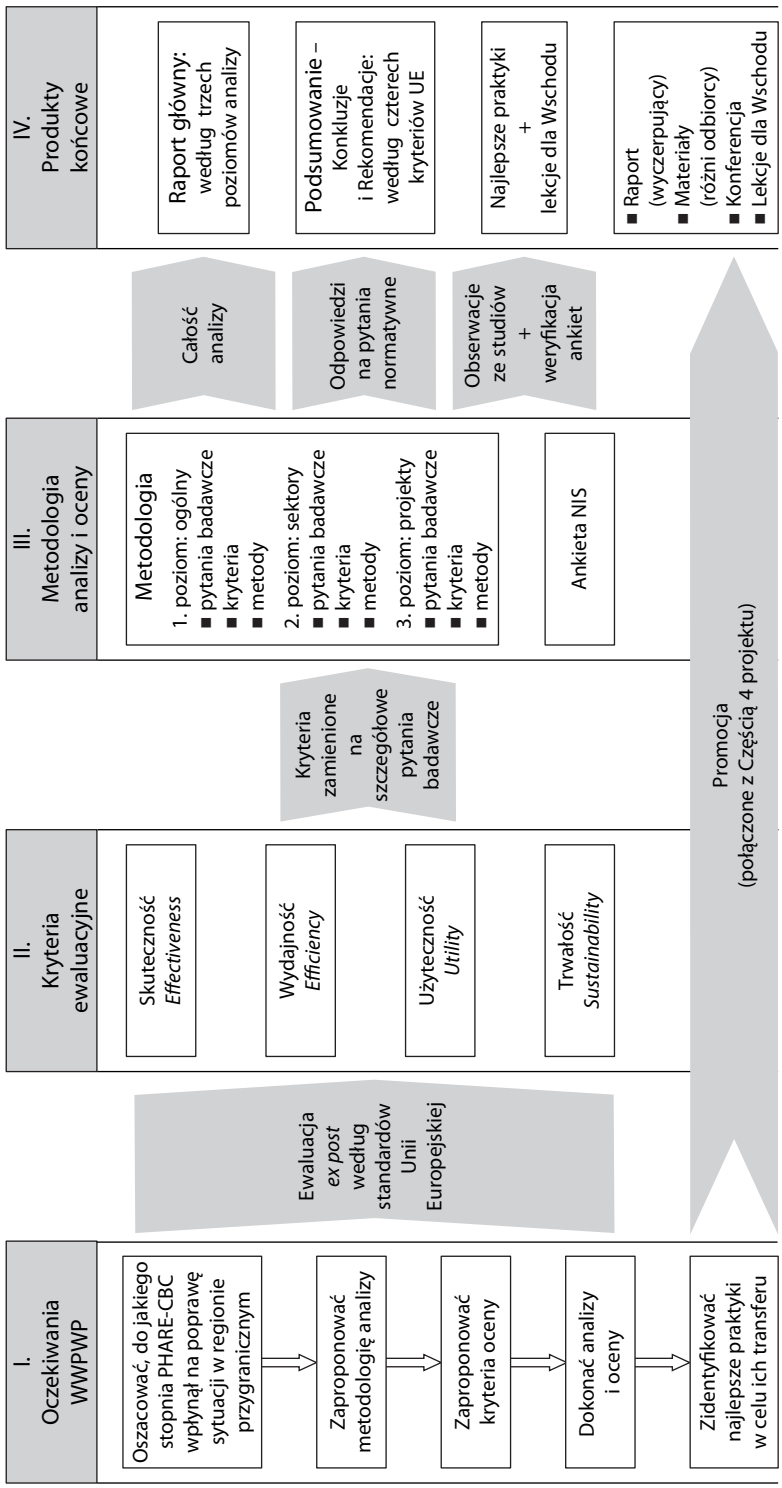
4.1. Jaką metodologię zastosowano w badaniu?

W ramach ewaluacji *ex post* głównym celem metodologii było zapewnienie informacji na temat procesów przebiegających w programie oraz rodzajów projektów, które przyniosły najlepsze wyniki w tego rodzaju programach. Zgodnie z tym zasoby przeznaczone na ewaluację dały zespołowi ewaluacyjnemu EUROREG możliwość opracowania nowatorskiej metodologii, łączącej *desk-research*, analizę związków przyczynowo-skutkowych z bardziej interaktywnymi technikami w celu ewaluacji skutków na poziomie zarówno całego programu, jak i indywidualnych projektów.

Mimo to ewaluacja stanęła w obliczu istotnych wyzwań. Głównym wyzwaniem było opracowanie metodologii, obejmującej szeroki zakres tematów i rodzajów projektów, wchodzących w skład programu PHARE CBC (obejmujących wszystkie duże projekty realizowane w ramach programu oraz mniejsze projekty infrastrukturalne – łącznie 118).

Opracowana metodologia składała się z czterech etapów (zob. rysunek 1). W etapie pierwszym określili ponownie głównie potrzeby klientów. W etapie drugim ewaluatorzy skoncentrowali się na głównych kryteriach ewaluacji, odpowiadających niniejszemu badaniu *ex post*.

W etapie trzecim kryteria te zostały przekształcone w szczegółowe pytania ewaluacyjne (opierające się na pytaniach z „drzewa pytań ewaluacyjnych" przedstawionych w niniejszej książce w rozdziale 1.1).



Rysunek 1. Struktura badania ewaluacyjnego

Źródło: EUROREG, PriceWaterhouseCoppers PL, IRS (2006).

W celu określenia stopnia złożoności programu podjęto decyzję o przyjęciu trzech poziomów analizy:

- Poziomu programowego (analiza ogólna kontekstu ekonomicznego, prawno-instytucjonalnego, społecznego oraz wpływu Programu).
- Poziomu sektorowego (analiza szczegółowych grup tematycznych projektów – transport, środowisko, turystyka) oraz grup typów projektów (duże, małe). Przeprowadzone w związku z powyższym badania opierały się na danych statystycznych i modelowaniu oraz wywiadach w połączeniu z wizytami na miejscu.
- Poziom projektu – analiza pytań szczegółowych związanych z projektem (wpływ na warunki lokalne, wskaźniki, problemy zarządcze, proces uczenia się).

Pytania (łącznie 40 – opisowych i normatywnych) opracowano dla każdego poziomu analizy.

Na tym etapie eksperci przypisywali do każdego pytania także zbiór metod i źródeł, wymaganych do rozwiązania problemu. Zarówno źródła danych, jak i metody poddano triangulacji.

W ostatnim etapie opracowano mapę produktów badania. Warto tutaj podkreślić, że kryteria ewaluacji były użyteczne nie tylko przy opracowywaniu pytań badawczych. Okazały się one także niezwykle istotne w ostatnim etapie prezentacji wyników. W raporcie skróconym, zamiast prezentować odpowiedź na każde 40 pytań, wnioski połączono i przedstawiono według czterech głównych kryteriów ewaluacji (skuteczności, wydajności, użyteczności, trwałości).

Metodologia przewidywała także opracowanie narzędzi najlepszych praktyk do zastosowania w programach przygranicznych w Polsce wschodniej. Wymagało to zaangażowania w proces ewaluacji przyszłych interesariuszy tych nowych programów, reprezentujących władze regionalne Polski wschodniej. Mimo że regiony wschodnie leżały poza obszarem objętym programem, pozyskanie partnerów z tych regionów, zaangażowanych w ewaluację uznano za istotne z powodu ich oczekiwanych korzyści z opracowywania podobnych programów na wschodzie kraju. Negatywne doświadczenia z poprzedniej ewaluacji wskazywały jednak na to, że pozyskanie zaangażowania wschodnich partnerów będzie wyzwaniem[6]. Opracowano więc specjalną strategię włączania interesariuszy programu Polska–Białoruś–Ukraina (zob. ramka 1).

W omawianej ewaluacji, poza opisanym w ramce 1 narzędziem transferu wiedzy do programów wschodnich, zastosowano także kompleksową metodologię oceny efektów programu PHARE-CBC Polska-Niemcy. Wykaz metod przedstawiamy w ramce 2.

---

[6] Podczas ewaluacji programów okresu 1994–1999 rozesłano do wschodnich województw zaproszenia na konferencje końcowe prozentujace wyniki badania. Niestety odzew ze strony instytucji publicznych wschodnich regionów Polski był znikomy.

Ramka 1

Narzędzie służące transferowi wiedzy z Zachodu na Wschód

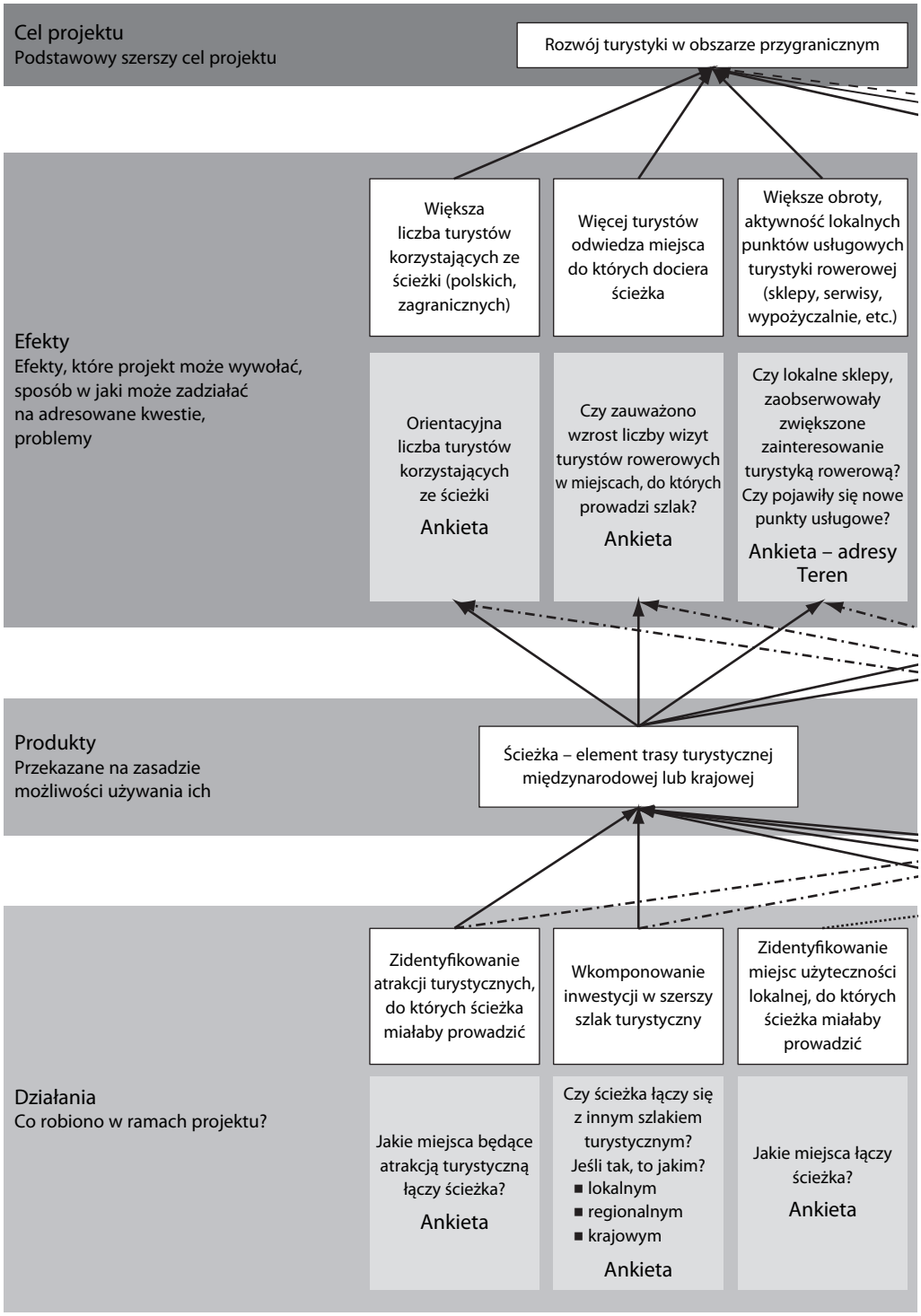
Ewaluatorzy opracowali ankietę, która została rozesłana do potencjalnych polskich beneficjentów programu Polska–Białoruś–Ukraina. W ankiecie pytano o główne potencjalne wyzwania i obawy, związane z przyszłym uczestnictwem w programach przygranicznych (z perspektywy instytucji aplikujących o fundusze). Do uczestników PHARE-CBC wysłano taką samą ankietę, aczkolwiek była to już ocena *post factum* – z perspektywy zakończonego programu; dodatkowo tych beneficjentów pytano o to, w jaki sposób przezwyciężyli oni napotkane problemy. Po przeprowadzeniu analizy odpowiedzi ewaluatorzy opracowali mapy kluczowych wyzwań, problemów oraz ich rozwiązań. Dla instytucji – przyszłych wnioskodawców ze Wschodu – opracowano specjalną broszurę, przedstawiającą główne wyzwania, sposoby stawiania im czoła oraz ilustracje faktycznych rozwiązań, stosowanych w projektach przygranicznych, realizowanych na Zachodzie.

Ramka 2

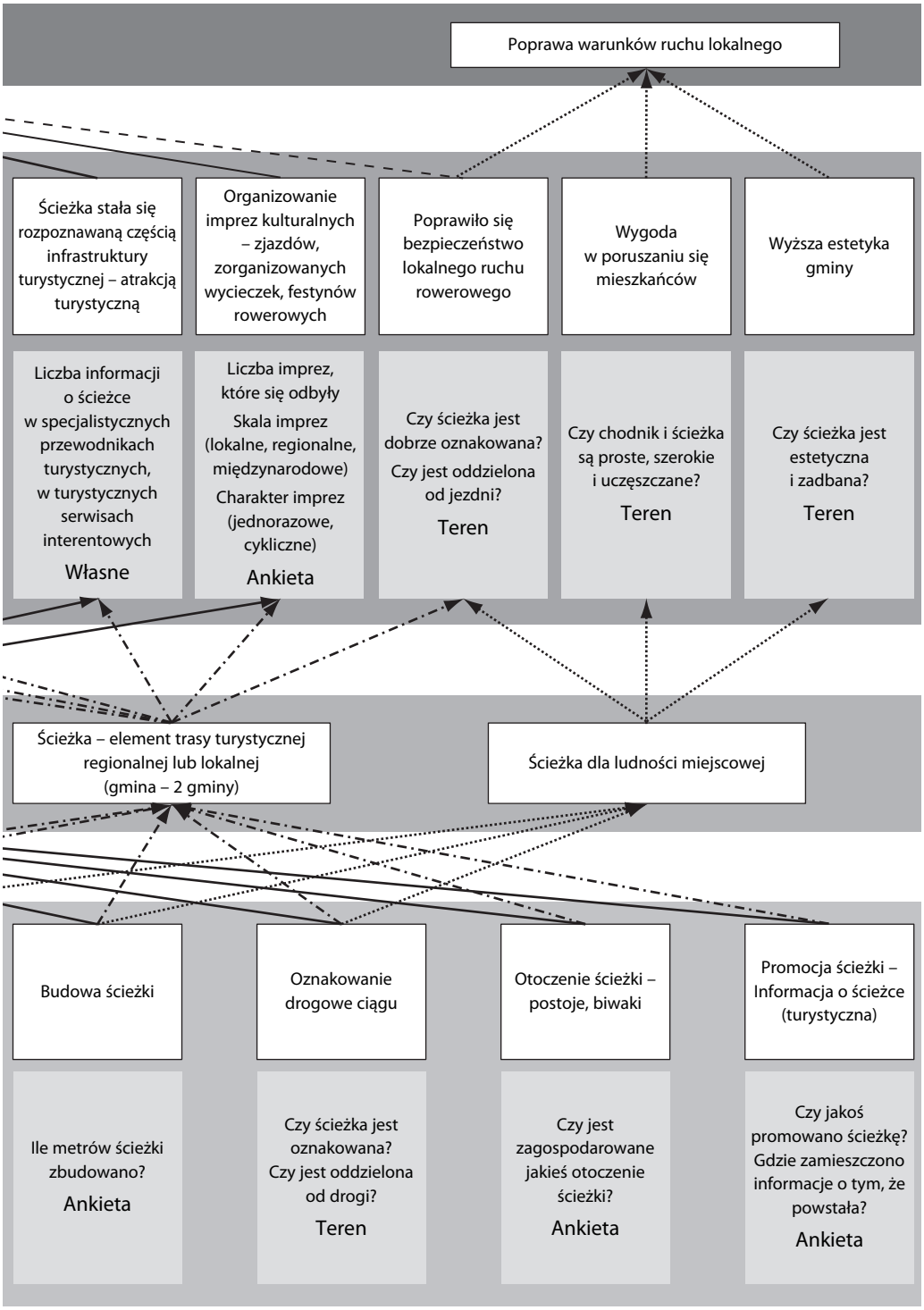
Skrócony opis metodologii ewaluacji PHARE-CBC Polska–Niemcy

- przegląd dokumentów programowych, dokumentacji z monitoringu;
- przegląd literatury tematu, badań poświęconych społeczno-gospodarczym aspektom rozwoju pogranicza polsko-niemieckiego;
- wywiady w instytucjach centralnych i regionalnych;
- wizyty studyjne (ok. 300) wszystkich zrealizowanych projektów;
- badania ankietowe beneficjentów programu;
- badania ankietowe potencjalnych beneficjentów (instytucje starające się o środki, odrzucone w trakcie konkursu oraz instytucje, które mogły, ale nie aplikowały o fundusze);
- badania ankietowe firm z obszaru przygranicznego;
- badania ankietowe firm wykonawców, realizujących kontrakty PHARE CBC;
- badania opinii społeczności obszaru przygranicznego (w tym badania młodzieży licealnej ze szkół przygranicznych „miast podzielonych”);
- kwerenda i analiza lokalnej prasy;
- matryce i modele logiczne;
- analizy statystyczne i przestrzenne;
- panele ekspertów (ocena wszystkich projektów), połączone z samoocenią beneficjentów;
- studia przypadków.

Źródło: EUROREG Warsaw University et al. 2006.



Rysunek 2. Analiza logiki i efektów projeków – Ścieżki rowerowe


Poprawa warunków ruchu lokalnego
Ścieżka stała się rozpoznawaną częścią infrastruktury turystycznej – atrakcją turystyczną
Organizowanie imprez kulturalnych – zjazdów, zorganizowanych wycieczek, festynów rowerowych
Poprawiło się bezpieczeństwo lokalnego ruchu rowerowego
Wygoda w poruszaniu się mieszkańców
Wyższa estetyka gminy
Liczba informacji o ścieżce w specjalistycznych przewodnikach turystycznych, w turystycznych serwisach interentowych
Własne
Liczba imprez, które się odbyły
Skala imprez (lokalne, regionalne, międzynarodowe)
Charakter imprez (jednorazowe, cykliczne)
Ankieta
Czy ścieżka jest dobrze oznakowana?
Czy jest oddzielona od jezdni?
Teren
Czy chodnik i ścieżka są proste, szerokie i uczęszczane?
Teren
Czy ścieżka jest estetyczna i zadbana?
Teren
Ścieżka – element trasy turystycznej regionalnej lub lokalnej (gmina – 2 gminy)
Ścieżka dla ludności miejscowej
Budowa ścieżki
Oznakowanie drogowe ciągu
Otoczenie ścieżki – postoje, biwaki
Promocja ścieżki – Informacja o ścieżce (turystyczna)
Ile metrów ścieżki zbudowano?
Ankieta
Czy ścieżka jest oznakowana?
Czy jest oddzielona od drogi?
Teren
Czy jest zagospodarowane jakieś otoczenie ścieżki?
Ankieta
Czy jakoś promowano ścieżkę? Gdzie zamieszczono informacje o tym, że powstała?
Ankieta

Użycie tak rozbudowanej metodologii – powiązanie podejścia jakościowego z ilościowym, uwzględnienie tematycznych grup ewaluacyjnych oraz przyjęcie wielopoziomowej perspektywy (programowa, sektorowa, projektowa) – było w tym czasie w Polsce podejściem nowatorskim.

Omawiając metodologię badania, należy jednak zwrócić uwagę na kilka istotnych kwestii. Po pierwsze, analiza ilościowych aspektów procesu badawczego została ograniczona z powodu luk w dostępnych statystykach bazowych i monitoringowych[7]. Po drugie, rozmiar finansowania interwencji (relatywnie mały w stosunku do wielkości objętego interwencją obszaru) uniemożliwiał przypisywanie zmiany w regionach wyłącznie działaniom programu. Ewaluatorzy próbowali więc prześledzić związki przyczynowo-skutkowe między projektami a ich efektami na niszym poziomie – lokalnym lub podsektorowym. Zamiast całościowych statystyk regionalnych zastosowano więc modele logiczne (przykład modelu logicznego ścieżek rowerowych przedstawiono na rysunku 2). Po czwarte, transfer wiedzy z Zachodu na Wschód był, mimo użytych strategii, z natury rzeczy ograniczony do szerszych kwestii (pomijał kwestie czysto techniczne – jak np. wzór wniosku, procedura finansowania). Wynikało to z tego, że współpraca prowadzona w obszarach przygranicznych na Wschodzie i Zachodzie miała inny kontekst społeczno-ekonomiczny oraz zupełnie odmienne procedury finansowe.

4.2. Jak przebiegał proces badawczy?

W prezentowanej ewaluacji *ex post* PHARE CBC 1999–2001 udało się zbudować wzajemne zaufanie pomiędzy instytucją zlecającą badanie a ewaluatorami. Wynikało to z tego, że obie strony miały już wcześniejsze pozytywne doświadczenia we współpracy (ewaluacja poprzedniego programu PHARE CBC 1994–1999) (Gorzelak et al. 2004). Dzięki temu badacze mieli też dobre rozeznanie w specyfice programu. Zlecający dał badaczom dużo swobody w opracowaniu metodologii szczegółowej oceny wyników programu, był za to na bieżąco włączany w proces – na szeregu spotkań roboczych przedstawiał uwagi na temat metodologii, opracowywanej przez EUROREG.

---

[7] Wskaźniki bazowe oraz postępu projektów służą określeniu postępów i wpływu programu oraz szczegółowemu informowaniu o postępach w realizacji poszczególnych projektów. Ponieważ program posiadał je w szczątkowej formie i były one niespójne, ewaluatorzy próbowali sami wypełnić luki. Odtworzenie *post factum* sytuacji sprzed rozpoczęcia projektów było często jednak niemożliwe.

5. PRODUKTY I WYKORZYSTANIE EWALUACJI

5.1. Jakie produkty przyniosło badanie?

Produktami badania były: główny raport z ewaluacji; wersja skrócona przedstawiająca główne wyniki ewaluacji (w układzie zgodnym z głównymi kryteriami ewaluacji); oraz krótszy raport określający dobre praktyki, skierowany do beneficjentów Programu Sąsiedztwa (EUROREG Warsaw University et al. 2006; Gorzelak et al. 2006)[8].

Główny raport z ewaluacji zawierał serię wniosków płynących z odpowiedzi na pytania szczegółowe dla każdego z 3 poziomów analizy (całościowej, sektorowej, poziomu projektów). Dominawały wnioski o charakterze koncepcyjnym. Były to głównie spostrzeżenia na temat mechanizmu współpracy przygranicznej oraz rzeczywistych form współpracy w ramach PHARE-CBC, pokazane na tle szerzych teorii współpracy ponadgranicznej i wyjściowych założeń programu. Główną obserwacją było to, że program miał przygraniczny, a nie ponadgraniczny charakter (ograniczony zakres równoległych projektów wzdłuż obszarów przygranicznych czy projektów, które w rzeczywistości przekraczały granice). Raport przedstawiał także wnioski z poszczególnych sektorów działania programu (ich podsumowanie, w kontekście szerszej analizy współpracy polsko-niemieckiej i w zestawieniu z wynikami ewaluacji pierwszych edycji PHARE-CBC można znaleźć w: Olejniczak 2008). Wniosków czysto tchnicznych w odniesieniu do PHARE-CBC było niewiele, z racji skończonego już wdrażania programu i braku jego kontynuacji w dotychczasowej formie.

Natomiast raport określający dobre praktyki dla przyszłego wschodniego Programu Sąsiedztwa miał dużo sugestii technicznych, zorientowanych na wdrażanie, prezentowanych jednak z perspektywy beneficjentów (instytucji wdrażających projekty, a nie cały program). Raport miał układ odzwierciedlający kolejne etapy starania się o środki i wdrożenia projektów (opracowanie i strukturyzacja koncepcji projektu, zbieranie dokumentacji, wdrażanie projektów, finalizacja projektów). Wszystkie kwestie uwzględnione pod tymi hasłami zostały zilustrowane studiami przypadków, pochodzącymi z programu PHARE CBC.

Dzięki temu, że w trakcie badania prowadzono ścisłe konsultacje pomiędzy ewaluatorami a WWPE, Zlecający był zadowolony z jakości technicznej produktów końcowych. Dostępność raportu końcowego, a zwłaszcza opracowanie skróconego opisu ustaleń zawartych w raporcie głównym zostały także ocenione pozytywnie. Dobre rozpowszechnienie

[8] Zob. http://www.wwpe.gov.pl/index.php?params[section_id]=4&params[category_id]=150

wyników było szczególnie ważne w przypadku lekcji i zaleceń dla Programu Sąsiedztwa. Wyniki tej części ewaluacji – prócz prezentacji na konferencji końcowej – wysłano bezpośrednio pocztą elektroniczną (w formie krótkiego przewodnika) do potencjalnych beneficjentów oraz prezentowano w regionach wschodnich, na seminariach dla instytucji aplikujących o środki.

5.2. W jaki sposób wykorzystano tę wiedzę?

Ewaluacja *ex post* nie wywiera oczywiście bezpośredniego wpływu na program, którego realizacja już się zakończyła. Warto jednak zwrócić uwagę na szerszy wpływ (wychodzący poza konkretny program i obszar działania), jaki wywarło to badanie na instytuję i zarządzane przez nią programy. WWPE przyjęła część zaleceń praktycznych przedstawionych w ewaluacji, uwzględniając je w innych programach (na przykład postulat stworzenia systemu budowania i monitoringu wskaźników projektów). Ponadto proces ewaluacji wywarł silny wpływ na kształtowanie własnej kultury organizacyjnej WWPE. Włączenie personelu instytucji w proces badawczy wzbogaciło ich doświadczenie ewaluacyjne. Co więcej, badanie po raz pierwszy zbliżyło do siebie dwa zespoły WWPE – odpowiedzialny za PHARE CBC i za programowanie INTERREG. Ewaluatorzy z EUROREG także wykorzystali ewaluację do wzbogacenia swojej wiedzy o rozwoju regionalnym oraz wzmocnienia doświadczenia w ewaluacji programów UE. Zespół opracował serię prezentacji, raportów i artykułów naukowych, które przyczyniły się do rozpowszechnienia ewaluacji. Ewaluacja zdecydowanie przyczyniła się do podniesienia stanu wiedzy o mechanizmach współpracy przygranicznej w Polsce. Ostateczny raport podniósł interesujące kwestie w odniesieniu do tego rodzaju współpracy oraz związanych z nią wyzwań[9].

W kategoriach transferu wiedzy na Wschód, można stwierdzić, że wnioski i lekcje z Zachodu niewątpliwie dotarły do grupy najbardziej zainteresowanych i potencjalnych użytkowników – przedstawicieli instytucji szczebla regionalnego i lokalnego na Wschodzie Polski. Trudno jest natomiast ocenić stopień wykorzystania tych rekomendacji. Ewaluacja była jednym z kilku działań informacyjnych, podejmowanych w ramach uruchamiania Programu Sąsiedztwa, trudno więc określić jej wpływ jednostkowy.

---

[9] Przykładami takich kwestii mogą być: przygraniczność zamiast ponadgraniczności projektów czy orientacja polskiej strony na projekty infrastrukturalne wynikająca z różnic w rozwoju między partnerami. Całościową dyskusję wniosków z obydwu ewaluacji *ex post* programów PHARE-CBC w kontekście teorii współpracy ponadgranicznej zaprezentowano w Olejniczak 2008.

6. WNIOSKI ZE STUDIUM PRZYPADKU

6.1. Jakie funkcje spełniło badanie?

Odnosząc opisany przypadek do szerokich funkcji ewaluacji opisanych w rozdziale 1.1 niniejszej książki, można stwierdzić, że badanie spełniło dwie funkcje – wzmocnienia odpowiedzialności oraz poznawczą.

Ewaluacja *ex post* wyraźnie pokazała zarówno skuteczność programu, jak i jego rzeczywiste efekty. Co ciekawe, ewaluatorzy odeszli od technicznego porównywania celów programowych ze stopniem ich wypełnienia, na rzecz kompleksowego porównania rzeczywistych efektów z sytuacją sprzed programu (ang. *Goal-free approach*). Badanie dało więc społeczeństwu i głównym interesariuszom wyraźną informację o efektach interwencji publicznej.

Z tą funkcją szła w parze funkcja poznawcza (wyjaśniająca). Badanie zidentyfikowało mechanizmy współpracy ponadgranicznej (czy w tym wypadku raczej przygranicznej) oraz ich główne determinanty. Co więcej, przyczyniło się do rozwoju krajowej metodologii badań efektów w polu współpracy terytorialnej. Wsparcie procesów uczenia się w sektorze publicznym nastąpiło również przez transfer wiedzy do programu na granicy wschodniej, a także przez rozpowszechnianie wiedzy w publikacjach naukowych. Trzeba jednak stwierdzić (o czym szerzej piszemy przy omówieniu czynników uczenia się), że budowanie wiedzy zostało poważnie osłabione przez zmiany instytucjonalne na poziomie centralnym.

Funkcja wzmocnienia partnerstwa i poczucia współwłasności uczestników była natomiast minimalna i pojawiła się nie kontekście badanego programu, a w kontekście zbliżania polskich interesariuszy (potencjalnych beneficjentów) Programu Sąsiedztwa Polska – Białoruś – Ukraina.

Funkcja poprawy wdrażania nie miała w tym wypadku żadnego zastosowania, ponieważ badany program nie był kontynuowany w dotychczasowej formie. Ostatnia edycja PHARE-CBC (2003) wciąż jeszcze wdrażana w momencie ukończenia ewaluacji *ex post* edycji 1999–2001 różniła się jednak w sposób znaczący w kwestiach proceduralnych. Podobnie było w przypadku nowych narzędzi współpracy na granicy polsko-niemieckiej po rozszerzeniu UE – Programy INTERREG IIIA Polska–Niemcy były zorganizowane w zupełnie odmienny sposób, według innych procedur i w innym kontekście instytucjonalnym. Nie było więc żadnego pola do technicznych rekomendacji związanych z procedurami i procesami. Podobnie rzecz się miała z funkcją poprawy jakości planowania. Nowa logika i sytuacja społeczno-gospodarcza znacznie różniły się od sytuacji z lat wdrażania PHARE-CBC. Jedynym potencjalnym polem wsparcia planowania przez wykonane badanie mogłyby być wnioski natury koncepcyjnej (roli i optymalnych form współpracy terytorialnej).

Takie wnioski – odnośnie do planowania – były stawiane przez ewaluatorów, nie ma też dowodów, by zostały użyte przez interesariuszy nowych programów INTERREG.

6.2. Jakie czynniki wpłynęły na wykorzystanie ewaluacji?

W wypadku niniejszego studium przypadku, kluczowym czynnikiem determinującym zastosowanie ewaluacji był kontekst instytucjonalny. Odnosi się to przede wszystkim do stabilności instytucjonalnej zlecającego badanie (Władzy Wdrażającej programy Europejskie) i ewaluatora oraz związków między nimi. Doświadczenie z poprzednich ewaluacji umożliwiło Zlecającemu opracowanie racjonalnych, ramowych wytycznych zamówienia na badanie. Doświadczenie ewaluatorów, połączone z elastycznym i otwartym podejściem Zlecającego, dało możliwość rozwinięcia innowacyjnego i szerokiego podejścia metodologicznego, łączącego potrzeby dwóch programów (PHARE-CBC i Programu Sąsiedztwa). To z kolei przyniosło wykorzystanie produktów poza zakresem programu.

Kolejny czynnik determinujący użycie dotyczył przebiegu procesu ewaluacji (przede wszystkim kompleksowej metodologii – różnych poziomów analizy i mieszanki metod ilosciowych i jakościowych). Opracowanie wszechstronnych informacji na temat wpływu projektów przygranicznych, zaangażowanie interesariuszy w proces badawczy, a zwłaszcza ukierunkowanie na partnerów zaangażowanych w rozwój przyszłych programów, umożliwiły wykorzystanie zakończonego programu jako narzędzia uczenia się dla programów nadchodzących. Ewaluatorzy stanęli jednak wobec istotnych wyzwań, które wywarły wpływ na wykorzystanie raportu. Po pierwsze, rzetelne wnioski były ograniczone brakiem wiarygodnych danych bazowych i monitoringowych. Po drugie, proces ewaluacji pokazał, że partycypatywne, interaktywne stosunki z interesariuszami są trudne w zastosowaniu w programach UE. W tak kompleksowych, wielowarstwowych systemach nieuniknione są napięcia między licznymi interesariuszami. Trudno jest pogodzić ze sobą zróżnicowane narzędzia, interesy i perspektywy. Im więcej uczestników obejmuje ewaluacja i im więcej przewiduje konsultacji z interesariuszami, tym trudniej jest zarządzać procesem konsultacji.

6.3. Jakie siły i ograniczenia miało badanie?

Za główną siłę prezentowanego badania należy uznać jego całościowe, wielopoziomowe podejście do analizy, połączone z rozbudowaną, lecz spójną metodologią badawczą. Ewaluatorzy podjęli próbę badania relacji przyczynowo-skutkowej między szeroką interwencją (przypomnijmy: realizowaną w trzech regionach, w odniesieniu do 6 różnych sektorów

i przy użyciu różnej skali projektów) a jej efektami. Co więcej, korzystając z różnych inspiracji metodologicznych, rozwinęli własne podejście do analizy efektów programów współpracy ponadgranicznej. To pogłębione i naukowe podejście może być punktem wyjścia dla kolejnych badań ewaluacyjnych nowej generacji programów z pola współpracy terytorialnej.

Drugim silnym i wartym podkreślenia elementem badania było zastosowanie dodatkowego narzędzia transferu wiedzy ze starego, zamkniętego już programu do nowych inicjatyw w polu współpracy terytorialnej.

Głównym ograniczeniem badania były natomiast trudności w odtworzeniu sytuacji sprzed interwencji oraz oszacowywaniu zakładanych rezultatów i oddziaływania projektów (co wynikało przede wszystkim z luk w systemie monitoringu oraz zmian w statystyce). To ograniczenie zespół ewaluatorów częściowo minimalizował przez – po pierwsze, samodzielne uzupełnianie danych (bezpośrednio od beneficjentów), a po drugie – koncentrowanie się wyłącznie na ocenie rzeczywistych, a nie planowanych efektów. Brak danych poważnie ograniczał jednak rzeczowe oszacowanie racjonalności wyjściowych planów beneficjentów.

Kolejnym ograniczeniem badania było jego skoncentrowanie na stronie polskiej (co wynikało z natury badanej interwencji – przeznaczonej wyłącznie dla polskich beneficjentów). Ewaluatorzy zrealizowali punktowe badania po stronie niemieckiej, jednak aby prowadzone w porównaniu do strony polskiej – w bardzo ograniczonym zakresie.

Wreszcie ostatnim elementem, który ograniczył siłę oddziaływania badania, były zmiany instytucjonalne. Prezentacja wyników przypadła na okres zmian kompetencyjnych i tworzenia nowego systemu zarządzania (programami INTERREG). Te zmiany (w kontekście całej polityki), sprawiły, że to drugie już z kolei doświadczenie w ewaluacji PHARE-CBC miało ograniczony wpływ na budowanie wiedzy na poziomie instytucji centralnych.

Podsumowując, analizowany przykład pokazuje, że dla całościowej ewaluacji efektów kluczowe znaczenie ma przemyślana, dostosowana do specyfiki programu metodologia (wsparta czasem budżetem), która musi iść w parze dostępnością danych (przede wszystkim z poziomu monitoringu). Dla wykorzystania wniosków z takiej szerokiej ewaluacji ważna jest natomiast ciągłość instytucjonalna i przemyślana strategia zlecającego w wykorzystaniu wyników badania.

BIBLIOGRAFIA

EUROREG Warsaw University, PricewaterhouseCoppers, IRS (2006), *An ex post evaluation of the Phare Cross-Border Co-operation Programme Poland-Germany 1999–2001*, Warsaw, Implementing Authority for Phare-CBC Poland.

Gorzelak G., Bachtler J., Kasprzyk M. (2004), *Współpraca transgraniczna Unii Europejskiej: doświadczenia polsko-niemieckie*, Warszawa, Wydawnictwo Naukowe Scholar.

Gorzelak G., Kozak M., Olejniczak K., Smętkowski M. (2006), *Doświadczenia Programu Współpracy Przygranicznej PHARE Polska–Niemcy 1999–2001. Lekcje dla Wschodu*, Warszawa, Władza Wdrażająca Program Współpracy Przygranicznej PHARE.

Olejniczak K. (2008), *Rola Programu Współpracy Ponadgranicznej PHARE w rozwoju lokalnym województw zachodnich*, w: Gorzelak G. (red.), *Polska regionalna i lokalna w świetle badań EUROREG-u*, Warszawa, Wydawnictwo Naukowe Scholar.

ROZDZIAŁ 5.4.

Agnieszka Haber
Jacek Pokorski

EWALUACJA *EX POST* PROGRAMU PHARE 2003 SSG KOMPONENT MSP[1]

W poniższym rozdziale omówione zostaną projekt, realizacja oraz wykorzystanie wyników ewaluacji *ex post*, na przykładzie końcowej oceny programu Phare 2003 SSG[2] i realizowanego w ramach tej edycji programu, na obszarze 14 województw – regionalnego komponentu MSP (*2003/004-379-05.04 Regionalny program wsparcia małych i średnich przedsiębiorstw).*

Projekt ten zasługuje na uwagę przede wszystkim ze względu na zastosowane wówczas nowatorskie podejście do pomiaru efektów programu (badanie efektów netto), wykorzystanie metody PSM (*Propensity Score Matching*) oraz fakt, że projekt ten kumulował kilkuletnie

[1] Niniejszy rozdział został opracowany na podstawie doświadczeń i opracowań autorów artykułu, kierujących (po stronie Zamawiającego) procesem ewaluacji oraz na podstawie wyników badania ewaluacyjnego, dostępnych na stronach www.parp.gov.pl. Ponadto w rozdziale wykorzystano fragmenty: raportu metodologicznego z ewaluacji komponentu MSP Phare 2003 SSG, opracowanego na zlecenie PARP przez Pentor Research International; ekspertyzy prof. J. Górniaka (Uniwersytet Jagielloński) dla PARP, dotyczącej ewaluacji końcowych programów wsparcia sektora MSP; dokumentacji przetargowej PARP (szczególnie wybrane fragmenty Szczegółowego Opisu Przedmiotu Zamówienia, opracowanego przez Sekcję Ewaluacji).

[2] W krajach takich jak Polska – wstępujących do UE w 2004 roku i objętych programem Phare SSG – 2003 był ostatnim rokiem programowania Funduszu Phare, jednak harmonogram wdrażania ustalony w Memorandum Finansowym przewidywał realizację projektów Phare 2003 w niektórych przypadkach nawet do końca 2006 roku.

doświadczenia Jednostki Wdrażającej, w zakresie ewaluacji *ex post* programów skierowanych do sektora MSP.

Ze względu na powyższe uwarunkowania, główna uwaga zostanie skoncentrowana na zastosowanym przez Polską Agencję Rozwoju Przedsiębiorczości podejściu i doświadczeniach płynących z realizacji w ramach oceny *ex post* jednego z funduszy dotacji dla przedsiębiorców – Funduszu Dotacji Inwestycyjnych Phare 2003 SSG. Ponadto celem niniejszego rozdziału będzie zapoznanie czytelnika ze specyfiką badań ewaluacyjnych w obszarze programów wsparcia sektora małych i średnich przedsiębiorstw.

1. KONTEKST I PRZEDMIOT EWALUACJI

1.1. Krótko o Phare SSG

Program Phare[3] był instrumentem pomocy przedakcesyjnej Unii Europejskiej – kompleksowym programem wspierającym przygotowanie kraju do członkowstwa i absorpcji środków z funduszy strukturalnych UE w przyszłości. Jeden z komponentów programu zakładał wsparcie spójności społecznej i gospodarczej przez rozwój społeczno-gospodarczy regionów i podniesienie ich konkurencyjności. Phare Spójność Społeczna i Gospodarcza (SSG)[4] wpisywał się zatem w ramy polityki regionalnej UE, realizowanej w latach 2000–2006, zakładającej wyrównywanie różnic między regionami, w tym przypadku regionami krajów kandydujących oraz podwyższanie standardów życia w tych regionach. W ramach rocznych alokacji 2000–2003 Phare SSG współfinansowane były inwestycje (projekty) podobne do tych, które obecnie wspierane są w państwach członkowskich w ramach Europejskiego Funduszu Rozwoju Regionalnego i Europejskiego Funduszu Społecznego.

Środki Programu Phare SSG mogły być przeznaczone na:

- wzrost konkurencyjności sektora produkcyjnego[5] – **komponent rozwój małych i średnich przedsiębiorstw (MSP)**;
- wzmocnienie potencjału ludzkiego – **komponent rozwój zasobów ludzkich (RZL)**[6],

---

[3] Następca Programu PHARE (ang. *Poland and Hungary Assistance for Reconstructing of their Economies*) utworzonego w 1989 roku w celu wpierania transformacji społeczno-gospodarczej krajów Europy Środkowo-Wschodniej.

[4] ang. *Phare Economic and Social Cohesion* (*ESC*).

[5] Komponent MSP zakładał działania mające na celu pomoc w różnicowaniu gospodarki, rozwoju sektora prywatnego oraz restrukturyzację i modernizację przemysłu i usług.

[6] Komponent RZL wspierał wzrost zatrudnienia, rozwój przedsiębiorczości, poprawę zdolności adaptacyjnych firm i ich pracowników oraz wyrównywanie szans na rynku pracy.

- poprawę infrastruktury związanej z prowadzeniem działalności gospodarczej[7] – **komponent infrastruktura**[8].

Program realizowany był w ramach czterech edycji – 2000, 2001, 2002 oraz 2003 – o różnym zasięgu terytorialnym[9]. Był swoistym pilotażem zarówno dla administracji krajowej i regionalnej, jak i beneficjentów korzystających z tej pomocy. Oprócz wyrównywania dysproporcji w rozwoju społeczno-gospodarczym regionów, miał bowiem na celu również przekazanie beneficjentom praktycznych umiejętności, które w przyszłości miały ułatwić pozyskiwanie środków pochodzących z funduszy strukturalnych[10].

2. WKŁAD W BADANIE

2.1. Zakres ewaluacji i zastosowana metodologia

Ewaluacja końcowa obejmowała 4 regionalne fundusze dotacji dla przedsiębiorców Phare SSG: Fundusz Dotacji Inwestycyjnych (zwany dalej FDI) oraz trzy fundusze wsparcia na doradztwo – Program Rozwoju Przedsiębiorstw (PRP), Program Rozwoju Przedsiębiorstw Eksportowych (PRPE) oraz Technologie Informatyczne dla Przedsiębiorstw (TIdP) – które w latach 2005–2006 realizowała Polska Agencja Rozwoju Przedsiębiorczości, we współpracy z Regionalnymi Instytucjami Finansującymi (RIF)[11].

Główny cel ewaluacji został sformułowany stosunkowo ogólnie, jako uzyskanie pogłębionej wiedzy o efektach wybranych instrumentów

---

[7] Komponent Infrastruktura zakładał wsparcie projektów infrastrukturalnych, sprzyjających bezpośrednio wzrostowi aktywności sektora produkcyjnego i poprawie warunków prowadzenia działalności gospodarczej na szczeblu lokalnym i regionalnym.

[8] Por. Haber A., Witowski W. (red.) (2007), *Phare Spójność Społeczno-Gospodarcza. Podsumowanie Programu*, Warszawa, Ministerstwo Rozwoju Regionalnego i Polska Agencja Rozwoju Przedsiębiorczości.

[9] Liczba regionów objętych pomocą Phare SSG była zróżnicowana zarówno między poszczególnymi edycjami, strumieniami wsparcia (regionalny, sektorowy), jak i pomiędzy komponentami.

[10] Por. *Raport zamykający wdrażanie Programów Współpracy Przygranicznej, Sprawiedliwości i Spraw Wewnętrznych oraz Spójności Społeczno-Gospodarczej Phare realizowanych przez Władzę Wdrażającą Program Współpracy Przygranicznej Phare/ Władzę Wdrażającą Programy Europejskie*, Władza Wdrażająca Programu Europejskie, Warszawa 2007.

[11] Ewaluacja została przeprowadzona w okresie od września 2007 r. do lutego 2008 roku na zlecenie Polskiej Agencji Rozwoju Przedsiębiorczości przez Pentor Research International, przy współudziale ekspertów z zakresu ewaluacji oraz analizy sektora MSP. Zakres niniejszej ewaluacji oraz jej metodologia (pełny projekt ewaluacji) zostały opracowane przez Sekcję Ewaluacji PARP. Była ona elementem szerszego projektu obejmującego łącznie dwa programy wsparcia MSP Phare 2003 SSG: *2003/004-379-05.04 Regionalny program wsparcia małych i średnich przedsiębiorstw* i *2003/004-379-05.02 Wsparcie konkurencyjności polskich przedsiębiorstw – Inwestycje w MSP w Technologie i Innowacje*.

Tabela 1. Wybrane pytania ewaluacyjne, w podziale na kluczowe kryteria oceny

Kryterium ewaluacyjne	Pytania ewaluacyjne
Trafność	Czy cele ocenianego programu zostały sformułowane w sposób precyzyjny i wymierny? Czy zapewniały osiągalność, przy założonych metodach działania i zaangażowanych zasobach (finansowych, ludzkich, instytucjonalnych)? Czy kwantyfikacja celów, rezultatów i produktów w postaci wskaźników przyjętych w programie i zaplanowanych przy ich pomocy poziomów osiągnięć, oddawała istotę postawionych celów? Czy przyjęty zestaw wskaźników był kompletny i zrównoważony, a wskaźniki właściwie ujmowały założone produkty, rezultaty oraz oddziaływanie programu i miały jasną interpretację normatywną?
Skuteczność	Jakie były efekty programu u beneficjentów? Czy można powiedzieć, że zmiany zaobserwowane na poziomie badanych przedsiębiorstw mogą zostać przypisane podjętym interwencjom? Jakie czynniki zewnętrzne miały wpływ na te zmiany w przedsiębiorstwach? Jakie były efekty netto projektów zrealizowanych w ramach FDI Phare 2003 SSG u beneficjentów, w porównaniu z grupą nieobjętą wsparciem FDI Phare 2003 SSG i w porównaniu z efektami netto projektów zrealizowanych w ramach FDI Phare 2002 SSG? Które rozwiązania formalnoprawne i organizacyjne w poszczególnych funduszach dotacji dla MSP (Phare 2003 SSG) najbardziej sprzyjały efektywnemu osiągnięciu założonych celów? Czy mogą zostać uznane za dobre praktyki możliwe do upowszechnienia? Jakie rozwiązania można uznać za słabe strony programu? Dlaczego?
Efektywność	Jaki był średni koszt (wydatki programu i poszczególnych funduszy dotacji) uzyskania jednostkowego efektu brutto i netto (ustalonego w odniesieniu do próby kontrolnej wobec beneficjentów FDI), po stronie wzrostu nakładów inwestycyjnych MSP (o 1 PLN), wzrostu nakładów na IT (o 1 PLN), wzrostu przychodów (o 1 PLN), wzrostu zatrudnienia (o jedno miejsce pracy), przyrostu przychodu z eksportu (o 1 PLN), stopy wzrostu przychodów z eksportu (o 1%). Jak można zinterpretować uzyskane wyniki?
Użyteczność	Czy efekty projektów satysfakcjonowały beneficjentów (zwłaszcza w relacji do poniesionych kosztów i założonych celów projektów)? Czy zrealizowane działania odpowiedziały na ich ówczesne potrzeby? Jak beneficjenci, z perspektywy czasu, oceniali przydatność i jakość zrealizowanych projektów?
Trwałość	Czy w MSP objętych programem, w 12 miesięcy po jego zakończeniu, utrzymały się pozytywne efekty wsparcia, pod względem nakładów na inwestycje, nakładów na IT, eksportu, zatrudnienia, liczby miejsc pracy wyposażonych w technologie informatyczne, innowacji? Czy zmiany wynikające z realizacji projektów w przedsiębiorstwach (beneficjentach programu) możemy uznać za trwałe czy raczej tymczasowe? Jakie czynniki warunkują trwałość efektów we wspartych przedsiębiorstwach?

Źródło: opracowanie własne na podstawie *Szczegółowego Opisu Przedmiotu Zamówienia*.

wsparcia MSP, wdrażanych w ramach Phare 2003 oraz lepszego zrozumienia mechanizmów osiągania (bądź nieosiągania), przez te instrumenty, oczekiwanych efektów końcowych programu. Jednak cele bezpośrednie ostatecznie odwoływały się do wyczerpujących i stosunkowo szczegółowo sformułowanych pytań ewaluacyjnych, na które ewaluacja miała dostarczyć odpowiedzi. Pytania te były pogrupowane według pięciu klasycznych kryteriów ewaluacyjnych – trafności, skuteczności, efektywności, użyteczności i trwałości.

Była to ewaluacja o charakterze podsumowującym doświadczenia z wdrażania regionalnego komponentu MSP Phare 2003 SSG i poprzednich edycji programu. Wnioski z niej miały jednak służyć zarówno instytucjom centralnym (głównie PARP), jak i regionalnym (zarządy województw, Agencje Rozwoju Regionalnego, w tym Regionalne Instytucje Finansujące), przygotowującym instrumenty pomocowe, skierowane do sektora przedsiębiorstw w nowej perspektywie finansowej 2007–2013 w ramach Programu Operacyjnego Innowacyjna Gospodarka i 16 Regionalnych Programów Operacyjnych. Ewaluacja ta miała zatem pozwolić na zidentyfikowanie dobrych praktyk oraz słabych stron realizowanych działań. Na podstawie zgromadzonej wiedzy i zebranych doświadczeń uczestników, dokonana została niezależna ocena końcowa oraz wypracowane wnioski i zalecenia dla poprawy analogicznych instrumentów bezpośredniego (poprzez dotacje) wsparcia MSP.

Badanie ewaluacyjne miało charakter pełny i konkluzywny (było realizowane w **12 miesięcy po zakończeniu programu**), będąc jednocześnie logiczną kontynuacją ewaluacji końcowych poprzednich edycji,

Tabela 2. Zastosowane wskaźniki oraz podstawowe źródła pozyskania informacji

Wskaźniki	Źródło
wskaźniki bazowe i docelowe	dokumenty programowe (fiszki projektowe)
charakterystyki projektów i beneficjentów	bazy danych wnioskodawców i beneficjentów
wskaźniki monitoringowe (produkty i rezultaty)	okresowe raporty monitorujące; sprawozdania
wskaźniki ewaluacyjne (oddziaływanie)	badania prowadzone przez Jednostkę Wdrażającą oraz Narodowego Koordynatora Pomocy
wskaźniki makroekonomiczne	statystyka publiczna (GUS)
wskaźniki porównawcze (czasowe i przestrzenne)	badania sektorowe, bazy z badań pokrewnych, badania z wykorzystaniem grupy kontrolnej
wskaźniki subiektywne (kwestionariuszowe)	narzędzia badawcze zastosowane w ewaluacji

Źródło: opracowanie własne.

Tabela 3. Przegląd metod badawczych[12]

	Wywiady bezpośrednie wspomagane komputerowo (CAPI)	Wywiady telefoniczne (CATI)	Zogniskowane wywiady grupowe (FGI)	Indywidualne wywiady pogłębione (IDI)
Beneficjenci	395 beneficjentów FDI (948)	90 beneficjentów PRP (191) 44 beneficjentów PRPE (87) 92 beneficjentów TIdP (162)		po 5 wywiadów z beneficjentami każdego funfduszu
Nieskuteczni wnioskodawcy	322 nieskutecznych wnioskodawców FDI (3523)			
Wielokrotni beneficjenci				20 wywiadów
PARP				8 wywiadów
RIF			1 wywiad grupowy	10 wywiadów

Źródło: Prezentacja wyników ewaluacji Regionalnego komponentu MSP Phare 2003 SSG, przygotowana na zlecenie PARP przez Pentor Research International, Warszawa 2008.

[12] Liczby zamieszczone w tabeli oznaczają liczby zrealizowanych wywiadów w ramach określonych technik; liczby w nawiasach oznaczają łączną liczbę beneficjentów danego funduszu dotacji.

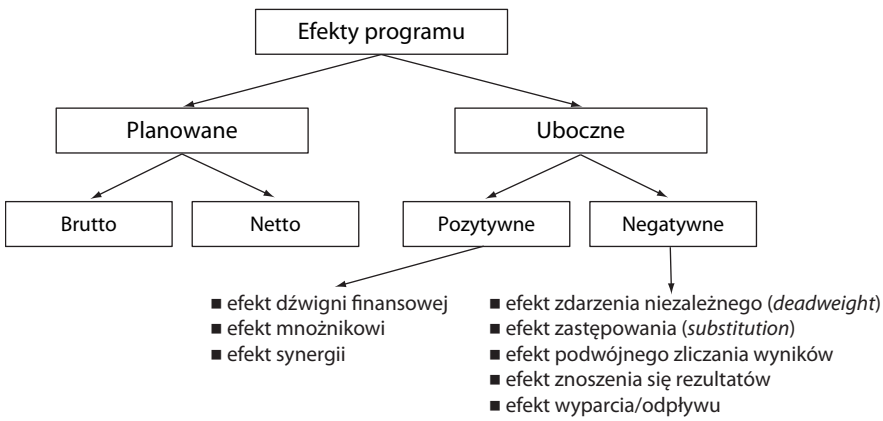
analogicznych regionalnych funduszy dotacji, wdrażanych w ramach Phare 2000–2002 SSG. W tabeli 2 wymienione są podstawowe wskaźniki oceny programu.

W myśl zasady triangulacji skorzystano zarówno z metod ilościowych – dostarczających „twardych", wiarygodnych i porównywalnych wyników – jak i z metod jakościowych – pozwalających na zrozumienie motywacji beneficjentów, mechanizmów zachodzących w programie i poznanie efektów niewymiernych.

Łącznie w ramach ewaluacji komponentu MSP Phare 2003 SSG przeprowadzono 943 wywiady kwestionariuszowe z przedstawicielami przedsiębiorstw, 58 indywidualnych wywiadów pogłębionych, zarówno z przedsiębiorcami, jak i pracownikami instytucji zaangażowanych we wdrażanie programu (PARP, RIF-y) oraz 1 zogniskowany wywiad grupowy z pracownikami RIF. Ponadto, poza wywiadami kwestionariuszowymi CATI i CAPI oraz wywiadami typu IDI i FGI, przeprowadzono badanie danych zastanych i analizę dostępnych dokumentów programu (*desk research*). Tabela 3 przedstawia dobór technik badawczych oraz przypisane im wielkości prób.

2.2. Założenia pomiaru efektów netto

Jednym z kluczowych założeń niniejszej ewaluacji było sprawdzenie, w jakim stopniu poczynione nakłady i zrealizowane działania przyczyniły się do osiągnięcia założonych celów programu. Dlatego też podjęto próbę zbudowania takiego projektu ewaluacji, który pozwoliłby na sprawdzenie związku przyczynowo-skutkowego pomiędzy nakładami i podjętymi działaniami programu a osiągniętymi efektami. Projekt oszacowania oddziaływania programu w odniesieniu do jego celów i wpływu nieplanowanego, został oparty na schemacie przedstawionym na rysunku 1.



Rysunek 1. Schemat oszacowania oddziaływania programu

Źródło: opracowanie własne.

Jednak próba ta, ze względu na dostęp do niezbędnych do tego celu danych, została zrealizowana jedynie w przypadku ewaluacji Funduszu Dotacji Inwestycyjnych[13] Phare 2003 SSG. Była ona również wynikiem doświadczeń z badanych efektów netto, zastosowanych po raz pierwszy w Polsce przez Polską Agencję Rozwoju Przedsiębiorczości, w ewaluacjach *ex post* Phare 2002 SSG. Badanie ewaluacyjne FDI 2003 miało zatem pozwolić zarówno na porównawczą ocenę skuteczności i efektywności netto funduszu z jego poprzednią edycją, jak i na zweryfikowanie uzyskanych wówczas wyników oraz zidentyfikowanych mechanizmów oddziaływania zrealizowanych projektów na grupę eksperymentalną (wspartą w ramach FDI) i kontrolną (bliźniaczo podobną grupę przedsiębiorstw, które nie uzyskały dotacji)[14].

Projekt zakładał, że ewaluacja powinna udzielić możliwie precyzyjną odpowiedź na pytanie o rzeczywiste efekty pomocy publicznej udzielonej przedsiębiorstwom. Jeśli efekt ten byłby pozytywny (netto), miała być to przesłanka do kontynuowania analogicznych form wsparcia przedsiębiorców. W przypadku braku efektów lub wystąpienia efektów negatywnych, były silne przesłanki do rekomendowania ograniczania tego typu wsparcia i wzięcia pod uwagę innych możliwości wykorzystania środków publicznych.

Skoncentrowano się przy tym głównie na efektach (wskaźnikach) w rozwoju przedsiębiorstw, jakie zaszły w firmach (beneficjentach programu) co najmniej rok od rozliczenia projektu. Rozwój przedsiębiorstwa, na potrzeby niniejszej ewaluacji, został operacyjnie ujęty na podstawie – między innymi – następujących wskaźników szczegółowych:

- wzrost wartości nakładów inwestycyjnych (ze środków niepochodzących z dotacji publicznych);
- wzrost liczby miejsc pracy we wspartym przedsiębiorstwie;
- wzrost wartości przychodów ze sprzedaży ogółem.

Opierając się na powyższych wskaźnikach, planowano ustalić, jak wygląda kondycja firmy po wykorzystaniu środków uzyskanych z dotacji na projekt inwestycyjny. Zakładano przy tym wyodrębnienie bezpośrednich efektów programu spośród wszystkich zaobserwowanych zmian w działalności przedsiębiorstwa w ciągu roku od uzyskania

---

[13] Fundusz Dotacji Inwestycyjnych zakładał bezzwrotną pomoc dla przedsiębiorców w formie dotacji, którzy złożyli poprawne formalnie i merytorycznie projekty inwestycyjne – przede wszystkim dotyczące rozbudowy parku maszynowego (zakupu i zainstalowania aparatury, oprzyrządowania, sprzętu komputerowego z oprogramowaniem czy specjalistycznych środków transportu) oraz wyposażenia firmy związanego z prowadzoną działalnością. Dotacje wynosiły od 30% do 50% wartości projektu inwestycyjnego i musiały zmieścić się w wyznaczonych przez fundusz limitach 2000–50 tys. EUR (Por. *Wytyczne dla wnioskodawców Funduszu Dotacji Inwestycyjnych*, http://archiwum.parp.gov.pl/dotacjedoc/phare2003/2003fdi/dotphare3fdi01.rtf_).

[14] Ze względu na nowatorskie podejście do pomiaru oddziaływania oraz zastosowaną metodologię, dalsza część niniejszego studium przypadku skoncentruje się właśnie na ocenie Funduszu Dotacji Inwestycyjnych.

wsparcia. Tak sformułowany zakres pomiaru oddziaływania programu musiał uwzględnić różnice parametrów ekonomicznych, osiągniętych we wspartych przedsiębiorstwach w 12 miesięcy po, z analogicznymi wskaźnikami zaobserwowanymi w grupie niepoddanej interwencji (grupie kontrolnej).

W związku z brakiem aktualnej i możliwie szerokiej informacji o firmach sektora MSP i ich aktywności, w ewaluacji *ex post* regionalnego Phare 2003 SSG (a zwłaszcza ewaluacji Funduszu Dotacji Inwestycyjnych) oparto się przede wszystkim na danych zgromadzonych przez Polską Agencję Rozwoju Przedsiębiorczości. Obejmowały one następujące grupy przedsiębiorstw, interesujące z punktu widzenia ewaluacji:

- mające rozliczone umowy o dofinansowanie w ramach Phare 2003 – **grupa zasadnicza**;
- bezskutecznie aplikujące o wsparcie z Phare 2003 – **grupa kontrolna**.

Za wykorzystaniem jako przypadków kontrolnych firm, które zgłosiły się do udziału w FDI, ale nie zostały pomyślnie zakwalifikowane, przemawiał bardzo istotny czynnik: motywacja do udziału w funduszu. Firmy, które otrzymały dotację i te, które jej nie otrzymały, były podobne pod względem wstępnego zainteresowania poszukiwaniem środków publicznych na rozwój i aktywności w przygotowaniu aplikacji. Ten czynnik predestynował je zatem do występowania w roli przypadków kontrolnych[15]. Zakładano, że podobieństwo pod względem innych cech będzie kontrolowane dzięki przyjętemu schematowi doboru próby na podstawie techniki *Propensity Score Matching* (PSM)[16].

3. PROCES EWALUACJI

Ujmując kwestie realizacji z punktu widzenia uczestnictwa Zamawiającego w niniejszej ewaluacji na osiach responsywne – partycypatywne, należy przyjąć, iż miała ona charakter zróżnicowany na poszczególnych etapach – silnie partycypatywny na etapie konceptualizacji i operacjonalizacji, oraz responsywny na etapie realizacji badania ewaluacyjnego i raportowania (etap analizy i oceny). Centrum badania, wokół którego

[15] Por. Górniak J., *Efekt netto i PSM w ewaluacji* ex post, prezentacja przedstawiona 9 maja 2007 roku na spotkaniu w Polskiej Agencji Rozwoju Przedsiębiorczości, poświęconym wynikom ewaluacji *ex post* komponentu MSP Phare 2002 SSG.

[16] W celu oszacowania stopnia podobieństwa badanych przedsiębiorstw (a tym samym znalezienia właściwych do porównań odpowiedników) w metodzie PSM wykorzystuje się Indeks Skłonności (*Propensity Score*). Indeks ten ma charakter syntetyczny i niesie informację o prawdopodobieństwie przyznania przedsiębiorstwu dofinansowania. Jego wartość dla indywidualnych firm została obliczona za pomocą modelu regresji logistycznej, gdzie zmienną zależną był fakt przyznania (bądź nie) dofinansowania, a zmiennymi niezależnymi – określone cechy przedsiębiorstwa.

budowano koncepcję pomiaru i analiz, było podejście zorientowane na cele programu oraz metodę *Propensity Score Matching* (PSM)[17].

3.1. Realizacja ewaluacji

Cechami kontrolowanymi w modelu PSM, na podstawie bazy danych PARP, były m.in.: fundusze dotacji, z których skorzystały przedsiębiorstwa przed aplikowaniem do FDI, siedziba wojewódzka firmy, zatrudnienie, obroty, aktywa, sektor działalności (PKD), rok rozpoczęcia działalności oraz cechy dodatkowe, zidentyfikowane w trakcie procesu ewaluacji[18].

Wykonawca, w ramach prac badawczych, miał za zadanie oszacować efekty netto, porównując obie badane grupy przedsiębiorców, związane z Funduszem Dotacji Inwestycyjnych. Dla ich oszacowania porównywał wskaźniki konkurencyjności (określone w pytaniach badawczych) w grupie objętej wsparciem i grupie kontrolnej, złożonej z przedsiębiorstw nieskutecznie ubiegających się o wsparcie, o cechach najbardziej zbliżonych do cech beneficjentów w chwili realizacji projektu. Dzięki temu uzyskał możliwie najbliższą stanowi faktycznemu odpowiedź na pytanie o rzeczywisty efekt wydatkowanych środków.

Dzięki stosunkowo dużej liczebności beneficjentów FDI Phare 2003 SSG, w ewaluacji można było również dokonać weryfikacji wpływu czynnika regionalnego. Jednak liczebności beneficjentów nie były na tyle duże, aby można było dobrać odpowiednią próbę w ramach każdego regionu. Dlatego 14 województw, na terenie których były realizowane projekty FDI, połączono w pary o najbardziej podobnym poziomie PKB/os. i liczbie podmiotów gosp./100 tys. ludności[19]. Następnie do populacji beneficjentów FDI, z wykorzystaniem techniki *Propensity Score Matching*[20] dobrano „statystycznych bliźniaków" (przypadki kontrolne) spośród przedsiębiorców, którzy ostatecznie nie uzyskali wsparcia

---

[17] Por. Konarski, Kotnarowski 2007; Strawiński 2008.

[18] W modelu PSM uwzględniono ponad **70 różnych cech wnioskodawców** (skutecznych i nieskutecznych) i ich projektów. Ponadto zdecydowano, że w badaniu będą mogły wziąć udział wszystkie firmy, niezależnie od tego, czy korzystały (bądź obecnie korzystają) z jednej, czy z kilku dotacji – nałożenie warunku o korzystaniu wyłącznie z FDI działałoby wysoce ograniczająco na liczebność próby. Jednak w przypadku firm, które już wcześniej korzystały, bądź starały się o przyznanie dotacji, prawdopodobieństwo jej otrzymania wzrasta, po części ze względu na zebrane doświadczenia, po części dlatego, że procedura przyznawania dotacji przewidywała preferencje dla tych firm, które aplikowały o fundusze już wcześniej. Dlatego też, aby uwzględnić ten efekt, do modelu włączone zostały również zmienne informujące o korzystaniu z dotacji innych programów – liczby złożonych wniosków oraz liczby rozliczonych umów przed złożeniem wniosku do FDI Phare 2003, dla każdego programu osobno.

[19] Wskaźnik podobieństwa województw był średnią arytmetyczną zmiennych PKB/os. oraz liczba podmiotów gosp./100 tys. ludności w danym regionie.

[20] Opis tej metodologii znaleźć można m.in. w publikacji: Bryson, http.

Tabela 4. Pary województw wraz z liczbą obserwacji z bazy danych

Województwo		Liczba obserwacji w każdym z województw		SUMA
1)	2)	1)	2)	1)+2)
mazowieckie	śląskie	288	568	856
wielkopolskie	dolnośląskie	420	289	709
pomorskie	zachodniopomorskie	256	285	541
łódzkie	kujawsko-pomorskie	190	226	416
małopolskie	warmińsko-mazurskie	481	253	734
świętokrzyskie	podlaskie	122	209	331
lubelskie	podkarpackie	373	475	848

Źródło: Raport metodologiczny z ewaluacji komponentu Phare 2003 SSG, opracowany przez Pentor Research International na zlecenie PARP, Warszawa 2008.

(nieskutecznych wnioskodawców FDI). Na podstawie wyliczonego indeksu, dokonano połączenia województw w pary (tabela 4).

Ostateczna baza, która posłużyła do wylosowania próby zasadniczej i kontrolnej, składała się z 906 beneficjentów FDI oraz 3529 nieskutecznych wnioskodawców FDI Phare 2003 SSG. Do każdego respondenta z populacji beneficjentów FDI, przy użyciu techniki PSM dobrany został odpowiadający mu „statystyczny bliźniak" z próby kontrolnej (nieskuteczny wnioskodawca)[21].

Wykorzystując dostępne cechy przedsiębiorstw, oszacowano model regresji logistycznej. Obliczony na podstawie przewidywanego przez model prawdopodobieństwa sukcesu, Indeks Skłonności (*Propensity Score*) posłużył następnie do wyłonienia par składających się z firmy-beneficjenta oraz firmy z grupy kontrolnej. Łączenie par przebiegało na podstawie metody najbliższego sąsiedztwa z zadeklarowanym minimalnym dystansem (*Nearest Neighbour Caliper Matching*). Oznacza to, że każdemu dofinansowanemu przedsiębiorstwu przyporządkowana została taka firma niedofinansowana, która charakteryzowała się najwyższą zbliżoną wartością Indeksu Skłonności. Procedura łączenia w pary

[21] Wstępnie zakładano, że każda firma z próby kontrolnej zostanie wykorzystana jako odpowiednik tylko jednej firmy z próby zasadniczej. Taki wymóg podyktowany był chęcią zapewnienia niezależności obserwacji, które znajdą się w próbie (ze względu na to, że większość statystyk stosowanych na późniejszym etapie prac zakłada niezależność obserwacji w próbie). Niemniej jednak, dopuszczano możliwość, że niewielki odsetek firm z bazy kontrolnej będzie wykorzystany dwukrotnie, jeśli okaże się, że po oszacowaniu wartości indeksu podobieństwa, nie można jednoznacznie dobrać wystarczającej liczby par firm.

przeprowadzona została w oparciu o jeden indeks dla wszystkich regionów ekonomicznych (w oparciu o jeden model regresji logistycznej), co miało swoje uzasadnienie merytoryczne.

Biorąc pod uwagę spodziewany wysoki poziom odmów w przypadku próby kontrolnej, na etapie losowania, od razu, oprócz „statystycznego bliźniaka", wylosowano 3 respondentów zastępczych, maksymalnie zbliżonych pod względem wartości wskaźnika PSM do „statystycznego bliźniaka"[22].

Po przeprowadzeniu dobierania par, beneficjenci, dla których nie udało się dobrać pary, zostali usunięci z próby, a spośród pozostałych, już dobranych par, została wylosowana właściwa próba, dla której zostały przeprowadzone wywiady. Struktura baz przed „parowaniem" oraz po usunięciu beneficjentów, dla których nie udało się dobrać „statystycznych bliźniaków", została porównana, a w przypadku dużych rozbieżności przeprowadzono czynności korygujące (procedura ważenia uzyskanych wyników).

W przypadku ewaluacji Funduszu Dotacji Inwestycyjnych zakładano liczebność próby na poziomie 770. Ogółem wybranych miało zostać 385 par „statystycznych bliźniaków":

- 385 wywiadów w próbie zasadniczej (7 par województw × 55 respondentów)
- 385 wywiadów w próbie kontrolnej (7 × 55 respondentów).

W efekcie przeprowadzonego łączenia województw, w każdej warstwie wyodrębnionego regionu ekonomicznego (para województw) uzyskano wystarczającą liczbę beneficjentów do zrealizowania 55 wywiadów.

Po zakończeniu badania proces dopasowania „statystycznych bliźniaków" został przeprowadzony powtórnie, już w ramach zrealizowanej próby, w celu zapewnienia maksymalnego podobieństwa dobranych par, biorąc pod uwagę wartość indeksu podobieństwa. Działanie takie było niezbędne, ze względu na fakt, że próba kontrolna została w części zrealizowana z „zastępcami", a nie z pierwotnie wylosowanymi odpowiednikami firm, co mogło spowodować, że pierwotnie optymalne pary w próbie zakładanej, nie będą takie w próbie zrealizowanej. Dlatego przeprowadzono procedurę sprawdzającą, czy – już w ramach zrealizowanych wywiadów – nie jest możliwe znalezienie lepiej dobranych par firm beneficjentów i nieskutecznych wnioskodawców. Przebieg procedury był taki sam jak w przypadku pierwotnego parowania.

Aby realizacja badania przebiegała jak najbardziej optymalnie (wysoki stopień realizowalności próby zasadniczej oraz jak najwyższy stopień realizowalności „statystycznych bliźniaków"), zastosowano następujący schemat postępowania:

---

[22] Miało to przede wszystkim usprawnić proces realizacji – ankieter w przypadku uzyskania odmowy niezwłocznie udawał się do kolejnego, wylosowanego wcześniej respondenta z grupy kontrolnej.

- Do realizacji skierowano dobrane komplety respondentów (pary przedsiębiorstw).
- Komplet składał się z jednego wywiadu w ramach próby zasadniczej, odpowiadającego mu bliźniaka z próby kontrolnej oraz najbliższego respondenta „zastępczego".
- Ankieter rozpoczynał realizację pary od wywiadu z próby zasadniczej – w przypadku absolutnej odmowy na bieżąco (zaraz po przekazaniu takiej informacji) otrzymywał z działu realizacji badań kolejny komplet.
- Gdy realizacja wywiadu z respondentem z próby zasadniczej powiodła się, ankieter przystępował do realizacji wywiadu w ramach próby kontrolnej, zaczynając od statystycznego bliźniaka, a gdy dotarcie do niego było niemożliwe, realizował wywiad z jego „zastępcą".

Dzięki zastosowaniu metody CAPI (*Computer Assisted Personal Interviewing*) realizacja próby była kontrolowana na bieżąco, na poziomie każdej z par województw. Zastosowanie droższej metody realizacji badania umożliwiło nie tylko utrzymanie prawidłowego scenariusza wywiadu, weryfikację poprawności wprowadzanych danych, gwarancje logicznego układ kwestionariusza (zgodnie ze specyfiką kolejnych odpowiedzi respondenta), lecz przede wszystkim umożliwiło skrócenie procesu przenoszenia danych, szybszą realizację badania i lepszą możliwość zarządzania logistyką całego procesu realizacji.

Szacowanie wpływu dofinansowania na poszczególne podmioty odbywało się przez obliczenie różnic w zmiennych obliczanych na podstawie odpowiedzi w ankiecie, charakteryzujących cechy operacyjne pomiędzy przedsiębiorstwami, którym przyznano dofinansowaniem oraz przedsiębiorstwami, którym dofinansowania nie przyznano. W celu oszacowania tych różnic, podmioty, którym zostało przyznane dofinansowanie, zostały porównane z grupą maksymalnie podobnych im podmiotów, przyporządkowanym wcześniej metodą PSM na podstawie wartości Indeksu Skłonności (*Propensity Score*). Przy testowaniu hipotezy o dodatnim wpływie dotacji na wyżej wymienione wskaźniki, zastosowane zostały jednostronne testy istotności różnicy średnich. Testy te pozwalały na potwierdzenie lub odrzucenie hipotezy o dodatnim wpływie (efekcie netto) uzyskanych dotacji na rozwój przedsiębiorstw ze względu na mierzone wskaźniki ekonomiczne.

4. PRODUKT I WYKORZYSTANIE EWALUACJI

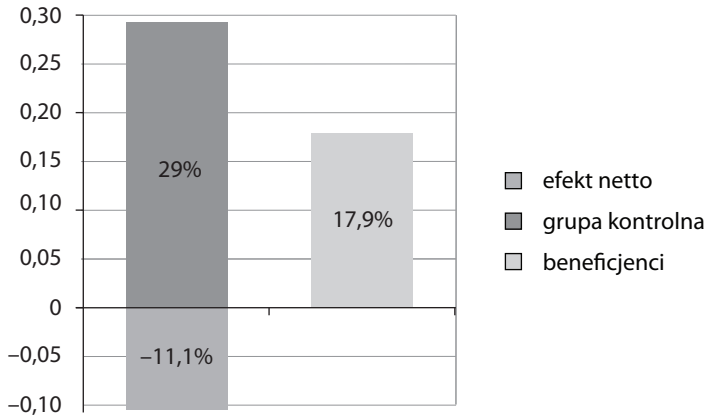
Otrzymany produkt końcowy – raport zbadania – był kumulacją doświadczeń z realizacji ewaluacji w badanym obszarze tematycznym, zarówno Zamawiającego (komórki ewaluacyjnej PARP), jak i Wykonawcy (zespołu ewaluatorów), który realizował dla PARP badanie ewaluacyjne dwóch kolejnych edycji komponentu MSP Phare SSG (2002, 2003).

Wiedzę, jaką uzyskano dzięki badaniu, można określić, w wymiarze oceny programu, głównie jako „spodziewaną" o charakterze diagnostycznym. Natomiast w wymiarze realizacji projektu, zastosowanej metodologii oraz konstrukcji przyszłych projektów badawczych jako techniczno-koncepcyjno-preskryptywną, dostarczającą konkretnych rozwiązań praktycznych.

4.1. Wiedza w wymiarze oceny programu – najważniejsze wyniki (efekty netto)

W wyniku przeprowadzonych badań i analiz zaobserwowano następujące wyniki:

- Średnie przyrosty zatrudnienia w etatach: różnica wskaźnika pomiędzy grupą beneficjentów i grupą kontrolną wyniosła zaledwie 1,6 etatu i nie była istotna statystycznie (jedynie w regionie łódzkie/kujawsko-pomorskie efekt był istotny statystycznie i wyniósł przyrost 17,4 etatu).
- Średnie przyrosty zatrudnienia w procentach: różnica wskaźnika pomiędzy grupą zasadniczą i kontrolną wyniosła 25% i była na granicy istotności statystycznej (efekt istotny statystycznie w regionach: pomorskie/zachodniopomorskie i łódzkie/kujawsko-pomorskie).
- Średni wzrost sprzedaży ogółem: różnica nie była istotna statystycznie, tj. beneficjenci i nieskuteczni wnioskodawcy notowali podobne wzrosty sprzedaży na poziomie kilkunastu procent (nie występowały w tym zakresie zróżnicowania regionalne).
- Wzrost nakładów inwestycyjnych: różnica pomiędzy grupą zasadniczą i kontrolną wyniosła –11% (minus jedenaście procent) i była



Rysunek 2. Zmiana nakładów inwestycyjnych – efekty netto

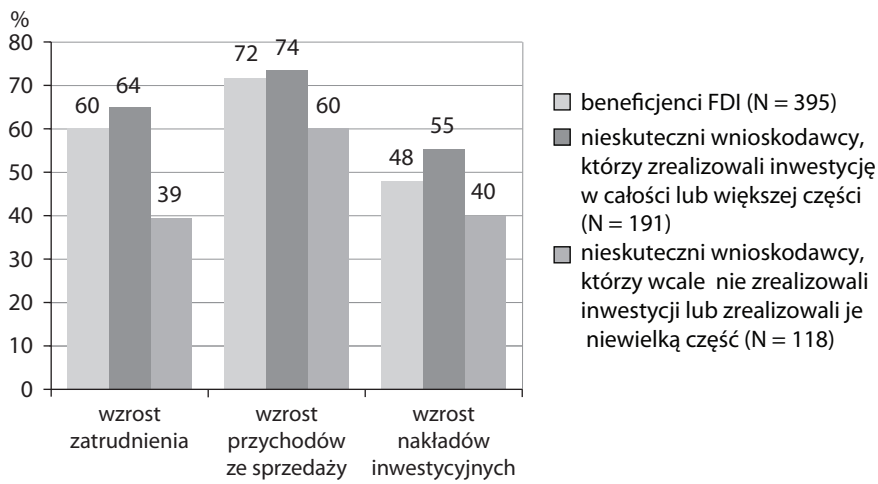
Źródło: Prezentacja wyników przygotowana dla PARP przez Pentor Research International, Warszawa 2008; http://www.parp.gov.pl/index/index/110

istotna statystycznie. Oznacza to, że nieskuteczni wnioskodawcy notowali w tym zakresie lepsze efekty niż beneficjenci FDI[23].

Warto jednak w przypadku tego wskaźnika podkreślić kilka istotnych czynników:

- zaspokojenie potrzeb inwestycyjnych beneficjentów po realizacji projektów;
- zadłużenie firm beneficjentów w wyniku realizacji projektów (blisko 1/3 wspartych firm, finansowała inwestycję z kredytów), co ograniczało aktywność inwestycyjną w ciągu roku po uzyskaniu wsparcia;
- większe rozłożenie w czasie procesu inwestycyjnego nieskutecznych wnioskodawców.

Ponadto zaobserwowano, że większość nieskutecznych wnioskodawców (59%) zrealizowała inwestycję, którą zamierzano sfinansować z FDI (wszystkie zaplanowane działania – 33%; większość działań – 26%). Pozwala to sądzić, że **predykatorem pozytywnych zmian w przedsiębiorstwie była realizacja samej inwestycji, a nie otrzymana dotacja**. Tezę tę w pewnym sensie potwierdza analiza obu grup, przy uwzględnieniu dodatkowego rozbicia grupy kontrolnej na przedsiębiorstwa, które zrealizowały inwestycję od tych, które jej nie zrealizowały. W przypadku trzech wybranych parametrów ekonomicznych okazało się, że tam, gdzie była realizowana inwestycja, wskaźniki osiągnięć były porównywalne (niezależnie od uzyskania bądź nieuzyskania dotacji).



Rysunek 3. Zmiany w przedsiębiorstwach (wspartych i nieskutecznie wnioskujących)

Źródło: Prezentacja wyników przygotowana dla PARP przez Pentor Research International, Warszawa 2008. http://www.parp.gov.pl/index/index/110

[23] Ze względu na to, że w większości badanych wskaźników, efekty netto nie były istotne statystycznie (lub były ujemne), zrezygnowano z szacowania wskaźników efektywności netto i poprzestano w tym zakresie na obliczeniach brutto.

Porównanie kondycji beneficjentów i nieskutecznych wnioskodawców pokazało, iż otrzymanie dotacji stało się zbyt słabym czynnikiem różnicującym kondycję ekonomiczną tych firm. Obie grupy firm były podobnie dobrze przygotowane do finansowania i realizowania zamierzonych inwestycji – reprezentowały postawę rozwojową i ekspansywną, miały dobrą kondycje finansową i wyrażały ogólną chęć inwestowania (przez aplikowanie o wsparcie). Ponadto, obie grupy poniosły znaczące nakłady finansowe i podjęły prace przygotowawcze poprzedzające realizację inwestycji – dokonały analizy jej opłacalności, przygotowały biznesplan na potrzeby udziału w programie oraz zapewniły wkład własny na prefinansowanie jej w całości.

Ocena efektów netto FDI Phare 2003 zidentyfikowała zatem „efekt uboczny" programu – stymulację procesu inwestycyjnego w całej grupie wnioskodawców – zarówno tych wspartych dotacją FDI, jak i nieskutecznie o nią aplikujących.

5. WIEDZA W WYMIARZE REALIZACJI PROJEKTU, ZASTOSOWANEJ METODOLOGII BADANIA

5.1. Ograniczenia procesu ewaluacji

Realizacja badania ilościowego z wykorzystaniem próby nieskutecznych wnioskodawców do oszacowania efektów netto programu, ze względu na liczebności przedsiębiorców uczestniczących w programie, możliwa była praktycznie tylko w przypadku Funduszu Dotacji Inwestycyjnych (ograniczenia założeń techniki PSM). W przypadku pozostałych funduszy regionalnego Phare 2003 SSG zrealizowano badania kwestionariuszowe tylko na próbie beneficjentów tych funduszy, co nie pozwala w sposób jednoznaczny ocenić wpływu interwencji na zaobserwowane wśród beneficjentów zmiany[24].

Wydaje się, że lepszym rozwiązaniem metodologicznym mogłoby być dodatkowe zbadanie zmian zachodzących pomiędzy grupą nieskutecznych wnioskodawców a grupą firm w ogóle nieaplikujących do FDI (reprezentantami całego sektora MSP). Problemem w przyjęciu takiego podejścia jest jednak brak odpowiednio szczegółowych baz firm niewnioskujących. Utworzenie takich baz byłoby z pewnością możliwe w wyniku pogłębionych badań sektora MSP, jednakże koszty, które należałoby w tym celu ponieść, byłyby zbyt duże w stosunku do możliwych do uzyskania efektów, przekraczając granice ekonomicznej racjonalności.

[24] Podejście to nie pozwala na wyizolowanie wpływu interwencji od wpływu innych potencjalnie oddziaływujących czynników, np. zmian w makrootoczeniu przedsiębiorstw, ogólnych tendencji w sektorze MSP, preselekcji do programu przedsiębiorstw gwarantujących sukces.

Warto podkreślić, że zaprojektowany schemat przeprowadzania wywiadów powodował bardzo duże trudności realizacyjne dla instytutu badawczego. W tym przypadku efektywnie przeprowadzona obserwacja obejmowała nie jeden wywiad z przedsiębiorcą, a realizację 2 wywiadów w ramach dobranej pary: beneficjent + nieskuteczny wnioskodawca FDI. Uniemożliwiało to równoległe uruchomienie w terenie obu prób. Badania musiały być prowadzone sekwencyjnie: najpierw wywiad z beneficjentem, później z jego bliźniakiem. Wymagana ścieżka realizacji, powodująca istotne wydłużenie etapu zbierania danych, musiała zostać uwzględniona w harmonogramie realizacji projektu. Dużą korzyść w tym zakresie przyniosła metoda CAPI, która pozwalała na bieżące monitorowanie zrealizowanych wywiadów z beneficjentami i uruchamiania w dziale realizacji odpowiednich przypadków kontrolnych.

Dodatkowym ograniczeniem był wysoki poziom odmów w zakresie danych dotyczących wskaźników finansowych, który ograniczał możliwości analityczne związane z obliczaniem wskaźników efektywności[25]. Dzięki zastosowaniu formularza z danymi metryczkowymi, przesyłanego respondentom w momencie umawiania wywiadu (jeszcze przed wizytą ankietera), oraz dzięki odpowiednim zapisom w umowach o dofinansowanie (zgoda beneficjentów na udział w badaniach ewaluacyjnych dotyczących programu), udało się częściowo ograniczyć negatywny wpływ tego utrudnienia.

5.2. Wykorzystanie wyników

W celu dotarcia do jak największej liczby odbiorców wyniki omawianej ewaluacji prezentowane były na kilku etapach. Pierwsze dwa etapy miały charakter konsultacyjny oraz służyły wypracowaniu jak najlepszych rekomendacji. Omówienie wyników miało charakter spotkań, w których brali udział ewaluatorzy, pracownicy szczebla operacyjnego, następnie szczebla decyzyjnego zaangażowanych instytucji[26]. Kolejne etapy to zaprezentowanie raportu szerokiemu gronu osób w środowisku wdrażania i zarządzania programem Phare SSG, w środowisku akademickim oraz środowiskach odpowiedzialnych za projektowanie, zarządzanie oraz wdrażanie interwencji o podobnym charakterze. Ostatni etap to upublicznienie najważniejszych wyników w postaci prezentacji na stronach PARP.

Kwestia wykorzystania wyników pozostaje w przypadku tej konkretnej ewaluacji zagadnieniem złożonym. Z jednej strony, jej wyniki, ze

---

[25] Zwłaszcza w przypadku średnich kosztów jednostkowego efektu brutto po stronie wzrostu nakładów inwestycyjnych i przychodów ze sprzedaży, gdzie odsetek odmów wahał się w przedziale od 26% do 43% w zależności od funduszu dotacji

[26] Odpowiedni dyrektorzy departamentów Ministerstwa Rozwoju Regionalnego, dyrektorzy bezpośrednio odpowiedzialni za wdrażanie oraz prezesi PARP.

względu na czas umiejscowienia w cyklu programowania w ramach kolejnych perspektyw finansowych – na lata 2004–2006 oraz 2007–2013 – nie mogły być uwzględnione w procesach konstrukcji nowych programów z drugiej zaś, w kontekście wyników przeprowadzonych badań ewaluacyjnych – FDI Phare 2002 SSG i FDI Phare 2003 SSG – w szczególności zidentyfikowania efektów netto programu bliskich zeru, zainicjowano debatę na temat faktycznej zasadności bezpośredniego (w formie bezzwrotnych dotacji) wspierania przedsiębiorstw oraz dyskusję na temat oddziaływania programu poza grupę bezpośrednio poddaną interwencji środków publicznych (np. przedsiębiorstwa nieskutecznie aplikujące o wsparcie, zyskiwały pozytywne efekty uboczne w pewnym sensie dzięki programowi). Stąd, choć funkcja ewaluacji w wymiarze poprawy wdrażania programów jest mocno ograniczona z przyczyn obiektywnych, to w wymiarze rozliczeniowym, poznawczym oraz wzmacniania odpowiedzialności, spełnia oczekiwania, jakie pierwotnie przed nią postawiono.

Pierwsze w Polsce badania efektów netto programów zaprojektowane przez Polską Agencję Rozwoju Przedsiębiorczości, w sposób bardzo silny wpłynęły na podejmowanie prób badań efektów netto w ramach ewaluacji przeprowadzanych przez inne instytucje. Odseparowane efekty programu spośród wszystkich zaistniałych zmian we wspartych podmiotach w określonym czasie, stały się dla decydentów i partnerów społeczno-gospodarczych bardzo interesujące, w przeciwieństwie do dotychczasowych wyników, opartych na pomiarze wskaźników brutto, polegających na prostym rejestrowaniu zmiany w czasie w obszarze docelowym. Wpłynęło to tym samym na podniesienie poprzeczki dla badań ewaluacyjnych, wywołało potrzebę stawiania pytań o relacje przyczynowo-skutkowe pomiędzy nakładami i podjętymi działaniami oraz wynikami. Niewątpliwie może się to tylko przyczynić do poprawy jakości projektowanych i obecnie realizowanych polityk publicznych.

Równie istotne oddziaływanie niniejszej ewaluacji oraz wykorzystanie jej wyników można dostrzec w obszarze doskonalenia podejścia stosowanego przez PARP w badaniach sektora przedsiębiorstw. Niespełna pół roku od zakończenia projektu, można zaobserwować bardzo wyraźną ewolucję podejścia i metod stosowanych przy kolejnych ewaluacjach programów wsparcia skierowanych do sektora MSP[27]. Zmiany te przyniosły wyraźne efekty w zakresie wysokiej jakości pozyskiwanych danych (zwiększyły się możliwości pozyskiwania wiarygodnych danych finansowych przedsiębiorstw, niezbędnych przy obliczaniu jakichkolwiek efektów ekonomicznych), budowy systemu oceny na lata 2008–2015, uwzględniającego szerszą perspektywę funkcjonowania programów publicznych oraz podniesienia efektywności projektów ewaluacyjnych.

---

[27] Ewaluacje dotyczące *Sektorowego Programu Operacyjnego Wzrost Konkurencyjności Przedsiębiorstw, lata 2004–2006* oraz *Programu Operacyjnego Innowacyjna Gospodarka, na lata 2007–2013*.

6. WNIOSKI ZE STUDIUM PRZYPADKU

Gdybyśmy – jako autorzy projektu oraz osoby zarządzające ewaluacją – mieli wskazać w kilku zdaniach najważniejsze wnioski, jakie rysują się nam po realizacji powyższego projektu, byłyby one następujące:

- przy tak skomplikowanym projekcie – technicznie, analitycznie oraz testującym nowe metody badawcze – należy oszacować czas realizacji z uwzględnieniem „bufora czasowego";
- zastosowany dobór grupy kontrolnej (wyłącznie spośród nieskutecznych wnioskodawców), mimo że istotnie przybliżył nas do faktycznych efektów programu, okazał się niewystarczający; dostęp oraz zakres gromadzonych danych (w tym jakość baz danych GUS) jest jedną z kluczowych kwestii, istotnych z punktu widzenia przyszłych ewaluacji;
- wykorzystanie wyników jest ściśle skorelowane z takimi czynnikami jak kultura zorientowana na zarządzanie wiedzą oraz kapitał intelektualny i społeczny danej instytucji, stąd w procesie ewaluacji należy uwzględniać również te czynniki, jako zwiększające absorpcję wiedzy.

BIBLIOGRAFIA

Bryson A. et al., *The Use of Propensity Score Matching in The Evaluation of Active Labour Market Policies" pod adresem internetowym*; http://www.dwp.gov.uk/asd/asd5/WP4.pdf

Haber A., Witowski W. (red.) (2007), *Phare Spójność Społeczno-Gospodarcza. Podsumowanie Programu*, Warszawa, Ministerstwo Rozwoju Regionalnego i Polska Agencja Rozwoju Przedsiębiorczości.

Konarski R., Kotnarowski M. (2007), *Zastosowanie metody propensity score matching w ewaluacji ex-post*, w: Haber A. (2007), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Warszawa, PARP, http://www.parp.gov.pl/files/74/81/158/Ewaluacja_ex-post.Teoria_i_praktyka_badawcza.pdf

New Deal for Lone Parents Evaluation: Findings from the Quantitative Survey, www.dwp.gov.uk/jad/2003/wae147rep.pdf

Strawiński P. (2008), *Quasi-eksperymentalne metody ewaluacji*, w: Haber A., Szałaj M. (red.) *Środowisko i warsztat ewaluacji*, Warszawa, PARP.

Wytyczne dla wnioskodawców Funduszu Dotacji Inwestycyjnych

http://archiwum.parp.gov.pl/dotacjedoc/phare2003/2003fdi/dotphare3fdi01.rtf

ROZDZIAŁ 5.5.

Grzegorz Gorzelak
Adam Płoszaj

EWALUACJA Z DŁUGĄ PERSPEKTYWĄ – PRZYKŁAD PROGRAMU INICJATYW LOKALNYCH[1]

1. KONTEKST I PRZEDMIOT EWALUACJI

Celem tego rozdziału jest zaprezentowanie przebiegu i wyników badania, które można rozpatrywać jako ewaluację z długą perspektywą czasową. Omawiane badanie nie było ewaluacją typową, choć niewątpliwie ma cechy, które pozwalają je uznać za ewaluację typu *ex post*. Impulsem

[1] Niniejszy tekst wykorzystuje fragmenty artykułu *Zewnętrzna interwencja jako czynnik rozwoju lokalnego (na przykładzie Programu Inicjatyw Lokalnych)* autorstwa Grzegorza Gorzelaka, który ukazał się w roku 2000 w kwartalniku „Studia Regionalne i Lokalne", nr 3(3)2000 (artykuł dostępny jest pod adresem: www.studreg.uw.edu.pl/pdf/SRIL03/Art07.pdf). Badania były finansowane ze środków indywidualnego grantu KBN pn. *Regionalne i lokalne aspekty polityk sektorowych*. W badaniach gmin uczestniczyli: Grzegorz Gorzelak, Bohdan Jałowiecki, Wojciech Roszkowski, Anna Tucholska. Witosław Klembowski zebrał ogólne informacje o założeniach całego programu i dokonał przeglądu lokalnych projektów.

podjęcia badań w gminach objętych Programem Inicjatyw Lokalnych była przede wszystkim naukowa ciekawość. Badanie nie zostało zamówione jako ewaluacja Programu. Zostało wymyślone, zaplanowane i przeprowadzone jako badanie naukowe, finansowane w ramach grantu KBN pn. *Regionalne i lokalne aspekty polityk sektorowych*. Jednym z celów badania było określenie roli czynników zewnętrznych (egzogennych) i wewnętrznych (endogennych) w rozwoju lokalnym (oraz regionalnym)[2]. Właśnie z tego powodu sięgnięto po przykłady gmin uczestniczących w programie – wychodząc z założenia, że oddziaływanie PIL było łatwym do uchwycenia bodźcem zewnętrznym. Gminy te były po prostu dogodnym pod względem metodologicznym materiałem empirycznym do zaplanowanego badania. Ponadto, dość długi odstęp czasu między zakończeniem programu (3 lata) a podjętym badaniem, dawał możliwość sprawdzenia oddziaływania programu i jego trwałości w dłuższym horyzoncie czasowym. W szczególności chodziło o odpowiedź na pytanie: czy dość istotna interwencja zewnętrzna w układy lokalne przyczyniła się do ich trwałej pozytywnej zmiany?

Badanie przeprowadzono w roku 1999, czyli 3 lata po zakończenia Programu, realizowanego w latach 1993–1996. Badanie nosi zatem niewątpliwie cechy ewaluacji typu *ex post* – należy jednak podkreślić, że horyzont czasowy był tutaj dłuższy niż w przypadku większości ewaluacji *ex post*, realizowanych zazwyczaj tuż po zakończeniu danego programu.

W założeniach badania głównymi odbiorcami jego wyników, obok szeroko pojętego środowiska naukowego zainteresowanego mechanizmami rozwoju regionalnego i lokalnego, były osoby praktycznie zarządzające rozwojem lokalnym (regionalnym), jak również instytucje projektujące i wdrażające programy rozwoju lokalnego (regionalnego).

Rozpatrując omawiane badanie jako ewaluację i przypisując je do poszczególnych typów ewaluacji, należy stwierdzić, że była to:

- ewaluacja *ex post* (badanie prowadzono 3 lata po zakończeniu programu);
- ewaluacja programu (przedmiotem badania były działania podejmowane w ramach programu PIL);
- ewaluacja zewnętrzna (badanie prowadzone było przez osoby niezwiązane z realizacją Programu i poszczególnych projektów).

---

[2] Najważniejsze endogenne czynniki rozwoju lokalnego, jakie zaobserwowano w Polsce na podstawie badań empirycznych, to: 1. lider lokalny, 2. elita lokalna, 3. instytucje lokalne, 4. zintegrowana społeczność miejscowych przedsiębiorców, 5. aktywna społeczność lokalna, 6. skłonność do współpracy międzygminnej (por. Gorzelak 2000). Czynniki egzogeniczne obejmują natomiast szeroki zakres zjawisk, takich jak: sytuacja gospodarki krajowej i światowej, uwarunkowania podatkowe, prawne, polityczne, technologiczne.

2. WKŁAD

2.2. Czym był Program Inicjatyw Lokalnych?

Program Inicjatyw Lokalnych na Rzecz Rozwoju Społeczno-Ekonomicznego, z budżetem 7,6 mln ECU, był częścią programu *Rozwój Społeczno-Gospodarczy* Phare, realizowanego ze środków Komisji Europejskiej. Celem PIL było „programowanie aktywnej odpowiedzi ze strony społeczności lokalnych na problemy rozwoju społeczno-ekonomicznego poprzez wspieranie przykładowych projektów oraz uruchomienie sieci współpracy i wymiany doświadczeń w całym kraju". W założeniach programu 10 gmin miało otrzymać wsparcie na realizację programów rozwoju lokalnego (w łącznej wysokości 6 mln ECU).

W wyniku wstępnej selekcji Ministerstwo Pracy i Spraw Socjalnych wytypowało 15 gmin, które charakteryzowały: znaczące zwolnienia grupowe, duży udział terenów popegeerowskich, peryferyjne położenie, jednorodny upadający przemysł państwowy. Odsetek bezrobotnych w wybranych gminach wynosił średnio 25%. W grupie znalazły się miejscowości małe i średnie (od 5000 do 60 tys. mieszkańców).

Kolejnym etapem selekcji była ocena strategii rozwoju gmin, w których prezentowały one pomysły na zagospodarowanie środków, jakie miały uzyskać z programu. Ostatecznie do programu zakwalifikowano: Biłgoraj, Działdowo, Kutno, Nidzicę, Lubawkę, Starachowice, Wicko, Zelów, oraz trzy gminy bieszczadzkie, wspólnie biorące udział w programie: Ustrzyki Dolne, Czarna i Lutowiska. Nie zakwalifikowały się: Białogard, Jawor, Krynki, Lubartów i Maków Mazowiecki.

W ośmiu gminach przewidywano uruchomienie funduszy gwarancji kredytowych dla miejscowych przedsiębiorców. Fundusze te skonsumowały od ok. dwóch piątych do dwóch trzecich przekazanych gminom środków. Wyjątkiem była Lubawka, w której połowa środków została przeznaczona na rewaloryzację starówki, a pozostałe środki na różne przedsięwzięcia związane z turystyką. Turystyka pojawiła się także w gminach bieszczadzkich i w Wicku. W pięciu gminach (Biłgoraju, Kutnie, Starachowicach, Wicku i Zelowie) planowano utworzenie instytucji wspierania małej przedsiębiorczości (inkubatora, ośrodka doradztwa, centrum promocji itp.). Kilka gmin przewidywało dofinansowanie konkretnych przedsięwzięć gospodarczych lub infrastrukturalnych, które mogłyby odgrywać rolę przykładów godnych upowszechnienia (gospodarstwo ekologiczne, plac zabaw, stacja obsługi i serwisu maszyn, pomoc gminnej spółdzielni itd.).

We wszystkich gminach powołano dwie instytucje: komitet lokalny, składający się z kilkudziesięciu osób reprezentujących różne sfery życia publicznego i gospodarczego w gminie, oraz agencję (spółkę akcyjną) lub fundację rozwoju lokalnego. Komitety przygotowywały programy, które stanowiły podstawę do ubiegania się o środki.

Tabela 1. Gminy biorące udział w PIL, środki finansowe i podjęte działania

Gmina	Województwo	Ludność 1994, w tys.	Budżet gminy w 1994 w mld starych złotych	PIL *per capita* w mln starych złotych	Udział PIL w budżecie w %	Najważniejsze podjęte (planowane) działania		
						funduszy gwarancji kredytowych	Instytucje wspierania przedsiębiorczości	Rozwój turystyki
Biłgoraj	lubelskie	26,2	61,0	607	26	×	×	
Działdowo	warmińsko-mazurskie	21,0	80,3	720	19	×		
Kutno	łódzkie	51,5	162,2	313	10	×	×	
Lubawka	dolnośląskie	12,7	45,2	1350	38			×
Nidzica	warmińsko-mazurskie	22,2	67,4	700	23	×		
Starachowice	świętokrzyskie	58,3	135,3	270	12	×	×	
Ustrzyki Dolne, Czarna, Lutowiska	podkarpackie	21,9	50,0	734	32	×		×
Wicko	pomorskie	5,4	19,9	3000	80	×	×	×
Zelów	łódzkie	16,4	41,2	980	38	×	×	

Źródło: opracowanie własne.

3. PROCES BADAWCZY

W badaniu wykorzystano przede wszystkim metody jakościowe: pogłębione wywiady indywidualne i analizy eksperckie, w tym analizę dokumentacji programowej. W pewnym stopniu uwzględniono także analizę danych statystycznych charakteryzujących badane gminy. Istotnym aspektem przyjętego podejścia badawczego była analiza porównawcza w zbiorze badanych gmin (zarówno zakwalifikowanych do PIL-u, jak i tych, które odpadły na ostatnim etapie kwalifikacji programu).

Badanie przebiegało w trzech głównych fazach. W pierwszej zebrano ogólne informacje o założeniach całego programu oraz dokonano przeglądu lokalnych projektów. Na tej podstawie opracowano wytyczne do pogłębionych wywiadów indywidualnych z interesariuszami w gminach objętych projektem. Drugą fazą były wywiady realizowane podczas wizyt w gminach. Ostatnią fazą było opracowanie zebranych materiałów i wywiadów, ich analiza i sformułowanie wniosków i rekomendacji. Wyniki badania zostały w syntetyczny sposób zaprezentowane w artykule opublikowanym w czasopiśmie naukowym.

Badanie gmin zorganizowano w ramach pięciu pól:

1) Stan gminy przed programem/sposób wejścia do programu.
2) Skład i funkcjonowanie komitetu lokalnego w trakcie i po realizacji programu.
3) Rola lidera.
4) Instytucje uczestniczące w programie oraz stworzone w wyniku jego realizacji.
5) Efekty gospodarcze realizacji programu.

W odniesieniu do każdego z pól sformułowano pytania badawcze:

- W jaki sposób dana gmina została włączona do programu (aktywność władz lokalnych czy szczególnie niekorzystna sytuacja na rynku pracy)?
- Jaki był skład i funkcjonowanie komitetu lokalnego oraz jego losy po zakończeniu programu, możliwość mobilizowania lokalnych elit oraz społeczności?
- Jaka była rola lidera w czasie realizacji programu i po jego zakończeniu?
- Czy inicjatywy podejmowane w toku realizacji programu zostały w trwały sposób zinstytucjonalizowane?
- Jaka była rola PIL-u w rozwoju gospodarczym badanych gmin?

Badaniem objęto wszystkie gminy ostatecznie zakwalifikowane do programu oraz 4 z 5 gmin, które odpadły w ostatnim etapie selekcji. Takie podejście umożliwiło porównanie gmin, które w ramach programu miały szansę jedynie na opracowanie strategii oraz tych, które uzyskały środki na realizację tej strategii.

Zakres czasowy badania określono szeroko. Za punkt wyjścia, określający sytuację w gminach przed realizacją programu, przyjęto okres od początku transformacji ustrojowej. Szczególną uwagę poświęcono okresowi realizacji programu, czyli latom 1993–1996. Natomiast długookresowe efekty PIL analizowano dla roku 1999. Krótko mówiąc, badaniem objęto lata 1989–1999.

Badanie zawierało zatem zarówno element identyfikacji, charakterystyki i oceny procesów, jak i analizę produktów programu – bezpośrednich i pośrednich. Dla wszystkich badanych gmin zebrano informacje dotyczące każdego z wymienionych wyżej pól, co w efekcie dało możliwość analizy porównawczej (wyróżnienia czynników sprzyjających – lub nie – sukcesowi programu w zbiorze badanych gmin). Objęcie analizą dość długiego okresu, z uwzględnieniem sytuacji przed realizacją i po realizacji programu, dało możliwość analizy dynamiki procesu wdrażania programu (wyróżnienie czynników, które determinowały przebieg programu w poszczególnych gminach).

Przy realizacji indywidualnych wywiadów pogłębionych starano się uwzględnić kluczowych aktorów z okresu realizacji badania (1999) oraz tych, którzy działali w czasie realizacji programu (1993–1996). Dotarcie do tych ostatnich nie zawsze było łatwe. W przypadku jednej z gmin (Wicko) okazało się, że w zasadzie nie ma już w niej osób dobrze zorientowanych w przebiegu programu. Inicjatorzy PIL-u wyjechali z gminy, pozostały jedynie osoby, których rola w programie była marginalna, tak więc ich wiedza nie była zbyt duża. Ponadto w tym konkretnym przypadku zatarciu wiedzy dotyczącej programu sprzyjało niepowodzenie jego realizacji. Brak informacji „z pierwszej ręki" utrudnił w znaczący sposób zgromadzenie informacji o przebiegu procesu realizacji programu.

Pewnym ograniczeniem zastosowanego podejścia było nieuwzględnienie w toku badania instytucji i osób odpowiedzialnych za wdrażanie programu. Jednakże w tym przypadku można to usprawiedliwić – badanie nie było typową ewaluacją, lecz miało charakter naukowy, zaplanowany tak, że uwzględnienie jedynie beneficjentów programu nie wpłynęło na jakość wyników. Należy jednak podkreślić, że jeśli byłaby to „zwykła" ewaluacja programu, należałoby – zgodnie z ogólnie przyjętymi standardami metodologicznymi – uwzględnić również szeroko pojęte instytucje zarządzające programem (zasada triangulacji, tj. spojrzenia na dany problem z kilku punktów widzenia).

4. PRODUKTY BADANIA

Produktem badania jest wiedza teoretyczna oraz praktyczne rekomendacje dla praktyki rozwoju lokalnego (regionalnego). W aspekcie teoretycznym uzyskano potwierdzenie tezy o prymacie czynników endogennych nad egzogennymi w rozwoju lokalnym. Na tej podstawie sformułowano

ogólne zalecenie dla programów pomocy układom lokalnym, które można streścić następująco: pomagać należy jedynie przygotowanym do wykorzystania tej pomocy, ponieważ skierowanie nawet znacznych środków do układów lokalnych, w których nie ma odpowiedniego klimatu do ich przyjęcia i racjonalnego zagospodarowania prowadzi do marnotrawstwa zasobów publicznych – gdzie indziej mogłyby być znacznie lepiej wykorzystane.

Realizacja badania pozwoliła ponadto wykazać skuteczność Programu Inicjatyw Lokalnych w długiej perspektywie czasowej – w większości gmin objętych programem osiągnięto znaczące i trwałe rezultaty.

Wyniki badania opracowano w formie raportu. Głównym sposobem rozpowszechniania wyników badania jest artykuł opublikowany w czasopiśmie naukowym „Studia Regionalne i Lokalne". Obecnie tekst artykułu dostępny jest nieodpłatnie na stronie internetowej kwartalnika (www.studreg.uw.edu.pl).

4.1. Jakie były wyniki badania?

Badanie potwierdził skuteczność programu. Ponadto dało możliwość wskazania czynników które wpłynęły pozytywnie bądź negatywnie na realizację PIL w poszczególnych gminach. W gminach, które osiągnęły w programie sukces, zbiegły się jednocześnie trzy pozytywne czynniki:

- ujawnił się niekwestionowany lider, *spiritus movens*, aktywności lokalnej;
- lider skupił wokół siebie wąską grupę osób najbardziej aktywnych i, wydaje się, najbardziej kompetentnych, które stymulowały prace szerszego zgromadzenia – komitetu lokalnego;
- szeroka elita była dość zwarta i uczestniczyli w niej lokalni przedsiębiorcy.

Jednocześnie w gminach tych nie wystąpiły dwa czynniki negatywne, obecne w innych przypadkach: konflikt polityczny oraz prywata, polegająca na tym, że środki zewnętrzne były traktowane jako zasilenie (czasem nawet 100-procentowe) prywatnych przedsięwzięć gospodarczych. Przeciwne proporcje owych czynników występowały w gminach, które albo nie uzyskały środków w ramach programu, albo też nie najlepiej je wykorzystały. Niekiedy czysto merkantylne podejście do możliwości uzyskania wsparcia finansowego powodowało konflikty w miejscowej elicie, której przedstawiciele zaczęli rywalizować o te środki. Z kolei konflikty polityczne w niektórych gminach paraliżowały dyskusję nad kierunkiem rozwoju gminy albo też – jeżeli występowały po okresie funkcjonowania programu – powodowały brak ciągłości w realizacji przyjętych wcześniej koncepcji.

Z porównania aktywności gmin przed przystąpieniem do programu ze skutkami, jakie przyniósł on w poszczególnych przypadkach, wyraźnie wynika, że stopień powodzenia jest proporcjonalny do stopnia

uprzedniego przygotowania elit lokalnych do podjęcia zadań zawartych w strategiach przygotowanych w ramach PIL. Spośród gmin ostatecznie niezakwalifikowanych do programu, jedynie w Lubartowie władze lokalne wykazywały dość znaczną aktywność w latach 1990–1993; żadna z gmin, w których PIL nie zakończył się sukcesem, nie przejawiała aktywności przed pojawieniem się programu. Odwrotnie – we wszystkich tych przypadkach, w których PIL przyczynił się do trwałej dynamizacji rozwoju lokalnego, samorząd terytorialny wykazywał znaczną – i kompetentną – aktywność już od samego początku funkcjonowania, a więc od 1990 roku.

Lider, elita, atmosfera współpracy to stosunkowo nietrwałe czynniki rozwoju lokalnego. Jest wysoce prawdopodobne, że dobremu wójtowi czy burmistrzowi prędzej czy później zostanie zaproponowane stanowisko w strukturach politycznych wyższego szczebla czy w biznesie; elita może zostać zdezintegrowana wewnętrznymi konfliktami, powstałymi wskutek pojawiających się sprzeczności interesów. Konflikty te mogą przenieść się do społeczności lokalnej, która z kolei może zachować się destrukcyjnie, odwołując radę w referendum. Sukces lokalny może zamienić się w chaos i gospodarczą stagnację.

Jedynym czynnikiem stabilizującym procesy rozwoju w gminie wydają się dwojakiego rodzaju instytucje: wspierania przedsiębiorczości i społeczeństwa obywatelskiego. W gminach PIL-owskich założone dla zarządzania programem agencje (w formie fundacji lub spółek akcyjnych) okazały się trwałą strukturą instytucjonalną, w większości przypadków zdolną do kontynuowania działalności po zaprzestaniu finansowania przez środki Phare. Sześć agencji funkcjonuje sprawnie i potrafi znaleźć popyt na swoje usługi. Agencja w Lubawce jest co prawda dość wątła kadrowo, lecz wykazuje znaczną aktywność w podejmowaniu inicjatyw społeczno-kulturalnych. Jedynie agencje w Kutnie i Wicku mają trudności z zapełnieniem portfela zamówień na szkolenia i doradztwo, co w tym ostatnim przypadku jest wynikiem stałego już nieporozumienia między zarządem agencji z kolejnym wójtami. Warto także podkreślić, że te z gmin, w których jeszcze przed przystąpieniem do programu działały agencje rozwoju lub inne instytucje wspierania przedsiębiorczości, środki PIL-u wykorzystały w najlepszy sposób, przynosząc największe efekty mnożnikowe i najtrwalsze skutki dla lokalnego rozwoju.

Ważną instytucją okazały się fundusze poręczeń, które uruchomiono w ośmiu gminach (jedynym wyjątkiem Lubawka). W dwóch przypadkach (w Biłgoraju i w Ustrzykach) fundusze przyjęły formułę gwarancji wzajemnych, polegającą na tym, że miejscowi przedsiębiorcy, wpłacając określone kwoty na rzecz funduszu, współdecydują o udzielonych poręczeniach. Formuła ta wydaje się mieć szczególnie duże znaczenie integracyjne.

Podsumowanie wyników badania w poszczególnych gminach zawiera tabela 2.

Tabela 2. Podsumowanie wyników badania w gminach

Gmina	Aktywność przed PIL	Aktywność i trwałość komitetu/elity	Obecność lidera	Trwałość instytucji	Efekty gospodarcze programu	Ogólna ocena sukcesu PIL i trwałości jego efektów	Udział PIL w budżecie w %
Gminy, które znalazły się w PIL-u							
Biłgoraj	+	+	+	+	+	++	26
Działdowo	–	+	–	+	+	+	19
Kutno	+	–	+	+	+	+	10
Lubawka	+/–	+	+	–	+/-	–	38
Nidzica	+	+	+	+	+	++	23
Starachowice	+	+/–	+	+	+	++	12
Ustrzyki Dolne, Czarna, Lutowiska	–	+	+	+	+	+	32
Wicko	–	–	+/–	+/–	–	–	80
Zelów	+	+	+	+	+	++	38

Gminy, które nie znalazły się w PIL-u							
Jawor	–	+/–	–	–	–	–	
Krynki	–	–	+/–	–	–	–	
Lubartów	+/-	+	+/-	+	+	+	
Maków Mazowiecki	–	+	–	–	+/–	–	
Pionki	–	–	–	–	+/–	–	
Białogard	nie był badany						

Źródło: opracowanie własne.

5. WNIOSKI ZE STUDIUM PRZYPADKU

Początkowe założenia badawcze zostały w pełni zrealizowane. Przebadano wszystkie gminy, które planowano przebadać. Przyjęte metody sprawdziły się i umożliwiły zebranie poszukiwanych informacji. Jak zakładano, przeprowadzone analizy pozwoliły na realizację głównego celu badania, czyli charakterystykę endogennych i egzogennych czynników rozwoju lokalnego. Niejako „przy okazji", i bez wyartykułowanej *explicite* intencji, przeprowadzono również swoistą ewaluację *ex post* programu PIL. Pokazuje to, że (przynajmniej niektóre) społeczno-ekonomiczne badania naukowe (rozwoju lokalnego/regionalnego) mogą mieć istotny wymiar aplikacyjny jako ewaluacja, oraz że ewaluacja, jako taka, może mieć charakter badania naukowego.

Największym problemem, jaki napotkano w toku realizacji badania, było dotarcie do wszystkich kluczowych aktów związanych z realizacją programu w gminach. Kilka lat po zakończeniu programu nie wszyscy z nich nadal aktywnie uczestniczyli w życiu badanych społeczności (niektórzy nawet opuścili daną gminę). W przypadku niniejszego badania, mimo pewnych trudności, szczególnie w przypadku jednej gminy, udało się uzyskać odpowiednie informacje. Niemniej jednak należy podkreślić, że w przypadku ewaluacji *ex post*, im dłuższy odstęp między końcem programu a podjęciem badania ewaluacyjnego, tym większe ryzyko wystąpienia problemów związanych z pozyskaniem informacji, szczególnie takich, których „nośnikami" są ludzie, czyli dotyczących przede wszystkim przebiegu wdrażania programu (lub projektu). Wydaje się, że im dłuższa perspektywa, z której analizowana jest skuteczność danego działania (programu, projektu), tym trafniej można ocenić jego skuteczność, efektywność oraz przede wszystkim trwałość. Jednak im więcej czasu dzieli badanie i zakończenie realizacji badanego działania, tym, ogólnie rzecz biorąc, słabszy dostęp do informacji, szczególnie informacji kontekstowych i tych związanych z doświadczeniami poszczególnych interesariuszy. Wyznaczając horyzont badawczy ewaluacji *ex post*, należy zatem poszukiwać „złotego środka" między możliwością oceny efektów i trwałości oraz dostępem do informacji. Należy podkreślić, że przyjęta w przypadku analizowanego badania trzyletnia perspektywa nie była zbyt długa, aby uniemożliwić zdobycie wymaganych informacji. Co więcej, wydaje się, że jej wydłużenie nie wpłynęłoby w znaczący sposób na ocenę efektów i trwałości, natomiast mogłoby w znaczący sposób utrudnić charakterystykę procesu wdrażania programu.

Spośród wielu różnych funkcji ewaluacji omawiane badanie spełniło przede wszystkim funkcję poznawczą. Związane było to głównie z tym, że nie było zaplanowane jako ewaluacja, lecz jako badanie o charakterze naukowym. Niemniej jednak, niejako przy okazji, zostały również spełnione inne funkcje ewaluacji. Wnioski z badania sformułowano tak,

że mogły zostać bezpośrednio wykorzystane do opracowywania nowych programów rozwoju lokalnego/regionalnego – zatem spełniono również funkcję racjonalizowania planowania. Chociaż w tym przypadku lepiej mówić o potencjalnym spełnieniu tej funkcji, gdyż ciężko wskazać konkretne przypadki wykorzystania wiedzy wypracowanej w badaniu do planowania nowych inicjatyw. Ponadto wyraźna jest też funkcja rozliczenia i wzmocnienia odpowiedzialności – charakterystyka i ocena programu zawarta w opracowanym na podstawie badań artykule była (i jest) dostępna dla wszystkich zainteresowanych.

BIBLIOGRAFIA

Gorzelak G. (2000), Zewnętrzna interwencja jako czynnik rozwoju lokalnego (na przykładzie Programu Inicjatyw Lokalnych), *Studia Regionalne i Lokalne* **3(3)**.

ROZDZIAŁ 5.6.

Martin Ferry

STUDIUM PRZYPADKU Z WIELKIEJ BRYTANII: ŚREDNIOOKRESOWA EWALUACJA SPECJALNEGO PROGRAMU PRZEJŚCIOWEGO HIGHLANDS & ISLANDS 2000–2006

1. KONTEKST I PRZEDMIOT EWALUACJI[1]

Program pt. *Highlands & Islands Special Transitional Programme* (HISTP) był realizowany w latach 2000–2006 i przewidywał wydatkowanie 308 mln EUR (ostatecznie przyznano dodatkowo rezerwę wykonawczą, która spowodowała wzrost wartości programu do 319,8 mln EUR). Wsparcie europejskie przyczyniło się znacząco do rozwoju regionu Highlands & Islands, wynikającego z poprzedniego programu realizowanego przez HISTP w ramach Celu 1 w latach 1994–1999. HISPT uznano jednak za ostatnią możliwość przyznania regionowi Highlands &

[1] Niniejsza część tego rozdziału powstała przede wszystkim na podstawie prac przeprowadzonych przez EPRC na zlecenie Komisji ds. Przedsiębiorczości i Kultury parlamentu Szkocji w celu zbadania wpływu programów współfinansowanych przez fundusze strukturalne UE w Szkocji w latach 1994–2006. Zob. Davies et al. 2007.

Islands dostępu do funduszy europejskich w tej skali. Z tej przyczyny podstawowym celem, przyjętym dla HISPT, było zapewnienie, by korzyści lub wartość dodana przez fundusze strukturalne mogły zostać określone w sposób umożliwiający utrzymanie długoterminowego postępu po 2006 roku.

Odpowiedzialność za wykorzystanie funduszy strukturalnych w Szkocji została powierzona Szkockiej Egzekutywie (rząd Szkocji od 2007 roku). Programami zarządzały Egzekutywy ds. Zarządzania Programami (Programme Management Executives PME). Były to odrębne organizacje, działające w ramach uprawnień delegowanych przez Szkocką Egzekutywę w celu zarządzania programami w imieniu partnerstw regionalnych. Działały one jako niezależne jednostki, wspierające wzrost interakcji z partnerami, szczególnie w dziedzinie opracowywaniu projektów oraz promujące większy dostęp do programu. Funkcję PME dla *Highlands & Islands Special Transitional Programme* pełnił program *Highlands & Islands Partnership Programme* (HIPP). Działalność ta była realizowana w ramach spółki z ograniczoną odpowiedzialnością o statusie spółki charytatywnej – Highlands & Islands (Scotland) Structural Funds Partnership Ltd. Podstawowym zadaniem PME było administrowanie programem, obsługa różnych Komisji i struktur Grupy Doradczej oraz ułatwianie organizacjom partnerskim dostępu do finansowania europejskiego. Partnerami były instytucje sektora publicznego oraz organizacje społeczne/charytatywne, aczkolwiek członkowstwo było dostępne dla wszelkich organizacji spełniających kryteria uprawniające do wnioskowania o finansowanie z funduszy EFS, EFRR czy EAGGF (trzeciej strony).

HISTP obejmował cztery fundusze:

- Europejski Fundusz Rozwoju Regionalnego (EFRR);
- Europejski Fundusz Społeczny (EFS);
- Europejski Fundusz Orientacji i Gwarancji Rolnej (EAGGF);
- Instrument Finansowy dla Rybactwa (FIFG).

Program posiadał 4 Priorytety, 14 Działań w ramach Priorytetów 1, 2 i 3 oraz dalszych 14 Poddziałań w ramach Priorytetu 4, z których jedno podlegało dalszemu podziałowi. Łącznie realizowano więc 4 Priorytety i 29 Działań lub Poddziałań.

- Priorytet 1: Wzrost konkurencyjności działalności gospodarczej – tworzenie zatrudnienia i zwiększanie dochodów;
- Priorytet 2: Tworzenie warunków dla konkurencyjności regionalnej;
- Priorytet 3: Rozwój zasobów ludzkich;
- Priorytet 4: Pomoc dla społeczności wiejskich;
  - 4.a Rozwój terenów wiejskich i rybactwa;
  - 4.b: Rozwój ekonomiczny Wspólnoty.

MTE pracowała pod kierunkiem Frasier's Associates, specjalistycznej firmy konsultingowej, ukierunkowanej na usprawnienia polityki

Tabela1. Alokacje finansowe według priorytetów w SPD (mln EUR)

	Koszt razem	Razem sektor publiczny	Razem UE
Poziom programu	853,114	675,837	308,502
Priorytet 1	187,68	153,68	66,427
Priorytet 2	315,81	281,28	106,06
Priorytet 3	120,15	116,15	58,076
Priorytet 4	220,13	115,38	73,261
Pomoc techniczna	9,35	9,35	4,675

Źródło: SPD 2000–06, pkt 8.1

rozwoju gospodarczego oraz na proces uczenia się i zastosowania lekcji, wynikających z wdrażania projektów i programów. W skład zespołu ewaluacyjnego wchodziło także Centrum Badań Polityk Europejskich (European Policies Research Centre, EPRC) – niezależny instytut badawczy, działający w ramach Business School of the University of Strathclyde w Glasgow w Szkocji. EPRC prowadzi badania akademickie i oferuje doradztwo oraz szkolenia w dziedzinie projektowania, wdrażania i ewaluacji polityki. W skład zespołu ewaluacyjnego wchodziła także spółka Rural Development Company, specjalizująca się w doradztwie w dziedzinie rozwoju gospodarczego terenów wiejskich.

2. WKŁAD

2.1. Potrzeby i zapytania informacyjne

Ponieważ MTE zostały zrealizowane w ramach wymogów Komisji Europejskiej, podstawowe Istotne Warunki Zamówienia (Terms of Reference; ToR) dla ewaluacji zostały opracowane na podstawie wytycznych Komisji. Ewaluacja średniookresowa programów realizowanych w ramach Celu Pierwszego i Celu Drugiego w latach 2000–2006 była wymagana na mocy postanowień art. 42 rozporządzenia Rady (WE) nr 1260/99. Na mocy tych przepisów, celem ewaluacji średniookresowej było stwierdzenie, w jakim stopniu program spełniał założone zadania; sprawdzenie ciągłości znaczenia strategii programu oraz zalecenie wszelkich działań, koniecznych do poprawy jakości i wyników programu.

W Wielkiej Brytanii podobną syntezę kluczowych celów Komisji przedstawiło Ministerstwo Transportu, Samorządów i Regionów (Department for Transport, Local Government and the Regions DTLR). Komisja wyraziła zgodę na ToR, które następnie zostały dostosowane do potrzeb przez Grupy Sterujące Ewaluacją (Evaluation Steering Groups) na poziomie lokalnym w celu spełnienia konkretnych kontekstów programowych. Zgodnie z nową koncepcją, Grupy Sterujące mogły raczej dodawać szczegółowe, a nie nowe obszary działania. Ponadto DTLR przyjęło indykatywne budżety ewaluacji programów. Mimo braku konkretnych danych liczbowych, budżety dla MTE były postrzegane jako „bynajmniej nie oszczędne". Podobno Komisja wyraziła zadowolenie z poziomu przydzielonych środków, które świadczyły o powadze, z jaką Wielka Brytania przystąpiła do tego działania.

Warto tutaj dodać, że przynajmniej w kontekście Wielkiej Brytanii MTE utworzono w innym celu niż w toku poprzednich ewaluacji. Tradycyjnie MTE działały na silnych podstawach formatywnych: zapewniały programom możliwości rewidowania ustaleń zarządczych i procesów wdrażania w toku realizacji, oferując przy tym zdecydowanie mniej informacji pod względem produktów/wyników czy wpływów. Jednakże w wypadku MTE dla Wielkiej Brytanii w latach 2000–2006 mamy do czynienia ze znaczącą różnicą, wynikającą z udzielenia znacznie większego priorytetu sumarycznej ewaluacji już osiągniętych produktów i wyników (Raines, Taylor 2002).

Przyczyn było kilka. Pierwszą była silna motywacja ze strony Rezerwy Wykonawczej (Performance Reserve). Była to nowa cecha, przyjęta w celu motywowania beneficjentów ostatecznych programów. Z czterech procent wartości funduszy strukturalnych przyznanych każdemu z państw członkowskich, utworzono rezerwę do 2003 roku, przeznaczoną do podziału najpóźniej do 31 marca 2004 roku pomiędzy programy, które osiągnęły najlepsze wyniki. Każde państwo członkowskie przedstawiało Komisji propozycje alokacji tej rezerwy na podstawie wskaźników monitorowania, wprowadzanych przez same państwa. Przez ustanawianie kluczowych wytycznych ministerstwa rządowe chciały zachować większą niż w poprzednich ewaluacjach kontrolę nad procesem ewaluacji. Ponieważ Rezerwa Wykonawcza zależała od wyników ewaluacji średniookresowej, pojawiły się głosy wyrażające wątpliwości co do obiektywności ewaluacji przez programy ich własnych wyników w sytuacji, gdy nagrodą były środki z Rezerwy Wykonawczej). Powszechnie uznano, że ocena postępu w realizacji programów w celu osiągnięcia celów Rezerwy Wykonawczej stanowiła jeden z najważniejszych problemów, związanych z ewaluacjami środterminowymi.

Drugim czynnikiem było przekonanie, że minął już wystarczający czas realizacji programów, by można było przeprowadzić pewne oceny wyników programów, wykraczające poza ewaluacje śródterminowe

w poprzednich okresach programowania. W konsekwencji kilku menedżerów programów z Wielkiej Brytanii uznało ewaluacje śródokresowe za szansę na przeanalizowanie konkretnych kwestii związanych z wynikami. Uznano także, że Komisja wywiera w swoich własnych wytycznych naciski na przeprowadzenie sumarycznej ewaluacji (w Dokumencie Roboczym Komisji/*Commission Working Paper* 8).

Stąd podstawowe ToR dla MTE w Wielkiej Brytanii obejmowały następujące wymogi:

- komentarz odnośnie do tego, czy i w jakim stopniu strategia programu określona w Jednolitym Dokumencie Programowania (*Single Programming Document*, SPD) oraz Uzupełnieniu Programu jest nadal dostosowana do warunków społeczno-ekonomicznych danego obszaru i pozostaje zgodna z innymi strategiami dla danego regionu;
- określenie postępu programu w kierunku osiągnięcia jego celów, ustalonych w SPD i Uzupełnieniu Programu;
- ocena jakości i skuteczności wdrożenia projektu i kierowania nim, w tym identyfikacja wszelkich słabych stron systemów monitorowania oraz przedstawienie zaleceń odnośnie do usprawnień;
- zbadanie i zgłoszenie wyników, osiąganych w stosunku do wskaźników stosowanych w rezerwie wykonawczej.

Dla osiągnięcia tych celów ewaluacje musiały przynieść odpowiedzi na następujące pytania:

- Czy potrzeby danego obszaru, zdefiniowane w dokumencie programowym, uległy zmianom od czasu zatwierdzenia programu oraz czy strategia programu jest nadal odpowiednia? Działanie to będzie wymagać szczegółowej oceny ciągłej ważności analizy *ex ante* programu oraz wyjaśnienia wszelkich zmian w obszarze, co do których można przypuszczać, że wymagają reakcji ze strony strategii programu.
- Jakie zmiany, o ile w ogóle, są konieczne dla strategii programu oraz planów jego realizacji? Mając na uwadze wszelkie potrzebne zmiany, określone powyżej, będzie to wymagać zbadania stopnia, w jakim priorytety wydatków i planowane działania, określone w SPD i Uzupełnieniu Programu, są nadal właściwe dla realizacji strategii programu.
- Czy uzgodnione priorytety horyzontalne – w tym zwłaszcza kwestia równość szans oraz ochrony środowiska – zostały z powodzeniem zintegrowane z programem; czy dokonano wkładu w Krajowy Plan Działania na rzecz Zatrudnienia?
- Jakie postępy osiągnięto w zakresie osiągnięcia celów ilościowych odnośnie do wydatków, produktów brutto i wyników? Będzie to wymagać analizy na poziomie działań programowych i musi obejmować: wyjaśnienie wszelkich różnic pomiędzy celami uzgodnionymi a faktycznie osiągniętymi; komentarz odnośnie do realizmu celów;

zalecenia wszelkich ulepszeń lub usprawnień w odniesieniu do kwantyfikacji celów i w pełni uzasadnionych szacunkowych kosztów jednostkowych na odpowiedniej próbie działań i większych projektów.

- Jakie postępy osiągnięto w celu zrealizowania planowanych wpływów programu? Będzie to wymagać poinformowanej i w pełni uzasadnionej oceny prawdopodobieństwa osiągnięcia planowanych wpływów programu na podstawie dotychczasowych wyników w osiąganiu celów operacyjnych. Należy także uwzględnić krótką analizę interakcji programu z innymi interwencjami w ramach odpowiedniej polityki lub programu, finansowanymi przez rząd centralny lub samorządy lokalne.
- Jakie wyniki osiągnięto w stosunku do wskaźników uzgodnionych w ramach Rezerwy Programowej odnośnie do skuteczności, zarządzania i wdrożenia finansowego? Należy tutaj uwzględnić zestawienie faktycznych osiągnięć z celami śródokresowymi, określonymi w SPD/Uzupełnieniu Programu (do daty wyraźnie określonej w raporcie); krótką analizę trendów w tych danych oraz opinię na temat zakresu, w jakim wyniki są reprezentatywne dla ogólnego wyniku programu.
- Jak skuteczne są procesy, stosowane w realizacji programu? Będzie to wymagać analizy procesu wyboru projektów, procesów oceny i zatwierdzania, zarządzania projektami i programami, udzielania zleceń i przetwarzania roszczeń, monitorowania wyników projektów, inspekcji finansowej i zgłaszania nieprawidłowości oraz działań uprawniających.
- Jaka była wartość dodana programu w tym regionie? Wymaga to określenia dodatkowych korzyści (o ile istnieją), wniesionych przez program, a które nie zostałyby osiągnięte bez tego programu. Przewiduje się, że reakcja będzie oparta na szczegółowej analizie zebranych wcześniej dowodów w opisanych powyżej sprawach.
- Jak wspomniano, podstawowe ToR zostały zaadaptowane do szczegółowych kontekstów programowych. W Szkocji, z HISTP włącznie, główne różnice dotyczyły znaczenia przywiązywanego do tematów horyzontalnych (trwałość środowiskowa oraz równość szans) oraz Wspólnotowej Wartości Dodanej (Community Added Value) programów. W odniesieniu do tej ostatniej kwestii, podkreślenie jej w szkockich ToR dotyczących ewaluacji odzwierciedla znaczenie ewaluacji we wnoszeniu wkładu do debaty po 2006 roku, gdy poziomy finansowania dostępnego dla Wielkiej Brytanii zostały znacząco ograniczone. Istotne znaczenie ma tutaj określenie wartości dodanej programów w celu zapewnienia zrównoważonego rozwoju, nawet w kontekście ograniczonego finansowania (władze angielskie już wcześniej zaplanowały szczegółową ewaluację wartości dodanej funduszy).

2.2. Kwestie metodologiczne w ToR

Znaczenie przypisywane ocenie postępów fizycznych i finansowych w realizacji Programu w ramach ewaluacji spowodowało, że kwestie metodologiczne uznano za bardzo ważne. MTE musiała spełnić wymóg zbadania prób projektów pod względem produktów i wyników. Działania te obejmowały analizę oddolną oraz odgórne badania scentralizowanych informacji projektowych na temat wydatków i podanych produktów. Należy tutaj zwrócić uwagę na fakt odłożenia znaczących budżetów na zapewnienie, by ewaluatorzy byli w stanie zrealizować intensywne prace empiryczne. ToR przewidywały rozległe bibliografie, kierunkujące ewaluatorów na ewaluacje kluczowych polityk krajowych oraz wytycznych rządu centralnego odnośnie do sposobu realizowania ewaluacji. Ponadto nie omawiano szczegółowej metodologii, a przede wszystkim przewidywano, że propozycje zostaną przedstawione przez samych ewaluatorów. Ewaluatorzy mieli możliwość adaptacji ToR do własnej wiedzy fachowej i doświadczenia. W wypadku MTE dla Highlands & Islands, zespół ewaluatorów łączył w sobie doświadczenie i wiedzę fachową w zakresie zarówno ilościowych, jak i jakościowych aspektów ewaluacji.

2.3. Dostępność danych i harmonogram

Jeden z głównych wniosków, wynikających z wcześniejszych ewaluacji śródterminowych, realizowanych w Wielkiej Brytanii i innych państwach, dotyczył osłabienia rzetelności ewaluacji z powodu problemów, związanych z monitorowaniem danych. Niespójne wskaźniki, niezweryfikowane roszczenia oraz problemy techniczne, związane z magazynowaniem danych, przyczyniły się do przedstawienia w wyniku kilku ewaluacji śródterminowych zaleceń zmian w systemie zbierania danych w ramach kilku programów. Menedżerowie programów zareagowali na ten fakt, a wprowadzenie serii zmian skutkowało zwiększeniem prawdopodobieństwa dostępności danych i ich wyższej jakości dla MTE 2000–2006.

Mimo to należy jeszcze rozwiązać wiele problemów. Systemy archiwizacji danych nie zostały do tej pory zreformowane zgodnie z uwagami ewaluatora z poprzednich ewaluacji. Żadne z terytoriów Wielkiej Brytanii nie opracowało zintegrowanych, rzetelnych systemów archiwizacji danych. W Szkocji budowa jednolitej bazy danych, obejmującej wszystkie szkockie programy, została opóźniona w ramach zarządzania przez Szkocką Egzekutywę. Programy takie jak HISTP musiały korzystać z tymczasowej bazy danych dotyczących roszczeń, co prawdopodobnie utrudni ewaluację. Ponadto, nadal istnieją pewne wątpliwości co do rzetelności definicji danych. Dzięki bardziej szczegółowej

definicji wskaźników dla programów osiągnięto znaczące usprawnienia; podjęto także kroki w kierunku publikacji tych definicji dla menedżerów projektu. Nadal jednak utrzymują się podejrzenia dotyczące problemów z interpretacją tych definicji danych przez menedżerów projektów.

Przetarg na ewaluację odbył się w czerwcu 2002 roku, ewaluatorzy zostali wybrani w czerwcu 2002 roku, a ewaluacja toczyła się do września 2003 roku.

3. PROCES

3.1. Podejście ewaluatora

Połączono analizę centralnie archiwizowanych danych, przeprowadzoną przez HISTP MTE oraz zastosowano empiryczną weryfikację próby projektów. Zespół ewaluacyjny MTE podzielił badania na sześć faz:

- Projektowanie materiałów badawczych, analiza wstępna postępów finansowych, analiza wstępna postępów fizycznych oraz wstępny przegląd kwantyfikacji.
- Konsultacja z interesariuszami, wprowadzanie i analiza danych, opracowanie raportu śródokresowego.
- Walidacja zgłoszonych wyników, analiza znaczenia strategicznego, przegląd procesów przewidzianych Programem, przegląd Wspólnotowej wartości dodanej, wkład w równoległe inicjatywy związane z polityką, przegląd integracji tematów horyzontalnych.
- Ostateczny przegląd postępów finansowych, oszacowanie wpływów na Program, prognoza postępów w kierunku spełnienia celów Rezerwy Wykonawczej oraz przegląd zmian w strategii.
- Opracowanie Projektu Raportu Końcowego oraz Raportu Końcowego.
- Przegląd zaktualizowanych szacunków postępów w realizacji celów Rezerwy Wykonawczej.

Zakończenie realizacji tych faz wiązało się z kilkoma istotnymi wyzwaniami dla ewaluatorów, które trzeba było rozwiązać w toku projektowania badań. Jak stwierdzono powyżej, wcześniejsze ewaluacje śródterminowe, realizowane w Wielkiej Brytanii i innych krajach napotykały problemy z monitorowaniem danych. Niespójne wskaźniki, niezweryfikowane roszczenia i problemy techniczne z archiwizacją danych pogarszały rzetelność ewaluacji. Jednakże w budżecie ewaluacji MTE przeznaczono znaczące środki na zapewnienie wszechstronnych badań empirycznych. Aby opracować rzetelne wnioski, ewaluatorzy podjęli konkretne działania w celu walidacji siły danych z monitoringu. Ewaluatorzy spotkali się z 45 spośród 70 posiadaczy projektów, by upewnić się, że dane monitoringowe są zbierane w sposób zapewniający

ich istotność i dokładność. Wskaźniki zostały opracowane centralnie, pojawiło się jednak ryzyko braku spójnego podejścia menedżerów projektów do definicji i zrozumienia, co ma być mierzone na poziomie projektu. Kolejny przykład z innego MTE w Wielkiej Brytanii dotyczy menedżera projektu, który w formularzu roszczeń powoływał się na utworzenie czterech atrakcji turystycznych, podczas gdy bliższa inspekcja wykazała, że chodzi o cztery budynki, wchodzące w skład jednej i tej samej atrakcji turystycznej. Obejmując dużą próbę projektów podczas wizyt, ewaluatorzy ograniczyli te niezgodności, uzyskując w konsekwencji bardziej rzetelny, śródokresowy obraz postępów w realizacji programu.

Innym podstawowym wyzwaniem było opracowanie sposobu zarządzania kompleksową, wieloaspektową ewaluacją, obejmującą „odgórną" analizę dokumentową oraz „oddolne" badania na poziomie projektu. Ewaluacja śródokresowa dotyczyła czterech głównych tematów: strategii, zarządzania, wyników i celów horyzontalnych (Scottish Executive 2003). W celu właściwego zarządzania, zespół ewaluacyjny postanowił zorganizować badania w formie serii Dokumentów Roboczych. W miarę kończenia kolejnych etapów badań ewaluatorzy rozpowszechniali raport, zawierający wyniki takiego konkretnego przedsięwzięcia.

Tematy Dokumentów Roboczych obejmowały: *Integrated Rural Development*; *Horizontal Objectives*; *Physical Progress*; *Contribution to the European Employment Strategy*; *Validation of Gross Outcomes*; *Quantification of the Programme*; *Community Added Value*; *Implementation Arrangements*; oraz *Progress Towards Performance Reserve Targets*. Koncepcję podziału pracy na serię dokumentów roboczych postrzegano jako szczególnie innowacyjną. Dzięki przyjęciu takiego podejścia wyniki każdego etapu badań mogły być rozpowszechniane niezwłocznie, bez potrzeby czekania na opracowanie Projektu Raportu Końcowego czy samego Raportu Końcowego. Z tej przyczyny podział badań na oddzielne sekcje skutkował zwiększeniem dostępności, lepszymi możliwościami rozpowszechniania, większą interakcją z interesariuszami (która już była istotna w wyniku wszechstronnego podejścia do dobierania prób projektów). W konsekwencji partnerstwo odegrało aktywną rolę w procesie badawczym z powodu zajęcia się tymi kluczowymi problemami przez ewaluatorów. Dzięki przyjęciu tego podejścia, obyło się bez większych niespodzianek dla interesariuszy pod koniec prac. Po zakończeniu badań ewaluacyjnych Dokumenty Robocze zostały zestawione jako rozdziały ostatecznego Raportu Technicznego, który towarzyszył Końcowemu MTE.

Wyzwania o bardziej jakościowym charakterze dotyczyły oceny Wspólnotowej Wartości Dodanej. Koncepcja ta jest postrzegana jako trudna do ewaluacji. Obejmuje ona ocenę „dodatkowego wpływu zasobów i procedur programu na organizacje krajowe" (Scottisch Executive:

60). W celu rozwiązania tego problemu ewaluacja śródokresowa przeznaczyła oddzielny rozdział kwestii Wspólnotowej Wartości Dodanej. Dzięki wiedzy i doświadczeniu zespołu ewaluacyjnego (a szczególnie EPRC) podjęto nowoczesne podejście do tej części ewaluacji, dzieląc koncepcję na sześć obszarów, w których HISTP mógł potencjalnie przynieść wartość dodaną organizacjom krajowym: opracowywanie strategii, partnerstwo, adycjonalność projektu, skutki finansowania, skutki polityki; oraz skutki administracyjne.

4. PRODUKTY

4.1. Jaką wiedzę wytworzono?

Ewaluacja średniookresowa skutkowała powstaniem wiedzy deskryptywnej i proskryptywnej oraz informacjami o charakterze jakościowym i ilościowym. Dotyczyło to czterech głównych tematów: strategii, zarządzania, wykonania i celów horyzontalnych (Scottish Executive 2003). W odniesieniu do **strategii**, uznano znaczący postęp w porównaniu do poprzedniego programu w kategoriach analizy, strategicznego ukierunkowania, celów, kwantyfikacji celów oraz wdrożenia celów horyzontalnych; za główną wadę uznano niemożność przeprowadzenia jednoznacznych wyborów priorytetów. W tym kontekście ewaluacja sprzyjała większej podregionalnej priorytetyzacji w uznaniu faktu, że program *Highlands & Islands* składał się ze zbioru względnie odrębnych gospodarek regionalnych. W odniesieniu do **zarządzania** uznano, że ustalenia wdrożeniowe ogólnie zadziałały prawidłowo; jednocześnie jednak uznano, że Komitet Monitorujący Programu powinien wzmocnić swoje myślenie i debatę strategiczną, natomiast proces i kryteria wyboru projektów powinny zostać uproszczone, a systemy informacji dla zarządu należy usprawnić. Jak już wspomniano, ewaluacja rozpatrywała miary wykonania w pewnym stopniu szczegółowości. Przeprowadzono szczegółowy przegląd postępów finansowych (w porównaniu z profilem finansowym SPD) – według Funduszu, Priorytetu i Działania. W ramach ewaluacji średniookresowej przeprowadzono przegląd postępów finansowych według Priorytetów w pewnym stopniu szczegółowości (zob. pkt 6 MTE). Postępy finansowe rozpoczęły się powodzeniem w ramach EFRR, a w ramach EFS były realizowane z wyprzedzeniem w stosunku do harmonogramu, zauważono jednak rosnące ryzyko obniżenia zaangażowania. W tabeli 2 przedstawiono podsumowanie postępów w tym okresie[2].

[2] Dane liczbowe podano w mln GBP według kursu wymiany 1:£0,66.

Tabela 2. Postępy finansowe według priorytetów[2]

Priorytet	SPD Alokacja (mln £)	Fundusze zaangażowane Oczekiwane: 58,6%		Fundusze – roszczenia Oczekiwane: 29,9%	
		(mln £)	(%)	(mln £)	(%)
1: Konkurencyjność działalności gospodarczej	43,842	29,511	67,3	10,517	24,0
2: Warunki dla konkurencyjności regionalnej	70,002	30,152	43,1	11,685	16,7
3: Rozwój zasobów ludzkich	38,330	27,792	72,5	15,646	40,8
4a: Rozwój obszarów wiejskich i rybactwa	42,282	27,226	64,4	6,527	15,4
4b: Rozwój ekonomiczny Wspólnoty	6,070	3,982	65,6	0,687	11,3
5. Pomoc techniczna	3,086	1,864	60,4	0,867	28,1
Razem	203,612	120,527	59,2	45,929	22,6

Źródło: MTE, Tabela 6.3.

MTE obejmuje także szczegółowy opis przyczyn takiego postępu (lub jego braku) według Działań. Wspomniane kwestie obejmują:

- spowolnienie rekodowania wydatków w ramach wsparcia opartego na schematach,
- charakter niektórych schematów (np. inżynieria finansowa),
- płaski przebieg rozwoju gospodarczego, szczególnie w niektórych sektorach,
- długie okresy oczekiwania w czasie realizacji dużych projektów infrastrukturalnych oraz związane z nimi uzależnienie od kilku głównych projektów,
- problemy współfinansowania, szczególnie w odniesieniu do wkładów z sektora prywatnego,
- konkurencję ze strony „rywalizujących" schematów krajowych o mniej uciążliwych wymogach administracyjnych,
- wykorzystanie pośredników w dystrybucji funduszy (szczególnie przez schematy),
- czynniki zewnętrzne, w tym przepisy UE i krajowe.

Przeprowadzono także przegląd postępów fizycznych według Priorytetów, z odróżnieniem produktów fizycznych od wyników śródokresowych. Dla każdego z czterech Priorytetów opracowano szczegółowe

tabele. W ramach produktów osiągnięto w zasadzie dobre postępy (w nieco mniejszym stopniu w przypadku Priorytetu 4); wyniki były jednak bardziej mieszane, z uwzględnieniem poglądu, że niektóre cele mogły zostać przyjęte w sposób konserwatywny. Na koniec, w odniesieniu do celów horyzontalnych, ewaluacja wykazała, że ich integracja była raczej nierównomierna.

Na poziomie bardziej jakościowym, ewaluacja poświęciła swój ostatni rozdział Wspólnotowej wartości dodanej – określonej jako „dodatkowy wpływ zasobów i procedur programu na organizacje krajowe" (Scottish Executive 2003: 60). Przegląd wartości dodanej przeprowadzono w odniesieniu do rozwoju strategii, partnerstwa, adycjonalności projektu, skutków finansowania, skutków polityki oraz skutków administracyjnych.

- **Rozwój strategii**. Aspekt ten jest często postrzegany jako istotna cecha pozytywna funduszy strukturalnych – także w ramach *Highlands & Islands*. Interesariusze postrzegali strategię głównie jako „ramy obejmujące oddzielne priorytety strategiczne w skali podregionalnej lub organizacyjnej" (Scottish Executive 2003: 60). Stwierdzono, że wynika to głównie z braku krajowego odpowiednika SPD. HISTP zapewnił platformę dla innowacji (szczególnie w odniesieniu do rozwoju gospodarczego wspólnoty oraz trwałego rozwoju), które zostały później wykorzystane w skali krajowej.
- **Partnerstwo**. Fundusze strukturalne obejmują partnerstwo zarówno w skali programów, jak i projektów. Na poziomie programów partnerstwo już od dawna funkcjonowało w programie *Highlands & Islands* i jest postrzegane jako skuteczne; jako czynnik sprawczy udzielenia głosu organizacjom mniejszym oraz jako odnoszące mniejsze sukcesy w angażowaniu sektora prywatnego. Partnerstwo w ramach funduszy strukturalnych jest podzielone w skali polityk krajowych. W konsekwencji doszło do pojawienia się społeczności interesów na poziomie lokalnym.
- **Adycjonalność projektów**. Niemal wszystkie projekty realizowane w ramach funduszy strukturalnych i poddane przeglądowi w ramach ewaluacji charakteryzowały się adycjonalności; fundusze wpłynęły na poprawę wykonalności, skali, jakości lub harmonogramów projektów. Odpowiedź na pytanie, czy wspierały naprawdę innowacyjne projekty, nie jest jasna (tzn. projekty, które nie uzyskałyby pomocy krajowej); jednakże niemal połowa rozmówców stwierdziła, że dziedzina projektu była nowa, a podejście do projektu – nowatorskie.
- **Skutki finansowania**. Fundusze strukturalne mogą wywierać efekt lewarowania (przez współfinansowanie) i mogą nawet „naginać" kierunek inicjatyw krajowych. W ramach programu *Highlands & Islands* nie stwierdzono jednak zbyt silnych dowodów takiego „naginania".
- **Skutki polityki**. Głównym wpływem okresu 2000–2006 była koncentracja równych szans i trwałego rozwoju. Wpływ ten jest jednak

głównie postrzegany jako raczej wzmacniający istniejącą praktykę (często w drodze projektów pilotowych) niż innowacyjny.

- **Skutki administracyjne**. Administracyjna wartość dodana funduszy strukturalnych może być ujemna, mając na uwadze zaangażowaną w ich wykorzystanie biurokrację. Z drugiej strony, podejście stosowane w ramach funduszy strukturalnych często wymaga bardziej przemyślanego uzasadnienia wszelkich projektów.

Poza dostarczaniem użytecznych informacji dla HISTP, to wszechstronne „rozpakowywanie" Wspólnotowej Wartości Dodanej przyczyniło się do przemyślenia tej koncepcji. Prace wykonane przez ewaluację skutkowały poinformowaniem innych zespołów badawczych o tematach realizowanych przez EPRC (Bachtler, Taylor 2003).

4.2. Proces rozpowszechniania

Jak stwierdzono, ewaluacja wymagała przyjęcia innowacyjnego podejścia do rozpowszechniania produkowanej wiedzy. Produkcja serii tematycznych Dokumentów Roboczych w toku całej ewaluacji oznacza, że wyniki każdego zadania badawczego mogły być rozpowszechniane niezwłocznie, bez potrzeby czekania na Projekt Raportu Końcowego czy sam Raport Końcowy. Dokument Roboczy był także postrzegany jako dokument dynamiczny, umożliwiający ewaluatorom zaangażowanie się w dialog ze stronami zainteresowanymi i interesariuszami w toku imprez upowszechniających, organizowanych w miarę udostępniania kolejnych Dokumentów. Ponadto zorganizowano taką imprezę w celu uruchomienia MTE. Zawartość Dokumentów Roboczych ewoluowała więc w toku badań, uwzględniając najnowsze materiały i analizy. Po zakończeniu badań ewaluacyjnych Dokumenty Robocze zostały zestawione jako rozdziały ostatecznego Raportu Technicznego, który towarzyszył Końcowemu MTE. Rozdziały te zawierają bardziej szczegółowy opis poszczególnych kwestii niż było to możliwe z powodu ograniczeń Raportu Głównego. Umożliwiło to interesariuszom i zainteresowanym czytelnikom dostęp – w ramach poszczególnych tytułów – do bazy dowodowej leżącej u podstaw syntezy, wniosków i zaleceń, zawartych w MTE.

5. WNIOSKI

Na poziomie podstawowym spodziewano się, że wyniki MTE będą wywierać istotny wpływ ze względu na rzeczywiste implikacje zasobowe. Ewaluacja przyczyniła się do alokacji funduszu rezerwy wykonawczej, co z kolei miało na Program wyraźny wpływ finansowy i związany z prestiżem. Ponadto wszechstronny zakres badań gwarantował ich użyteczność. Obejmowało to intensywne działania podejmowane przez ewaluatorów w celu ograniczenia rozbieżności w bazie danych HISTP przez

wszechstronne wizyty projektowe oraz ścisłą ocenę sposobu definiowania celów projektów. Jak stwierdzono, oczekuje się, że obecna generacja MTE opracuje wczesny obraz wpływów programu, a podejście przyjęte przez zespół ewaluacyjny oznaczało możliwość wyciągnięcia rzetelnych wniosków.

Podstawową funkcją MTE była poprawa wyników, jakości i planowania programu *Highlands & Islands*. MTE zapewnił wszechstronną analizę realizację programu, postępu finansowego, produktów oraz – tam, gdzie były dostępne dane – wyników. Zalecenia ewaluacji śródokresowej zostały szczegółowo uwzględnione przez specjalnie w tym celu utworzoną Grupę Przeglądową (Review Group). Grupa Przeglądowa doradzała Komitetowi Monitorującemu Program w sprawie planu działania związanego z realizacją zaleceń. Kluczowe znaczenie dla rekomendacji miał zakres korekt, wymaganych przez Program w celu uwzględnienia wszelkich zmian warunków społeczno-ekonomicznych w obszarze danego Programu. Uznano dalsze znaczenie ogólnych celów, priorytetów i działań Programu; panuje też ogólna zgoda co do konieczności korekt alokacji finansowych pomiędzy poszczególnymi działaniami. Zalecono korekty finansowe oraz alokację Rezerwy Wykonawczej, zatwierdzonej przez PMC i uzgodnionej z Komisją.

Fakt, że proces ewaluacji sprzyjał interakcji pomiędzy ewaluatorami a interesariuszami, przyczynił się także do wzrostu jego wpływu i użyteczności. MTE podjęto przy użyciu technik partycypacyjnych, które uwzględniały uczestnictwo partnerów w procesie diagnostycznym i rozwiązywania problemów. Wszechstronny zakres wizyt projektowych w czasie ewaluacji (możliwych dzięki hojnemu budżetowi) oraz innowacyjne podejście do rozpowszechniania (dzięki opracowaniu serii tematycznych Dokumentów Roboczych) skutkowały większą interakcją pomiędzy ewaluatorami a interesariuszami, podnosząc poczucie partnerstwa i własności.

Na koniec – w kategoriach poznawczych – nowoczesne podejście ewaluacji do oceny Wspólnotowej Wartości Dodanej skutkowało lepszym rozumieniem koncepcji w ramach HISTP. Podejście to wniosło także wkład do szerszej debaty na temat roli programów FS w Wielkiej Brytanii w okresie po 2006 roku. Ponadto odegrało rolę informacyjną wobec innych projektów badawczych, realizowanych w okresie późniejszym przez EPRC.

BIBLIOGRAFIA

Bachtler J., Taylor S (2003), *The Added Value of the Structural Funds: A Regional Perspective*, IQ-Net Thematic Paper on the Future of the Structural Funds, Glasgow, European Policies Research Centre, University of Strathclyde.

Davies S., Bachtler J., Gross F., Michie R., Vironen H., Yuill D. (2007), The Impact of Structural Funds Programmes in Scotland 1994–2006, *European Policy Research Papers* **60**, Glasgow, European Policies Research Centre, University of Strathclyde.

Raines P., Taylor S. (2002), Mid-term Evaluation of the 2000–06 Structural Fund Programmes, *IQ-Net Thematic Paper* **11(2)**, Glasgow, European Policies Research Centre, University of Strathclyde.

Scottish Executive (2003), *Mid-Term Evaluation of the Highlands & Islands Special Transitional Programme, Main Report (Final)*, July, produced by Fraser Associates in association with EPRC and the Rural Development Company.

CZĘŚĆ SZÓSTA

STRATEGICZNE OBSZARY TEMATYCZNE BADAŃ EWALUACYJNYCH

ROZDZIAŁ 6.1.

Bartosz Ledzion

WPROWADZENIE

Niniejsza, szósta część książki zawiera teksty o szerszym, eseistycznym charakterze. Sygnalizują one ograniczenia bieżących praktyk ewaluacyjnych, a także zarysowują wyzwania filozoficzne, koncepcyjne i metodologiczne w wybranych, głównych czterech polach tematycznych ewaluacji polityki spójności.

Pierwszy artykuł omawia wyzwania, jakie niesie za sobą obszar ewaluacji sektorowych programów i interwencji publicznych w zakresie infrastruktury technicznej i społecznej.

Kolejny tekst zwraca uwagę na specyfikę badań ewaluacyjnych w zakresie procesów innowacji, uważanych obecnie za jedne z najważniejszych czynników decydujących o produktywności i przewadze konkurencyjnej gospodarek.

W trzecim artykule autorka podejmuje próbę zdefiniowania głównych charakterystyk ewaluacji interwencji prowadzonych w obszarze rozwoju zasobów ludzkich.

Ostatni tekst przedstawia specyfikę jednego z ważniejszych obszarów strategicznych systemu ewaluacji Narodowego Planu Rozwoju (NPR) oraz Narodowych Strategicznych Ram Odniesienia (NSRO) tzw. *Rozwoju regionalnego i przestrzennego*.

ROZDZIAŁ 6.2.

Michał Wolański
Elżbieta Kozłowska

ROZWÓJ I MODERNIZACJA INFRASTRUKTURY

Cel 3 NSRO *Budowa i modernizacja infrastruktury technicznej i społecznej mającej podstawowe znaczenie dla wzrostu konkurencyjności Polski* jest osiągany na podstawie realizacji POIiŚ, RPO oraz PORPW. W ramach POIiŚ wsparcie koncentrować się będzie przede wszystkim na rozwoju i modernizacji podstawowej infrastruktury technicznej transportu. Towarzyszyć mu będą inwestycje mające na celu zapewnienie odpowiedniej jakości środowiska naturalnego w ramach poszanowania zasad zrównoważonego rozwoju.

Jak stanowią zapisy największego polskiego programu operacyjnego – *Programu Operacyjnego Infrastruktura i Środowisko*, członkostwo w Unii Europejskiej stanowi dla Polski zarówno szansę, jak i wyzwanie. Wraz z uczestnictwem we wspólnotowej polityce spójności istnieje szansa na zmniejszenie dystansu rozwojowego, jaki dzieli nasz kraj od „starych" państw członkowskich. Zapóźnienia rozwojowe mają przede wszystkim charakter luki infrastrukturalnej, która uniemożliwia optymalne wykorzystanie zasobów kraju oraz w dużym stopniu blokuje istniejący potencjał. Przezwyciężenie jej jest niezbędnym warunkiem wzrostu konkurencyjności i podniesienia aktywności inwestycyjnej.

W zakresie inwestycji infrastrukturalnych w okresie programowania 2007–2013 najwięcej środków skierowanych jest na inwestycje transportowe, związane przede wszystkim z infrastrukturą transportu drogowego, kolejowego i w znacznie mniejszym stopniu lotniczego oraz wodnego, a następnie inwestycje w ramach sektora Środowisko.

Celem niniejszego rozdziału jest przedstawienie najważniejszych wyzwań, typowych dla analizowanych dziedzin, związanych z kompleksową

analizą podejmowanych działań, przesłankami i poprawnością ich wyboru oraz skutecznością ich wpływu na realizację celów strategicznych.

1. SPECYFIKA EWALUACJI INWESTYCJI INFRASTRUKTURALNYCH SEKTORA TRANSPORTU

Największa część środków POIiŚ (prawie 20 z 28 mld EUR – MRR 2008a) przypada na trzy priorytety transportowe (szósty, siódmy i ósmy), zarządzane przez Ministerstwo Infrastruktury. Mają one za zadanie realizację celów strategicznych *Białej Księgi*, zawierającej europejską politykę transportową do 2010 roku (European Commission, 2001):

- wyważania podziału pracy przewozowej pomiędzy środkami transportu,
- eliminowania wąskich gardeł,
- stawiania użytkownika w centrum polityki transportowej,
- uwzględniania wyzwań globalizacji transportu.

Tak duża suma – rosnąca po dodaniu środków EFRR i FS do 26 mld EUR – wynika z faktu, że dobra sieć transportowa ma kluczowe znaczenie dla rozwoju gospodarki oraz pozwala istotnie obniżyć emisję zanieczyszczeń. Nie bez znaczenia jest również ogromna liczba wypadków na polskich drogach. W 2005 roku zginęło w ich wyniku 5,5 tysiąca osób, czyli aż 15 osób dziennie – najwięcej spośród wszystkich krajów Unii Europejskiej. *Per capita* zajmujemy niechlubne czwarte miejsce od końca w ramach EU-25, zaś w przeliczeniu na liczbę pasażerokilometrów – drugie od końca (Eurostat, 2007: 131).

W wymienionych priorytetach transportowych POIiŚ znaczną rolę odgrywają projekty bardzo duże, zawarte na liście indykatywnej. Dotyczą one budowy autostrad i dróg ekspresowych, przebudowy kilku ciągów kolejowych, portów morskich oraz portów lotniczych w największych miastach. Planowane są również linie kolejowe do trzech lotnisk, wdrażanie nowoczesnego i jednolitego w ramach Unii Europejskiej systemu sterowania ruchem kolejowym ERTMS, budowa drugiej linii metra w Warszawie oraz kilka punktowych inwestycji drogowych (MRR, 2008b).

Z punktu widzenia ewaluatora duża rola listy indykatywnej stanowi istotne wyzwanie, gdyż brakuje sformalizowanych kryteriów wyboru najtrafniejszych projektów. W latach 2004–2006 wątpliwości mógł budzić m.in. specyficzny wybór modernizowanych odcinków linii kolejowych – pierwotnie skupiano się na mało uczęszczanych trasach, takich jak Wrocław – Legnica – Zgorzelec/Bielawa Dolna. Obecnie wybierane projekty cechują się z pewnością większą trafnością. Przykładowo, w transporcie kolejowym z reguły łączą one największe aglomeracje Polski oraz są położone w obrębie tzw. międzynarodowych korytarzy

transportowych TEN-T, których modernizacja do określonych parametrów wynika ze zobowiązań międzynarodowych. Wciąż jednak można się spierać, czy słuszna jest np. modernizacja trasy Warszawa – Radom, a nie na przykład Gdynia – Poznań. Dodatkowych problemów w ocenie przysparza niestabilny kontekst strategiczny. Planowana – chociaż nie pewna – jest budowa linii dużych prędkości Warszawa – Łódź – Wrocław/Poznań, tymczasem remontuje się linie z Warszawy do Łodzi oraz z Wrocławia do Poznania. Podobne działania dotyczą dróg krajowych, na których przeprowadzało się uciążliwe remonty (np. odcinek DK nr 4 Bolesławiec – Zgorzelec), chociaż w tym przypadku dało się przewidzieć, że za kilka lat większość ruchu zostanie skierowana na równoległą autostradę.

Przy ocenie wyboru projektów ewaluator powinien zweryfikować, czy dokonywano konsultacji z potencjalnymi użytkownikami – zwłaszcza w przypadku transportu kolejowego, gdzie współdziałanie zarządcy infrastruktury oraz przewoźników i organizatorów przewozów gra kluczową rolę, o czym będzie mowa w dalszej części tekstu.

Jeszcze większe dyskusje budzi alokacja budżetu pomiędzy różne środki transportu – czy nie efektywniej byłoby położyć większy nacisk na kolej, która generuje znacznie mniej wypadków? A może wręcz przeciwnie – większe korzyści przyniosłaby sieć autostrad? System wyboru projektów, oparty na liście indykatywnej, nie ułatwia znalezienia odpowiedzi na tak postawione pytania.

Istniejące uzasadnienia – zawarte w omawianym dokumencie (MRR, 2008b) – są bardzo ogólnikowe, a czasami wręcz nieadekwatne. Przykładowo Euro 2012, nie jest dobrym powodem dla budowy wschodniego i zachodniego odcinka drugiej linii warszawskiego metra, gdyż zostaną oddane do użytku kilka lat po mistrzostwach. Oczywiście błędna argumentacja nie oznacza, że ta inwestycja nie jest potrzebna.

Przy ocenie trafności wyboru niezbyt pomocne jest porównywanie przeprowadzonych analiz ekonomicznych, gdyż te mogą być wykonywane przy różnych założeniach (np. dotyczących jednostkowych kosztów zewnętrznych transportu czy ogólnych prognoz przewozowych dla wariantu referencyjnego). Typowym zjawiskiem w sektorze infrastruktury jest również niedoszacowywanie kosztów, co także zawyża współczynniki efektywności.

Oczywiście trudno wymagać od ewaluatora powtórzenia analiz ekonomicznych w zunifikowany sposób – byłoby to bardzo czasochłonne i kosztowne, a pozyskane dane – jedynie częściowo użyteczne. Dobrym rozwiązaniem na etapie ewaluacji *ex post* jest ocena punktowa różnych kryteriów oraz wywiady z interesariuszami, jednak docelowo problem może być zredukowany jedynie dzięki zunifikowanej analizie ekonomicznej na etapie ubiegania się o dofinansowanie.

Przy konstrukcji listy indykatywnej niepokój może budzić również zaniedbanie dwóch trendów, widocznych już od wielu lat w krajach

„starej" Unii, związanych w zasadzie z integracją środków transportu – zwłaszcza publicznego.

Po pierwsze – chociaż na liście indykatywnej znajdują się projekty połączeń kolejowych pomiędzy aglomeracjami i portami lotniczymi, to jednak zaniedbuje się kwestie związane z tworzeniem powiązań transportowych pomiędzy lotniskami i regionami. Tymczasem tendencją europejską jest zapewnianie bezpośrednich połączeń dalekobieżnych, czego przykładem może być chociażby nowo wybudowana stacja pociągów dużych prędkości pod terminalem Charles de Gaulle w Paryżu, czy w rejonie lotniska we Frankfurcie nad Menem. Dotyczy to również mniejszych miast, takich jak Lyon, Kolonia, czy Duesseldorf. Trend ten wynika z faktu, że zrównoważony rozwój regionalny wymaga zapewnienia mniejszym miastom dobrych połączeń międzynarodowych, natomiast nawet budowa w nich lotnisk nie pozwoli zapewnić dobrej siatki połączeń, satysfakcjonującej klienta biznesowego. Tymczasem krótkie loty do najbliższego hubu (portu przesiadkowego) są bardzo kłopotliwe, gdyż zmniejszają możliwą liczbę startów i lądowań dla połączeń dalszego zasięgu. Stąd też w Niemczech, Francji, Belgii, czy Czechach linie lotnicze oferują nawet przejazd pociągiem na podstawie swojego biletu. Dotyczy to np. połączenia kolejowego Paryż–Bruksela.

Po drugie, wciąż brakuje projektów, które w sposób kompleksowy umożliwiałyby tworzenie systemów kolei miejskich, kluczowych dla zrównoważonego rozwoju przestrzennego polskich aglomeracji. Rozwiązania takie są podstawą transportu miejskiego w większości dużych metropolii europejskich, gdyż cechują się bardzo dużą efektywnością ekonomiczną, dzięki względnie wysokim prędkościom oraz wykorzystaniu istniejącej infrastruktury liniowej (chociaż konieczne są inwestycje, związane np. z budową nowych stacji). Systemy kolei miejskich funkcjonują m.in. w największych miastach Niemiec, Paryżu, czy Berlinie, natomiast dobrymi przykładami współczesnego dostosowania węzła kolejowego do potrzeb komunikacji aglomeracyjnej mogą być Madryt (por. Carillo 2007) i Londyn (przeprowadzany aktualnie projekt Overground).

Ocena efektywności i trwałości jest największym wyzwaniem w przypadku inwestycji kolejowej oraz w transporcie miejskim, gdyż – w przeciwieństwie do budowy dróg – nie tworzą one kompleksowej oferty dla użytkownika końcowego (pasażera). Powstaje ona z połączenia działań zarządcy infrastruktury, organizatorów przewozów[1]

---

[1] Organizatorem przewozów są samorządy miejskie (w transporcie miejskim) i wojewódzkie (regionalny transport kolejowy), a także minister właściwy do spraw transportu (międzywojewódzki i międzynarodowy transport kolejowy). W transporcie miejskim samorządy często powołują wyspecjalizowane zarządy transportu miejskiego, jak np. Zarząd Transportu Miejskiego w Warszawie. Powołanie podobnych organów postuluje się w transporcie regionalnym – kolejowym i autobusowym (por. Wyszomirski et al. 2007).

i przewoźników[2]. Dodatkowo, w przypadku kolei istnieje rozdzielenie pomiędzy właścicielem infrastruktury liniowej (PKP PLK SA) oraz budynków dworcowych (PKP SA).

Na efektywność kluczowy wpływ ma wybór optymalnego zakresu inwestycji, np. pod względem przepustowości i innych parametrów technicznych. Jak w każdej innej branży, wymaga to przeprowadzenia w studium wykonalności analiz wielu możliwych wariantów inwestycji („ustawianie" wariantów może prowadzić do poważnej nieefektywności), a następnie ich obiektywnego porównania, najlepiej z wykorzystaniem współczynników efektywności ekonomicznej.

Niestety, beneficjenci bardzo często traktują studium wykonalności nie tylko jako uzasadnienie z góry wybranej inwestycji, ale również z góry wybranego jej zakresu, co może się przejawiać w wybraniu skazanych na porażkę wariantów alternatywnych.

Wybór wariantu wymaga wyważenia pomiędzy uwzględnianiem efektów w skali całego korytarza a zróżnicowaniem potrzeb na poszczególnych odcinkach. Przykładowo, na linii kolejowej Warszawa–Gdynia w rejonie stolicy i Trójmiasta pojawia się niezależny od dalekobieżnego ruch aglomeracyjny, co nakazuje oddzielne, niezależne analizy wariantów inwestycji na tych odcinkach. Negatywne wyniki efektywności ekonomicznej modernizacji fragmentu linii mogą być jednak złudne, gdyż jej zaniechanie spowoduje brak konkurencyjności całego ciągu.

Przy wyborze wariantu inwestycji kluczowe jest przyjęcie właściwych założeń (np. dotyczących popytu) oraz dobra harmonizacja z planami przewoźników. Problemem są przede wszystkim kwestie taborowe – wiele linii modernizowanych jest do prędkości 160 km/h, co stanowi standard dla korytarzy objętych umowami międzynarodowymi. Niestety, prędkość ta jest niedostępna dla większości taboru w Polsce.

Ewaluator powinien zwracać tutaj uwagę na koordynację działań, prowadzonych w ramach innych funduszy unijnych (Sektorowy Program Operacyjny – Transport, ZPORR) oraz ze środków krajowych (w tym środki na zakup taboru, dotacje do przewozów kolejowych). Dotyczy to także budowy spójnych sieci komunikacji miejskiej.

Z punktu widzenia trwałości inwestycji bardzo istotne jest jej odpowiednie utrzymanie i konserwacja (np. regularne szlifowanie torów), lecz dużą rolę odgrywają również kwestie organizacyjne, takie jak cennik oraz zasady dostępu do infrastruktury drogowej i lotniczej, a także – w transporcie zbiorowym – odpowiednia oferta przewozowa.

---

[2] Przewoźnikiem może być np. spółka PKP Przewozy Regionalne, Koleje Mazowieckie, Metro Warszawskie, Miejskie Zakłady Autobusowe. Warto podkreślić, że istnieje tendencja do coraz częstszego zlecania przewozów firmom prywatnym. Przykładowo, w Warszawie 25% przewozów na zlecenie ZTM wykonują firmy prywatne, wyłonione w toku przetargu (zwykli pasażerowie praktycznie nie odróżniają ich autobusów). Podobne rozwiązanie zastosowano w kolejowych przewozach regionalnych w województwie kujawsko-pomorskim, gdzie zorganizowany przez samorząd regionalny przetarg na część linii wygrało prywatne konsorcjum firm PCC Rail oraz Arriva.

Niestety, również w tym względzie istnieją istotne zagrożenia dla realizowanych projektów. Przykładowo, aktualny cennik PKP PLK SA – zatwierdzony przez Urząd Transportu Kolejowego – wymaga od przewoźnika wnoszenia ponaddwukrotnie wyższych opłat za dostęp do infrastruktury, jeśli na danej linii można osiągnąć prędkość przekraczającą 120 km/h – niezależnie od tego, czy przewoźnik wykorzystuje te parametry (PKP PLK 2007)[3]. W konsekwencji – zwłaszcza w przewozach regionalnych – nowe inwestycje mogą powodować negatywne skutki dla przewoźnika, który płaci za efekty inwestycji, których *de facto* nie konsumuje. W sposób oczywisty wpływa to na ograniczenie oferty przewozowej.

Co więcej, niedofinansowanie eksploatacji infrastruktury z budżetu państwa powoduje, że stawki dostępu do niej – szczególnie w przypadku transportu towarowego – należą do najwyższych w Europie. Zmniejsza to pozytywne oddziaływanie inwestycji, gdyż zrażeni kosztami przedsiębiorcy wybierają droższy społecznie transport drogowy, zaś przepustowość linii kolejowej – nawet zmodernizowanej – pozostaje niewykorzystana. Zjawisko to potęgują kłopoty z dostępem do bocznic – zmonopolizowanych przez PKP Cargo – i brak odpowiednich terminali logistycznych. Może to nasuwać wniosek, że przeprowadzane projekty nie rozwiązują problemów braku konkurencyjności kolei, gdyż *de facto* nie dotyczą one jakości infrastruktury liniowej. Do podobnych wniosków można dojść w przypadku przewozów pasażerskich, chociażby obserwując stan dworców oraz ofertę przewozową.

Kończąc rozważania na temat infrastruktury, należy wspomnieć o typowych problemach kadrowych beneficjentów (duża rotacja pracowników i liczba wakatów), które bardzo często utrudniają przeprowadzanie badań ewaluacyjnych oraz pozyskanie wiarygodnych danych, a nawet dokumentów źródłowych. Obserwując lukę pokoleniową powstającą w niektórych jednostkach, należy przewidywać, że problem ten w kolejnych latach będzie miał jeszcze większe znaczenie.

Wszystkie wspomniane trudności i wyzwania wskazują na konieczność przeprowadzania badań ewaluacyjnych w sektorze infrastrukturalnym oraz wyciągania z nich właściwych wniosków przez decydentów.

2. SPECYFIKA EWALUACJI INWESTYCJI INFRASTRUKTURALNYCH SEKTORA ŚRODOWISKO

Sektor ochrony środowiska jest jednym z najszerszych i kompleksowych obszarów, w którym zmiany zależą od wielu trudnych do eliminacji

[3] Problem ten jest istotny dla ewaluatora również przy ocenie jakości przeprowadzanej analizy finansowej; ze względu na zmienność cennika PKP PLK SA wykonawcy studiów mogli nie przewidzieć wzrostu przychodów wskutek realizacji projektu. Może to stawiać pod znakiem zapytania nieefektywność finansową projektów.

czynników zewnętrznych oraz powiązane są z dużą niepewnością procesów zachodzących w środowisku naturalnym. Bogactwo i kompleksowość polityki dotyczącej tego obszaru odzwierciedlone są w 6. Programie Działań na Rzecz Środowiska. Dokument ten wyznacza cele i ramy dla polityki Unii Europejskiej do 2012 roku. Priorytetami 6. Programu Działań na Rzecz Środowiska są:

- zmiany klimatyczne,
- ochrona przyrody i bioróżnorodności,
- zdrowie i środowisko naturalne,
- zrównoważone wykorzystanie zasobów naturalnych i gospodarka odpadami.

Zgodnie zapisami dokumentów programowych UE, celem wsparcia w zakresie sektora Środowisko jest przede wszystkim poprawa jakości środowiska naturalnego. Cel ten jest także realizowany przez wypełnianie zobowiązań zawartych w Traktacie Akcesyjnym w obszarze „Środowisko". Celem projektów realizowanych w ramach sektora Środowisko nie może być tylko osiąganie efektów gospodarczych lub społecznych,

Dla stworzenia dogodnych warunków lokalizacji inwestycji firmy w danym układzie lokalnym, niezbędne jest, aby oprócz dostępności transportowej, zapewnić odpowiednie uzbrojenie terenów inwestycyjnych i dobrze rozwiniętą sieć infrastruktury technicznej. Ponadto, odpowiednia jakość środowiska i warunki zdrowotne mieszkańców, to elementy, które mogą wpłynąć na decyzję o lokalizacji firmy w danym miejscu. Aby w danym regionie istniały korzystne warunki dla osiedlania się wysoko wykwalifikowanych specjalistów, inwestycjom transportowym muszą towarzyszyć inwestycje w infrastrukturę ochrony środowiska. Efektywne ekonomicznie wsparcie publiczne dla inwestycji sektora Środowisko w ramach funduszy unijnych może przyczynić się także do zmniejszenia kosztów środowiskowych, co wpłynie na zwiększenie atrakcyjności inwestycyjnej regionu.

Brak podstawowej infrastruktury ochrony środowiska może stanowić barierę rozwojową dla sektora MSP. Brak możliwości podłączenia firmy do wodociągu i odprowadzenia ścieków do kanalizacji powoduje wzrost kosztów utrzymania firmy, lub też uniemożliwia jej lokalizację w danym miejscu. Ponadto, braki w infrastrukturze ochrony środowiska mogą wpłynąć na możliwość kooperacji z dużymi firmami, zwłaszcza zagranicznymi, które posiadają certyfikat ISO 14.001 i w związku z tym będą stawiały warunki wypełnienia wymogów ekologicznych przez swoich kooperantów. Jeżeli spełnienie tych wymogów nie będzie możliwe, taka współpraca może nie dojść do skutku.

jednak poprawa jakości środowiska naturalnego może pociągać za sobą powstanie dodanych ubocznych efektów społeczno-gospodarczych.

Ochrona środowiska naturalnego (osiągana także pośrednio przez wypełnianie zapisów Traktatu Akcesyjnego) jest więc podstawowym celem projektów realizowanych w ramach sektora Środowisko, lecz podczas ewaluacji tego typu przedsięwzięć nie należy zapominać o ich ekonomiczności, a także o uchwyceniu dodatkowych efektów, jakimi będą np. efekty gospodarcze i społeczne wiążące się z realizacją danej inwestycji.

Przy ocenie efektów środowiskowych oraz społeczno-gospodarczych danej interwencji należy wziąć pod uwagę następujące cechy tego sektora:

- wysoką złożoność; zachodzące procesy są silnie ze sobą powiązane, charakteryzują się wysokim stopniem kompleksowości;
- niepewność procesów zachodzących w środowisku (trudność oszacowania reakcji);
- konieczność interdyscyplinarnego podejścia do problemów środowiska;
- konfliktowość celów – w zależności od przyjętej perspektywy;
- różną skalę zjawisk (od lokalnych do globalnych problemów);
- efekty uboczne (które mogą zmodyfikować kierunek oddziaływania realizowanej interwencji;
- zagrożenie rozwiązania problemu przez jego przesunięcie.

Podejmując próbę oszacowania efektów interwencji sektora Środowisko, należy pamiętać, że niektóre efekty interwencji mogą być widoczne:

- w krótkim okresie (np. obserwowalna w krótkiej perspektywie poprawa jakości powietrza, uzyskana przez redukcję zawartości ołowiu w wyniku zastąpienia benzyny zawierającej ołów benzyną bezołowiową);
- w długim okresie (takim długofalowym procesem jest proces samooczyszczania się wód, uruchamiany na skutek zmniejszania dopływu ładunku zanieczyszczeń, np. azotanów do wody).

W niektórych przypadkach, ewaluacja wykonana w zbyt krótkim czasie po zakończeniu projektu lub programu może błędnie wykazać jego nieefektywność.

Przy analizie efektów i oddziaływania interwencji sektora Środowisko poważnym wyzwaniem badawczym jest problem z uchwyceniem, w jakim stopniu dane efekty są rezultatem badanej interwencji, a w jakim stopniu na zmiany otoczenia wpłynęły inne procesy i czynniki zewnętrzne (np. w ostatnich latach poprawiła się jakość wód powierzchniowych, ale w jakim stopniu jest to efekt realizowanych projektów, w jakim stopniu zmiany technologii, a w jakim stopniu przemian gospodarczych?). W związku z tym ewaluator powinien dokładnie przeanalizować procesy zachodzące w otoczeniu projektu, by uniknąć błędnego

wnioskowania oraz np. dowodzenia sytuacji, w której po uruchomieniu oczyszczalni ścieków, jakość wód powierzchniowych pogorszyła się, a faktycznym powodem zmiany jakości wód nie jest analizowana inwestycja, lecz wzrost intensyfikacji produkcji rolnej i związane z tym wpływy powierzchniowe bogate w składniki minimalne, pochodzące ze stosowanych nawozów.

Kolejnym wyzwaniem dla ewaluatora jest uchwycenie zmian w klasyfikacji jakości środowiska, jaka nastąpiła od roku 2004 w związku z akcesją Polski do UE. Jako przykład może służyć tutaj zmiana metodologii badania i oceny oraz klasyfikacji jakości wód powierzchniowych. W związku z tym, dane z lat 1999–2003 nie mogą być porównane z najnowszymi danymi.

Mając na uwadze powyższe czynniki oraz cykl i terminy prowadzenia ewaluacji, w większości przypadków uchwycenie oddziaływania inwestycji z uwagi na zbyt krótki czas od momentu zakończenia realizacji projektu będzie niemożliwe; w takiej sytuacji badanie i analizę należy ograniczyć tylko do szacowania rezultatów określonych, np. % redukcji wskaźnika zanieczyszczeń.

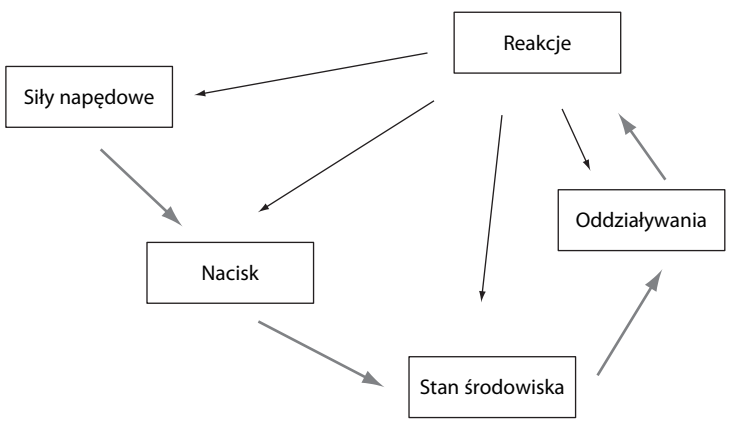
Ponadto, poszukując efektów i oddziaływania projektów sektora Środowisko, nie należy ograniczać zakresu analiz do poszukiwania zmian w jakości środowiska, ale także należy obszar analizy rozszerzyć o poszukiwanie efektów gospodarczych i społecznych, jakie mogą być wynikiem zrealizowanego projektu. Szczególnie istotnym czynnikiem, który powinien zostać zawarty w badaniu, jest spojrzenie na cele projektu z perspektywy celów strategicznych Unii Europejskiej, kraju oraz regionu, a także zidentyfikowanych potrzeb lokalnych. Obecnie, w polskich realiach przy ewaluacji sektora Środowisko ważną częścią analiz jest badanie stopnia wypełniania przez projekt (lub projekty) postanowień Traktatu Akcesyjnego.

W ewaluacji sektora Środowisko istotne jest interdyscyplinarne podejście do analizy zagadnień, a w procesie analiz należy wykorzystywać szeroki wachlarz metod i danych z różnych źródeł, a następnie dokonywać ich triangulacji. Szczególną uwagę trzeba zwrócić na możliwość powstania oddziaływań skumulowanych, a także analizowanie, czy dana interwencja, rozwiązując problem w danym obszarze, nie powoduje przesunięcia problemu na inne obszary (np. realizacja inwestycji mających na oczyszczanie ścieków, bez należytego rozwiązania problemy gospodarki osadowej, spowoduje kolejny problem związany z zagospodarowaniem lub utylizacją odpadów, które powstały w wyniku realizacji ewaluowanych projektów).

Z uwagi na powyższe wyzwania ewaluacji sektora Środowisko, źródła metodologiczne UE do ewaluacji polityk środowiskowych (European Commission, 2007) polecają model DPSIR. Model ten można wykorzystać do analizy ścieżki przyczynowo-skutkowej, śledzącej wpływ człowieka na środowisko, który ułatwia kompleksową i systemową ocenę

rezultatów i oddziaływania i zależności pomiędzy poszczególnymi elementami wpływającymi na stan środowiska.

W modelu DPSIR (Drivers – Pressure – State – Impact – Responses) siły napędowe, takie jak: zmiany populacji, wzrost ekonomiczny, urbanizacja, intensyfikacja rolnictwa, powodują emisję zanieczyszczeń oraz inne presje, które wpływają na stan środowiska. W rezultacie zmiana stanu środowiska oddziałuje na zdrowie człowieka i kondycje ekosystemów. Efektem tych zmian jest potrzeba reakcji, mającej na celu ograniczenie, niekorzystnych zmian, które zaszły w systemie. Reakcje mogą być kierowane zarówno na: siły napędowe (by zachorować przyczynę, np. przez zamknięcie uciążliwej fabryki w regionie); jak i na presję – naciski (np. przez założenie filtrów na kominy); lub pośrednio na stan środowiska i jego oddziaływanie (np. przez próby eliminacji danego zanieczyszczenia ze środowiska; wprowadzenie systemów chroniących oddziaływanie na człowieka, czego przykładem są ekrany akustyczne).



Rysunek 1. Model DPSIR

Źródło: European Commission, 2007.

Ewaluacja sektora Środowisko powinna skupiać się na problemie, jak efektywne są połączenia między reakcjami (a więc inwestycjami mającymi na ochronę lub poprawę stanu środowiska) a osiągniętymi rezultatami (osiągniętymi pośrednio lub bezpośrednio jako poprawa lub zachowanie stanu środowiska naturalnego).

Polityka w sektorze środowiska powinna skupiać się bardziej na oddziaływaniu na siły napędowe i naciski (eliminacja przyczyn), niż na bezpośrednim wpływie na stan lub oddziaływania. W niektórych przypadkach działania mające na celu zmianę stanu lub oddziaływań (np. wapnowanie zakwaszonej gleby, lub reintrodukcja gatunków) są jednak także konieczne.

Inwestycje w zakresie gospodarki wodno-ściekowej to przede wszystkim eliminacja „nacisków" (redukcje zanieczyszczeń trafiających do

środowiska), jednakże realizacja tych inwestycji zgodnie z zasadą „zanieczyszczający płaci", to już „reakcja" na „siły napędowe".

W przypadku inwestycji w zakresie gospodarki wodno-ściekowej przedsiębiorstw, możemy mieć do czynienia zarówno z inwestycjami mającymi na celu eliminację nacisków (przez budowę urządzeń mających na celu redukcję zanieczyszczeń) lub wpływających na siły napędowe (np. przez zmiany technologii na bardziej zasobooszczędne i zamknięcie obiegów wody).

Podsumowując, należy stwierdzić, iż z uwagi na kompleksowość, złożoność oraz niepewność procesów zachodzących w środowisku (analogiczna interwencja w różnych ekosystemach może przynieść różne skutki), wyzwaniem dla ewaluatora jest oszacowanie spowodowanej przez przedsięwzięcie reakcji w środowisku. Podobna złożoność dotyczy systemów transportowych, w których często dopiero współdziałanie wielu partnerów tworzy kompleksowy produkt.

W sektorze transportu ogromne znaczenie ma optymalizacja i trafność zaspokajania potrzeb pasażerów przy użyciu nowoczesnych, kompleksowych rozwiązań, dających trwały efekt oraz rozumienie podejmowanych działań technicznych jedynie jako cele. Zadania tego nie ułatwia alokacja dużej części środków za pomocą list indykatywnych, a także powierzchowność przeprowadzanych analiz kosztów i korzyści.

W sektorze środowiska ważne jest, by każdą ocenę rozpatrywać dla konkretnego projektu lub programu, który umiejscowiony jest w danym otoczeniu. Należy zwrócić uwagę na różną skalę zjawisk zachodzących w przyrodzie (od lokalnych do globalnych). Do analizy problemów środowiska należy więc podchodzić na różnych poziomach: od globalnych (np. zmiany klimatu) do lokalnych (zbiórka odpadów), a także konfiktowości celów w zależności od przyjętej perspektywy czy specjalizacji eksperta (np. spór co lepiej zabezpiecza przed powodzią: mała retencja vs duże zbiorniki wodne)

Należy także pamiętać o zachowaniu głównych zasad sektora Środowisko, a więc zasadzie przezorności, subsydiarności oraz „zanieczyszczający płaci". Ponadto – o analizie konsekwencji niewypełnienia zobowiązań wobec UE w przypadku braku realizacji danego projektu czy programu. Istotne jest także identyfikowanie efektów ubocznych, zarówno tych pozytywnych (efekty wzmocnienia oddziaływania czy efekty gospodarcze), jak i negatywnych (np. rozwiązanie problemu przez jego przesunięcie).

W obu przypadkach konieczne jest zastosowanie interdyscyplinarnego podejścia, zróżnicowanych metod badawczych oraz źródeł danych (ważne są nie tylko dane statystyczne, lecz także obserwowalne trendy, opinie specjalistów i ekspertów z różnych dziedzin). Przy ewaluacji projektów i programów w ramach sektorów infrastruktury i środowiska istotna jest więc triangulacja danych oraz analiza możliwości oddziaływań skumulowanych.

BIBLIOGRAFIA

Carillo A. (2007), Cercanias – metamorfoza kolei, *Biuletyn Komunikacji Miejskiej* **98**, 62–63.

European Commission (2001), *White Paper, European transport policy for 2010: time to decide*, Luxembourg.

Eurostat (2007), *Panorama of Transport*, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-DA-07-001/EN/KS-DA-07-001-EN.PDF (3.06.2007).

Ministerstwo Rozwoju Regionalnego (2008a), *Program Operacyjny Infrastruktura i Środowisko*, http://www.mrr.gov.pl/ProgramyOperacyjne+2007–2013/Infrastruktura+i+Srodowisko/, (3.06.2008).

Ministerstwo Rozwoju Regionalnego (2008b), *Lista projektów indywidualnych dla Programu Operacyjnego Infrastruktura i Środowisko 2007–2013*, http://www.mrr.gov.pl/NR/rdonlyres/36237A56-0A3E-428D-AE80-0847A9FA703E/43457/080201_Nowa_lista_projektow_indywidualnych_POIiS.pdf, (3.06.2008).

PKP PLK (2008), *Cennik stawek jednostkowych opłat za korzystanie z infrastruktury kolejowej zarządzanej przez PKP Polskie Linie Kolejowe SA obowiązujący od 9 grudnia 2007*, http://www.plk-sa.pl/fileadmin/pdf/reg0708_zm5/cennik_2007–08.pdf, (3.06.2008).

Wyszomirski O. et al. (2008), *Rola, zadania i zasady funkcjonowania regionalnego pasażerskiego transportu regionalnego w latach 2008–2015 w świetle przewidywanych zmian organizacyjnych, prawnych i społecznych*, Warszawa, PIGTSiS.

ROZDZIAŁ 6.3.

Michał Miedziński

WYBRANE ZAGADNIENIA EWALUACJI POLITYKI INNOWACYJNEJ

1. WPROWADZENIE

Wiedza i potencjał innowacyjny uważane są obecnie za jedne z najważniejszych czynników decydujących o jakości wzrostu gospodarczego, produktywności i przewadze konkurencyjnej gospodarek. Nie jest zatem zaskakujące, iż podobnie jak w większości państw UE i OECD, wzrost innowacyjności gospodarki znajduje się wśród głównych priorytetów polityki gospodarczej Polski. Wsparcie procesów innowacji staje się również jedną z ważniejszych pozycji planowanych wydatków z funduszy strukturalnych na lata 2007–2013, szczególnie w ramach Programu Operacyjnego *Innowacyjna Gospodarka* (z budżetem ponad 9,7 mld EUR, w tym ponad 8,2 mld EUR z Europejskiego Funduszu Rozwoju Regionalnego) oraz 16 regionalnych Programów Operacyjnych[1]. W tym kontekście rola ewaluacji działań państwa na rzecz podniesienia innowacyjności staje się niebagatelna.

Wprowadzając temat ewaluacji polityki innowacyjnej, należy podkreślić, iż obejmuje ona wiele zróżnicowanych inicjatyw, wykraczających

[1] Około jednej czwartej wsparcia wspólnotowego w ramach 16 regionalnych programow operacyjnych zostanie przeznaczonych na działania finansujace B+R, nowe technologie i innowacje. Finansowanie wyniesie od 160 mln EUR w Podlaskiem do 430 mln EUR w Mazowieckiem (por. Walendowski 2008).

poza zakres prac pojedynczych ministerstw czy agencji. Poza „flagowymi" działaniami wsparcia innowacji, inicjowanymi i promowanymi głównie przez Ministerstwo Gospodarki i PARP, na innowacyjność gospodarki oddziaływają między innymi polityka naukowa i badawcza, programy edukacyjne i szkoleniowe, polityka regionalna czy polityka fiskalna państwa (np. ulgi podatkowe dla firmy prowadzących B+R). Ogólny obraz wpływu interwencji publicznych na innowacyjność jest złożony i wymaga zrozumienia wielu, często nieskoordynowanych, równolegle prowadzonych działań. Ewaluator oceniający politykę innowacyjną państwa, dany program czy poszczególne działanie staje przed wieloma wyzwaniami związanymi z kompleksowością i wielowymiarowością tematu. Jednym z ważniejszych wyzwań jest zrozumienie specyfiki procesów innowacji i umieszczenie ewaluowanych działań w szerszym kontekście (np. regionu, sektora czy całego systemu innowacji). Innym, niezwykle istotnym wyzwaniem jest atrybucja efektów interwencji publicznej, szczególnie jeżeli chodzi o ocenę efektów polityki na gospodarkę i społeczeństwo.

Niniejszy rozdział jest wprowadzeniem do tematyki ewaluacji polityki innowacyjnej, przedstawiającym główne wyzwania stojące przed ewaluatorami. Część pierwsza wprowadza definicje, główne koncepcje i współczesne podejścia do analizy procesów innowacji. Część druga skupia się na zagadnieniach dotyczących logiki i zakresu polityki innowacji. Celem części trzeciej jest zarysowanie wybranych wyzwań stających przed ewaluatorami. Rozdział zamykają wnioski i podsumowanie. Rozdział zawiera liczne odniesienia do ważnych głosów w dyskusji na temat specyfiki i wyzwań ewaluacji innowacji, jak również do licznych raportów i opracowań z konkretnymi przykładami metod ewaluacji i ich zastosowań.

2. KONTEKST ANALIZY PROCESÓW INNOWACJI

2.1. Definicje innowacji i firmy innowacyjnej

Najczęściej używana definicja innowacji została rozpowszechniona przez tak zwany podręcznik z Oslo przygotowany wspólnie przez OECD i EUROSTAT (OECD 2005b). Podręcznik definiuje innowację jako wprowadzony na rynek nowy lub znacząco ulepszony produkt lub usługę, proces, technikę marketingową czy nową formę organizacji. W tym kontekście firma innowacyjna to firma, która (w czasie objętym badaniem) wdrożyła innowację. W zależności od potrzeb analizy definicje innowacji i firmy innowacyjnej można uściślać, biorąc pod uwagę przykładowo typ innowacji czy nowatorstwo rozwiązania (np. unikatowość w skali regionu, państwa, świata).

W literaturze rozróżnia się wiele typów innowacji. Podręcznik z Oslo (OECD 2005b) definiuje cztery główne typy innowacji: produktową, procesową, marketingową i organizacyjną. W analizach za innowacje uznaje się zatem nie tylko nowy produkt lub usługę, lecz także zmiany organizacyjne zarówno w zakresie zarządzania, produkcji, jak i rozpowszechniania produktów na rynku. Autorzy rozróżniają również innowacje radykalne i ewolucyjne (czy też inkrementalne). Innowacje radykalne są wprowadzane niezwykle rzadko, natomiast innowacje ewolucyjne, czyli wprowadzenie stopniowych ulepszeń w produkcie, usłudze czy też organizacji produkcji są bardziej powszechne. W obliczu złożoności tematyki, zarysowanie jasnych ram koncepcyjnych i uzgodnienie definicji innowacji w dokumentach strategicznych i programach wsparcia innowacyjności ma znaczenie fundamentalne dla procesu ewaluacji.

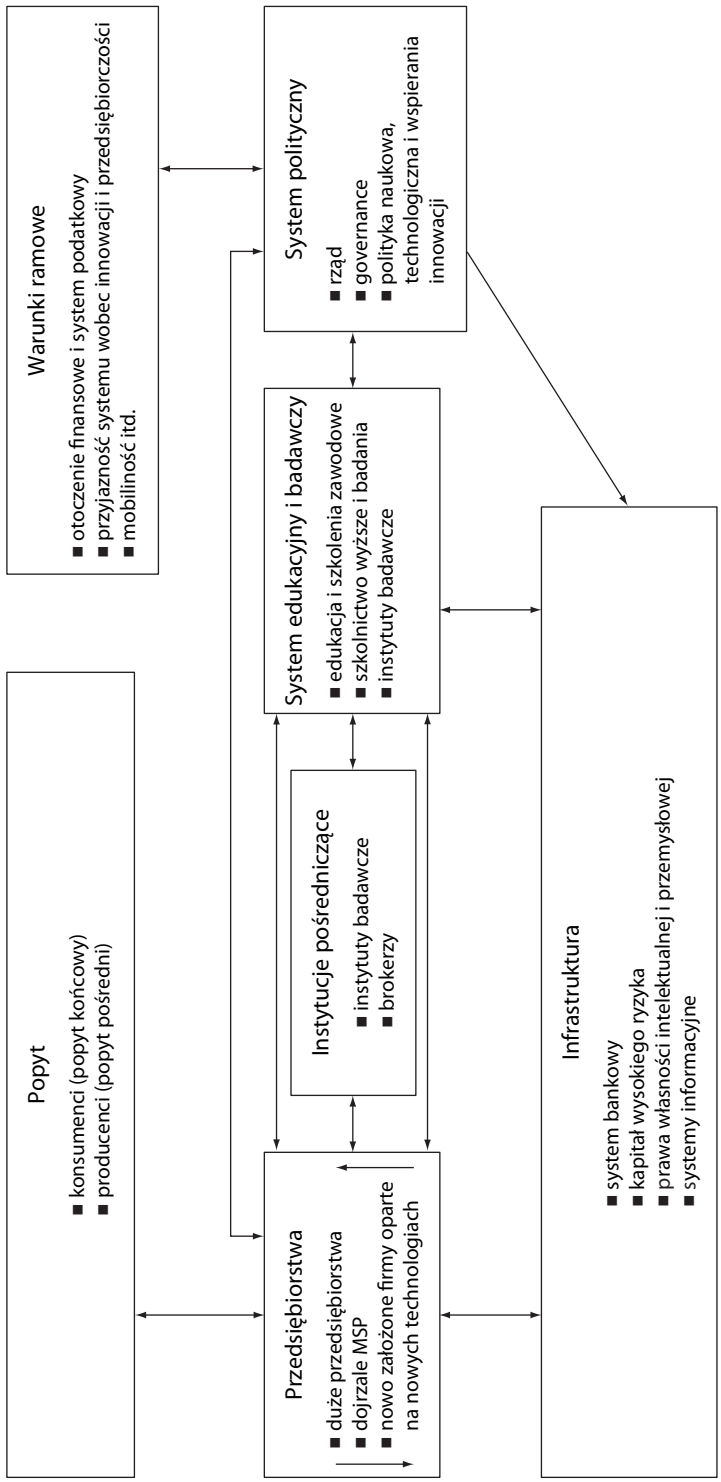
2.2. Zmiany podejścia do analizy procesu innowacji

Pod koniec lat 80. i w latach 90. nastąpiła fundamentalna zmiana rozumienia procesów innowacji i odejście od tzw. podejścia liniowego na rzecz ujęcia systemowego innowacji (por. Freeman 1987, Lundvall 1992, Nelson 1993, Smith 2000). Tradycyjnie proces innowacji pojmowany był jako przyczynowy proces liniowy, nazywany często podażowym lub zdeterminowanym przez naukę (*science-push model*)[2].

Podejście systemowe odstępuje od przyczynowości liniowej, podkreślając złożony i interaktywny charakter procesu innowacji. Freeman (1987) definiuje system innowacji jako „sieć instytucji sektora prywatnego i publicznego, których działania i wzajemne zależności prowadzą do wytwarzania, importu, modyfikacji i rozpowszechniania nowych technologii". Z kolei Lundvall (1992) w nieco szerszym ujęciu włącza do systemu „wszystkie części i aspekty struktury gospodarczej i układu organizacyjnego, mające wpływ na procesy uczenia się, badań i odkryć – system produkcyjny, system marketingowy oraz system finansowy jako podsystemy – w których mają miejsce procesy uczenia się". W ujęciu systemowym przepływy informacji i wiedzy zależą też od kontekstu społecznego, politycznego i kulturowego. Początkowo systemy innowacji rozpatrywane były w wymiarze narodowym (narodowe systemy innowacji), obecnie również w wymiarze sektorowym (Malerba 2005) i regionalnym (por. m.in. Cooke et al. 1998, Asheim i Gertler 2005).

Podstawy teoretyczne koncepcji systemu innowacji opierają się na ekonomicznych teoriach ewolucyjnych, podkreślających, że firmy nie są

[2] Cooke (1999) wymienia trzy wady modelu liniowego: przesadne podkreślenie roli badań podstawowych; wprowadzenie sztucznej hierarchii wiedzy stawiającej wiedzę wypływającą z czystych badań podstawowych ponad wiedzą techniczną i inżynieryjną; oraz brak uznania potrzeby ciągłej weryfikacji i oceny rzeczywistej stosowalności badań.



Rysunek 1. Schemat systemu innowacji

Źródło: Arnold i Kuhlman 2001.

w pełni racjonalnymi agentami, mając ograniczony dostęp do informacji i ograniczoną zdolność do interpretacji posiadanych zasobów informacji (*bounded rationality*) (Simon 1976 w: Arnold 2004, Nelson i Winter 1982, Smith 2000). Zdolność do wymiany informacji i współpracy z otoczeniem staje się decydująca dla potencjału innowacyjnego przedsiębiorstwa. Podejście ewolucyjne postrzega proces innowacji jako „zależny od ścieżki"; wiedza i technologie wytwarzane są przez kumulatywny proces uczenia się i interakcji pomiędzy wieloma organizacjami (Nelson i Winter 1982). Rysunek 1 przedstawia system innowacji, ilustrując powiązania pomiędzy głównymi aktorami oraz ich otoczeniem (popyt, infrastruktura, warunki ramowe).

3. CECHY POLITYKI INNOWACJI

3.1. Podstawy interwencji publicznej

Podejście neoklasyczne uzasadnia potrzebę publicznego finansowania działalności B+R niedoskonałościami rynku, które sprawiają, że inwestycje w proces innowacji stają nieopłacalne dla firm (Arrow 1962, Smith 2000, Arnold 2004, Dosi 2006). Arrow (1962, cyt. w Dosi 2006) podkreśla, iż wiedza związana z innowacją charakteryzuje się trzema cechami:

- niepodzielnością (*indivisibility*): część wiedzy na temat nowej technologii lub innowacji jest bezwartościowa; wartość ma tylko całość wiedzy (stąd: długotrwały proces bez możliwości zrealizowania zysku);
- niepewnością i ryzykiem (*non-probabilizable risk*): niemożność pełnego zaplanowania i przewidzenia efektów procesu innowacji;
- niemożnością zrealizowania pełnego zysku z innowacji przez jej twórców (*inappropriability*) w związku z tym, że wiedzy nie można całkowicie zatrzymać w firmie (np. mobilność badaczy, publikacje naukowe).

Pojawienie się ekonomicznych teorii ewolucyjnych i systemu innowacji rozszerzyło podstawy interwencji publicznej o perspektywę ograniczeń firmy oraz dążenie do zniwelowania systemowych niedoskonałości procesu innowacji (Smith 2000, Arnold 2004). Arnold (2004) wymienia cztery typy barier w systemie innowacji, włączające zarówno problemy na poziomie poszczególnych instytucji, jak i całego systemu:

- niski potencjał organizacyjny i technologiczny firm
  - niska zdolność firm do zmiany i dostosowania sie do zmieniających sie warunków prowadzenia działalności (np. niska zdolność absorpcyjna);
- niedoskonałość instytucji
  - niski potencjał uniwersytetów i instytutów B+R, urzędów patentowych i innych ważnych uczestników systemu innowacji;
- niedoskonałość powiązań sieciowych

Ryzyko innowacji – jakie konsekwencje dla polityki innowacyjnej i ewaluacji?

Z punktu widzenia firmy proces innowacji jest ryzykowny, m.in. w związku z trudną do przewidzenia reakcją rynku na nowy produkt lub usługę. Również publiczne finansowanie procesu innowacji łączy się z pewnym poziomem ryzyka. Stopień niepewności i ryzyka zależy m.in. od sytuacji rynkowej sektora, specyfiki danego obszaru technologicznego, wspieranej fazy procesu innowacji czy też stopnia nowości przedsięwzięcia. Ocena profilu ryzyka działań oraz wpisana w logikę programu niska bądź wysoka tolerancja dla niepowodzeń, mają kardynalne znaczenie dla procesu ewaluacji efektywności interwencji publicznej (np. dla sformułowania pytań ewaluacyjnych, zarysowania zakresu badania, zdefiniowania kryterium sukcesu).

Oczekiwanie wysokiej efektywności programów wspierających ambitne projekty innowacyjne oznacza niezrozumienie natury procesu innowacji (Boekholt et al. 2001, Arnold i Boekholt 2002, Perrin 2002). Niepowodzenia są bowiem niedłącznym elementem poszukiwań i eksperymentów mających ambicje wytworzenia innowacji radykalnych. Wyznaczenie zachowawczych, krótkookresowych celów prowadzi do unikania ryzyka i, co za tym idzie, wspierania mniej innowacyjnych projektów. Krótkookresowe cele są właściwe dla programów wspierających mniej ryzykowne innowacje ewolucyjne, np. związane z zakupem nowej technologii.

  - niedostatecznie rozwinięte powiązania i przepływy pomiędzy aktorami systemu innowacji (szczególnie pomiędzy nauką i przemysłem);
  - problemu transformacji i zamknięcia przedsiębiorstw i reżimów technologicznych (por. Smith 2000);
  - nieadekwatna struktura przemysłowa (np. brak rozwiniętych gron);
- nieadekwatne warunki ramowe
  - uregulowania prawne dotyczące prowadzenia działalności gospodarczej, prawo własności przemysłowej i intelektualnej itd. (Smith 2000);
  - uwarunkowania społeczne, popyt i postawy konsumenckie.

Niedoskonałości systemowe uzupełniają perspektywę neoklasyczną logiki interwencji publicznej. Potrzeba zniwelowania niedoskonałości na poziomie systemu – tzw. wąskich gardeł (Arnold 2004) czy też najsłabszych ogniw (Soete 2007) – uzasadnia interwencje państwa mające na celu poprawę funkcjonowania całego system innowacji, a nie tylko działalności B+R i innowacyjnej na poziomie poszczególnych grup celowych (Boekholt et al. 2001, Arnold 2004).

3.2. Cele, zakres i poziomy interwencji polityki innowacyjnej

Widziana z perspektywy systemowej polityka wspierania innowacyjności ma na celu poprawę procesów innowacji na poziomie przedsiębiorstw przez działania dążące do zniwelowania niedoskonałości i barier systemu innowacji (Smith 2000, Arnold 2004). Współczesna polityka innowacyjna rozwinęła się pomiędzy polityką naukową i technologiczną oraz polityką przemysłową i jako taka skupia się przede wszystkim na podniesieniu innowacyjności przedsiębiorstw (por. OECD 2005a). Należy zaznaczyć jednak, że interwencje publiczne, mające wpływ na innowacyjność gospodarki, wykraczają poza granice tak pojmowanej polityki. Instrumenty mające pośredni bądź bezpośredni wpływ na szeroko rozumianą innowacyjność wdrażane są w ramach wielu innych działań publicznych. Funkcjonalny charakter szeroko rozumianej polityki innowacyjnej nie ogranicza się zatem do programów skierowanych do przedsiębiorstw lub sektora B+R. Ważne w kontekście innowacyjności mogą okazać się na przykład polityka edukacyjna czy polityka ochrony środowiska (zob. tabela 1).

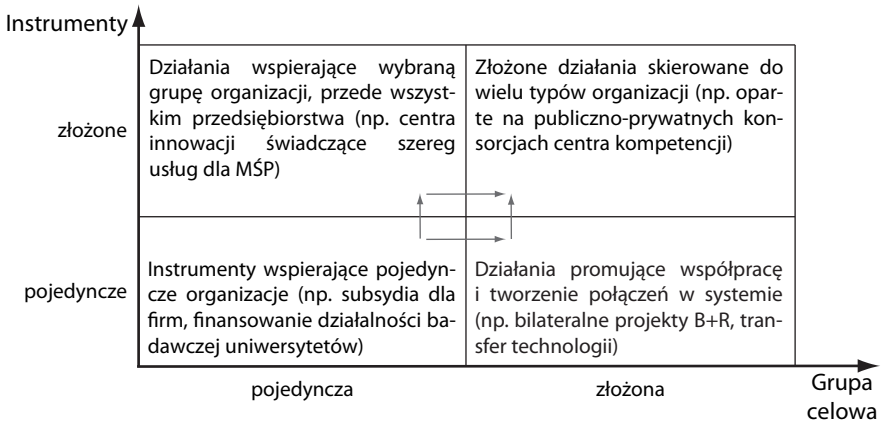
Arnold (2007) podkreśla, że w ostatnich latach następuje zmiana podejścia do działań polityki innowacyjnej (zob. rysunek 3). Tradycyjne instrumenty skupiały się na wsparciu poszczególnych instytucji przez pojedyncze działania, takie jak dofinansowanie działalności B+R poszczególnych firm. Następnie pojawiła się tendencja do wprowadzania bardziej złożonych działań, mających na celu podniesienie potencjału całych grup celowych, np. regionalne centra oferujące firmom usługi

Tabela 1. Taksonomia polityki innowacyjnej

Cele	Sektorowa polityka innowacyjna	Wielosektorowa polityka innowacyjna
Polityka innowacyjna Cel: wzrost innowacyjności gospodarki i wzrost gospodarczy	Polityka innowacyjna w wąskim rozumieniu (skupiająca się na polityce technologicznej i przemysłowej)	Zintegrowane polityki naukowa, technologiczna i innowacyjna
Polityka innowacyjna w szerszym rozumieniu Cel: wzrost gospodarczy i podniesienie jakości życia	Polityka innowacyjna w innych sektorach, np. zdrowie publiczne, ochrona środowiska	Horyzontalna/zintegrowana czy też spójna/systemowa polityka innowacyjna

Źródło: OECD 2005a.

Rysunek 2. Ewolucja instrumentów wspierania B+R i innowacji (Arnold 2007)



doradcze, oferty transferu technologii i informacje na temat finansowania innowacji (formuła *one-stop shop*). Pojawiły się również instrumenty kładące nacisk na współpracę pomiędzy różnymi typami instytucji, przede wszystkim biznesu i nauki (np. tworzenie sieci tematycznych).

Najnowsze instrumenty polityki innowacyjnej włączają złożone *portfolia* działań obejmujące wiele grup celowych: MAPs (*Multi-Actor, Multi-Measure Programmes*). Typowym przykładem takich działań są centra kompetencji, oparte na szerokich konsorcjach publiczno-prywatnych (por. Baumann et al. 2004). Tego typu działania dają możliwość oddolnego zdefiniowania szczegółowych celów i formy organizacji. Wymóg ścisłej współpracy nauki i biznesu w formule MAP oznacza, że wdrażane działania mają szanse być bliższe realiom rynku i rzeczywistym potrzebom społeczeństwa (por. Arnold 2007).

4. ZAGADNIENIA EWALUACJI POLITYKI INNOWACYJNEJ

4.1. Poziomy i zakres analizy publicznego wsparcia innowacyjności

Arnold (2004) rozróżnia trzy poziomy ewaluacji polityki innowacyjnej:

- pojedyncze interwencje publiczne (np. projekt, działanie lub program);
- ewaluacja podsystemów lub analizy „wąskich gardeł" systemu (np. analizy tematyczne na temat sektorów, grup celowych, gron);
- ewaluacja na poziomie systemu innowacji.

Ewaluacja na poziomie pojedynczej interwencji publicznej, najczęściej działania lub programu, pozostaje ważnym elementem oceny efektywności i skuteczności polityki innowacyjnej. Jednakże szersza refleksja nad odpowiedniością oraz użytecznością interwencji publicznej wymaga przeanalizowania danego instrumentu w kontekście jego roli i wpływu na system innowacji (Boekholt et al. 2001, Arnold 2004). Ewaluacja

podsystemów znajduje się pomiędzy ewaluacją pojedynczych interwencji i całego systemu, obejmując więcej niż jedną interwencję publiczną, wiele typów instytucji oraz ich zależności (Arnold 2004). Przykładowo, mogą to być przekrojowe studia sektorowe, analizy poświęcone wybranym obszarom technologicznym, opracowania na temat wpływu interwencji publicznych na wybrane typy instytucji, firm czy grup celowych.

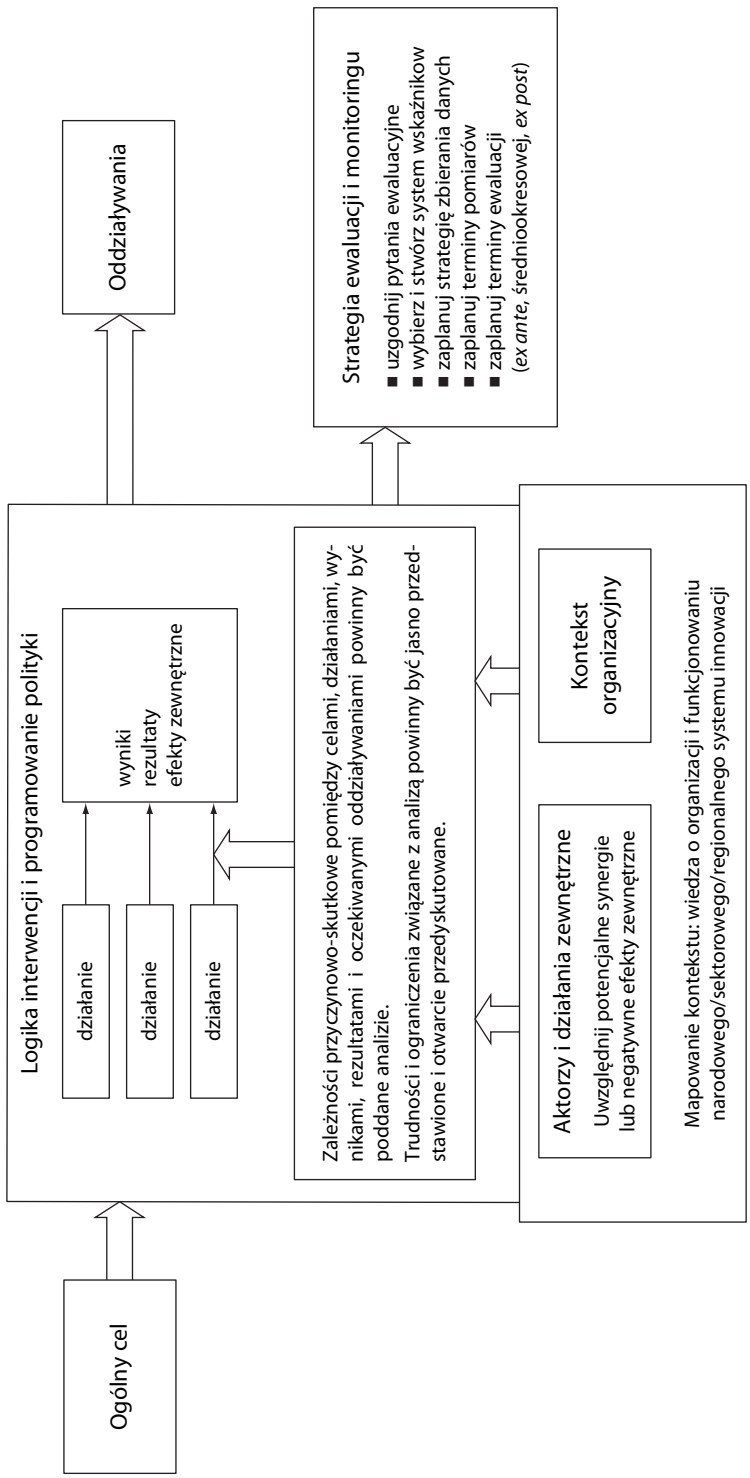
Szeroka analiza systemu innowacji skupia się na roli i zasobach poszczególnych uczestników systemu, dynamiki połączeń i przepływów pomiędzy nimi oraz warunków ramowych procesu innowacji. Analiza na poziomie systemu ocenia wpływ interwencji publicznych, skupiając się na efektach szeroko rozumianej polityki innowacyjnej państwa. Taka szeroka analiza staje się też punktem odniesienia dla studiów na poziomie działania czy programu oraz analiz podsystemów (ewaluacji tematycznych). Lepsze zrozumienie specyfiki całego systemu innowacji daje również solidniejszą podstawę priorytetyzacji działań państwa.

4.2. Wybrane wyzwania ewaluacji innowacji

4.2.1. Wyzwania strategiczne

Aby ewaluacja mogła pełnić funkcję formatywną i prowadzić do lepszego planowania przyszłych – czy też korygowania wdrażanych – działań, proces ewaluacji powinien być zintegrowany z planowaniem strategicznym, programowaniem i wdrażaniem polityki (Boekholt et al. 2001, Anderson et al. 2004). Wkomponowanie ewaluacji w planowanie i programowanie nadaje jej funkcję strategiczną w procesie ustalania priorytetów i ulepszania całego procesu wdrażania polityki (rysunek 3). Wnioski z ewaluacji powinny stać się również ważnym wkładem w proces formułowania długookresowej wizji i planowania. Tak rozumiana ewaluacja staje jednym z głównych narzędzi tworzenia zasobów strategicznej wiedzy, na których mogą zostać oparte decyzje publiczne (Fahrenkrog et al. 2002).

Nadanie ewaluacji funkcji strategicznej wymaga od ewaluacji wysokiej jakości. Oznacza to potrzebę poważnych inwestycji w ewaluację w celu podniesienia jakości procesu oraz otwarcia szerokiego publicznego dyskursu na temat wniosków z oceny jakości interwencji publicznej państwa. Otwartość takiego dyskursu jest jednym z istotnych elementów „kultury ewaluacji" i dobrego zarządzania państwem. W literaturze podkreśla się, że ewaluacja powinna być nastawiona na dostarczanie użytecznych informacji i wniosków dla wszystkich uczestników programu – nie tylko dla administracji publicznej, lecz także dla beneficjentów i potencjalnych grup celowych. Uczestnictwo przedstawicieli różnych grup celowych w procesie ewaluacji może prowadzić do lepszego społecznego odbioru polityki, poprawienia samego procesu ewaluacji oraz polepszenia jakości programów publicznych (Boekholt et al. 2001).



Rysunek 3. Integracja ewaluacji w proces programowania

Źródło: Appelquist 2004.

4.2.2. Wyzwania analityczne

Zrozumienie specyfiki procesu innowacji

Ewaluatorzy polityki innowacyjnej powinni znać specyfikę procesu innowacji zarówno na poziomie przedsiębiorstw, jak i całego systemu innowacji. Bez zrozumienia specyfiki obszaru badawczego ewaluacja nie może spełniać funkcji formatywnej i strategicznej, a wręcz grozi wyciąganiem błędnych wniosków i przecenieniem lub niedocenieniem pełnej wartości interwencji publicznej (Georghiou 2007).

Kompleksowy charakter procesu innowacji wymaga zarówno systemowego podejścia do polityki innowacyjnej, jak i systemowego rozumienia wpływu interwencji publicznych na procesy innowacji (Arnold 2004). Zasadniczym wyzwaniem analitycznym w tym kontekście jest włączenie podejścia systemowego do praktyki ewaluacji. Arnold (2004) podkreśla, że zidentyfikowanie niedoskonałości w powiązaniach i przepływach pomiędzy różnymi elementami systemu innowacji powinno należeć do kluczowych ról polityki innowacyjnej państwa. Głównym narzędziem analizy tych niedoskonałości są studia ewaluacyjne. Wiedza na temat funkcjonowania systemu innowacji pozwala rozróżnić jego podsystemy i główne procesy oraz wskazać niedoskonałości i bariery. Pozwala to lepiej zrozumieć kontekst ewaluowanych działań oraz odpowiednio dobrać poziom i zakres ewaluacji.

Poziomy i zakres ewaluacji polityki innowacyjnej

Dobór poziomu i zakresu analizy jest fundamentalnym elementem procesu ewaluacji. W celu dostrzeżenia szerszych efektów społeczno-gospodarczych interwencji publicznej należy oceniać je na odpowiednim poziomie. Praktycy podkreślają, że aby móc analizować efekty polityki, ewaluacja powinna objąć całe programy lub *portfolia* działań wchodzących w skład polityki innowacyjnej (Boekholt et al. 2002, Fahrenkrog et al. 2002, Georghiou et al. 2002, Appelquist 2004). Tylko wyższy poziom agregacji daje możliwość oceny polityki na poziomie systemu innowacji. Również w porównaniach efektywności różnych instrumentów polityki innowacyjnej każdy z nich powinien być analizowany w kontekście pozostałych instrumentów i ich relacji do systemu innowacji. Należy pamiętać, iż skuteczna polityka powinna składać się z synergicznych lub co najmniej komplementarnych działań (Boekholt et al. 2002).

Ewaluacja skupiona na poziomie pojedynczego projektu czy działania nie pozwala na analizę szerszych efektów. Podstawowym problemem jest tu niemożność jednoznacznej atrybucji efektów w szerokim kontekście funkcjonowania systemu innowacji. Nie oznacza to, iż ewaluacja powinna się skupić wyłącznie na poziomie *portfolio* działań czy większych programów. Ewaluacje poszczególnych interwencji (np. projektów czy działań) odgrywają istotną rolę głównie w ocenie efektywności i skuteczności, lecz nie mogą w pełni odpowiedzieć na strategiczne pytania

o adekwatność i efekty całej polityki czy relatywną ocenę jej różnych instrumentów.

Strategiczny proces ewaluacji powinien zatem równoważyć ewaluacje na poziomie instrumentów (np. programów czy działań), podsystemów (ewaluacje tematyczne) oraz całego systemu (ewaluacja polityki). Badania ewaluacyjne o różnych zakresach i prowadzone na różnych poziomach powinny wzajemnie się uzupełniać i wzmacniać, będąc częścią szerszego procesu uczenia się, mającego na celu polepszenie jakości interwencji publicznej (por. też Arnold 2004).

Refleksja nad dodatkowością interwencji publicznych

Dodatkowość interwencji publicznej jest podstawowym problemem ewaluacji. Pytanie o dodatkowość zmierza do oceny wartości dodanej przez ewaluowaną interwencję i odpowiedzi, czy stworzony efekt usprawiedliwia inwestycję publiczną. Ewaluacja tradycyjnie skupiała się na ocenie bezpośrednich efektów wsparcia na poziomie beneficjenta, takich jak wzrost obrotów czy redukcja kosztów uzyskana w związku z wprowadzoną innowacją. W zależności od celów ewaluacji analiza dodatkowości może być prowadzona na różnym poziomie agregacji (Geoghiou 2002).

W literaturze rozróżnia się kilka typów dodatkowości polityki innowacyjnej. Najczęściej mówi się o dodatkowości inwestycyjnej (*input additionality*), wynikowej (*output additionality*) oraz behawioralnej (*behavioural additionality*) (Georghiou 2002, OECD 2006). Głównym pytaniem w analizie dodatkowości inwestycyjnej brzmi: „Czy wsparcie publiczne jest dodatkiem, czy substytutem inwestycji firmy w dany projekt?". Analiza dodatkowości inwestycyjnej bywa problematyczna; ocena intencji firmy korzystającej ze wsparcia jest trudna (np. firma może uzyskać wsparcie na projekt, który wprowadziłaby niezależnie od uzyskania finansowania publicznego).

Dodatkowość wynikowa ocenia, jaka część wyników projektu nie mogłaby być osiągnięta bez wsparcia publicznego. Ocena nie jest skomplikowana jeżeli chodzi o wyniki projektu. Problem atrybucji pojawia się w przypadku oceny rezultatów i oddziaływań. Zaobserwowanie niektórych efektów projektów innowacyjnych możliwe jest dopiero lata po zakończeniu ewaluowanych działań (Georghiou 2002). Ponadto oceniane działania są często uzupełnieniem skomplikowanego i długotrwałego procesu. Prosta ocena dodatkowości wyjętego z kontekstu projektu nie bierze pod uwagę perspektywy specyfiki procesu innowacji i perspektywy beneficjenta. Problem ten znany jest w literaturze jako „pułapka projektu" (*project fallacy*) (Georghiou 2002, Georghiou et al. 2002, OECD 2006). Wyzwaniem w tym kontekście jest zrozumienie znaczenia interwencji publicznej z punktu widzenia beneficjenta (grupy celowej) czy procesu innowacji (może włączać więcej niż jedną organizację) oraz przeprowadzenie analizy na wyższym poziomie agregacji (np. sektora).

W praktyce ewaluacji pojawia się coraz więcej podejść poszukujących uniknięcia „pułapki projektu" i oceny pośrednich efektów interwencji

(Georghiou 2002, 2007, OECD 2006, Falk 2006, Autio et al. 2008). Analizy dodatkowości behawioralnej dążą do pełniejszego zrozumienia efektów interwencji publicznej na beneficjentów, włączając zmiany organizacji i zarządzania, podejścia do działalności innowacyjnej czy też współpracy z innymi organizacjami (OECD 2006). Co istotne, badania wskazują, że tego typu efekty są bardziej trwałe i potencjalnie głębsze niż bezpośrednie efekty subsydiów czy zwolnień podatkowych na wydatki na B+R (Georghiou 2007, Falk 2006).

Ważne jest również, by ewaluacja włączyła niezamierzone (pozytywne i negatywne) efekty interwencji publicznych. Niezamierzone efekty często umykają ewaluatorom, gdyż nie odpowiadają one bezpośrednio celom programu, wokół których zazwyczaj formułowane są pytania ewaluacyjne (Georghiou et al. 2002).

Profil przedsiębiorstw i innych grup celowych

Przedsiębiorstwa o różnym profilu (np. sektor działalności) mają skrajnie różny potencjał innowacyjny, jak również inne podejście do korzystania z pomocy publicznej. Ze zróżnicowaniem mamy do czynienia również w przypadku innych grup celowych polityki innowacyjnej (np. instytutami B+R). Ewaluacje powinny włączać analizę dotyczącą efektów interwencji publicznych na różne profile przedsiębiorstw i innych ważnych grup celowych. Dotyczy to zarówno działań horyzontalnych, otwartych dla wszystkich firm, jak i instrumentów bardziej restrykcyjnych (np. otrzymanie status centrum B+R). Analizy biorące pod uwagę wybraną grupę celową mogą być nie tylko elementem ewaluacji poszczególnego działania, lecz także oddzielnym przekrojowym badaniem efektów pełnego *portfolio* działań polityki innowacyjnej.

Ujęcie sektorowe

W literaturze przedmiotu zwraca się uwagę na zróżnicowanie procesów innowacji w różnych sektorach (Malerba 2005). Ocena rezultatów i efektów programów powinna zatem uwzględnić różnice pomiędzy różnymi sektorami, takie jak:

- ogólny profil sektora i jego firm (np. struktura przemysłowa, zaawansowanie technologiczne i zdolność absorpcyjna, poziom konkurencji);
- specyfika procesu innowacji (np. kapitałochłonność i wymogi techniczne procesu B+R, rola zasobów ludzkich, czas trwania procesu);
- przepływy i współpraca w sektorze (np. otwartość na współpracę z innymi firmami czy instytutami B+R oraz współpraca między różnymi sektorami);
- warunki ramowe (np. rola praw własności przemysłowej i praw autorskich).

W zależności od sektora produkcji proces badań i rozwoju odgrywa mniej lub bardziej istotną rolę w procesie wytwardzania innowacji. Warto tu podkreślić, że duża część procesów innowacji nie jest oparta na

B+R i zależy od kreatywności i kwalifikacji pracowników, współpracy z innymi przedsiębiorstwami i instytucjami badawczymi, oraz zarządzania i struktury organizacyjnej przedsiębiorstw. Zróżnicowanie sektorowe procesów innowacji – jak też inna „waga" systemowych „wąskich gardeł" w przypadku różnych sektorów – powinno być wzięte pod uwagę zarówno w programowaniu, jak i ewaluacji polityki wspierania innowacji.

Zróżnicowanie regionalne

Procesy innowacji są zróżnicowanie przestrzennie (Cooke 1998, Asheim i Gertler 2005). Regionalne systemy innowacji różnią się od siebie:

- profilem sektorowym i potencjałem uczestników procesu innowacji (np. strukturą przedsiębiorstw, potencjałem B+R, obecnością firm z kapitałem zagranicznym, obecnością instytucji naukowo-badawczych);
- jakością powiązań w systemie (np. poziom zaufania pomiędzy lokalnymi MŚP i instytutami B+R, poziom współpracy międzynarodowej);
- warunkami ramowymi (np. jakość infrastruktury badawczej czy doświadczenie w prowadzeniu regionalnej polityki innowacyjnej).

Cechy regionalnego systemu innowacji mogą predysponować region do uzyskiwania lepszych bądź gorszych rezultatów i efektów polityki innowacyjnej w różnych obszarach wsparcia. Podobnie jak w przypadku ujęcia sektorowego, pominięcie przestrzennego wymiaru ewaluacji może prowadzić do niezrozumienia zróżnicowania rezultatów polityki i w konsekwencji pominięcia tych aspektów w przyszłości.

4.2.3. Wyzwania metodologiczne

Właściwy dobór metod

Metodologia ewaluacji powinna zostać stworzona w taki sposób, by odpowiadała celom strategicznym, zakresowi i poziomowi analizy. Uwzględnienie systemowego charakteru procesów innowacji ma zasadniczy wpływ na dobór metod i narzędzi badawczych. Wybrane metody powinny dostarczyć informacji, pozwalających na ocenę danego działania w kontekście jego roli w systemie innowacji. Praktycy podkreślają, że w związku z kompleksowością tematyki ewaluacje programów publicznych w tym obszarze powinny opierać się na wielu uzupełniających się metodach badawczych. Obok metod ilościowych (najczęściej badań ankietowych), istotne jest włączenie metod jakościowych (np. studiów przypadku, paneli eksperckich itd.) (por. Georghiou 2007). Pozwala to stworzyć pełniejszy obraz efektów interwencji publicznej oraz, z drugiej strony, zmniejszyć ryzyko wyciągnięcia błędnych wniosków na podstawie jednostronnej lub powierzchownej analizy (Boekholt et al. 2001, Fahrenkrog et al. 2002).[3]

---

[3] Przykłady doboru i zastosowań różnych metod badawczych używanych w ewaluacjach polityki innowacyjnej można znaleźć m.in. w Georghiou i Roessner (2000), Boekholt et al. (2001), Fahrenkrog et al. (2002), Georghiou et al. (2002), Baumann et al. (2004).

Otwarty proces tworzenia metodologii ewaluacji

Podstawy metodologii ewaluacji interwencji publicznej powinny zostać uzgodnione w tym samym czasie, kiedy planowana jest interwencja publiczna. Zapewnia to odpowiednie przygotowanie do zbierania danych oraz akceptację procedur i kryteriów procesu ewaluacji przez instytucje wdrażające (Boekholt et al. 2001). W kontaktach pomiędzy ewaluatorami i administracją publiczną już podczas planowania systemu ewaluacji powinno się otwarcie dyskutować o możliwych nieścisłościach i trudnościach metodologicznych związanych z ewaluacją efektów danego programu. Typowe problemy dotyczą atrybucji szerszych efektów społecznych interwencji czy trudności z przedstawianiem efektów za pomocą wskaźników takich jak nowo powstałe miejsca pracy, wprowadzone na rynek produkty (por. Georghiou 1998, Boekholt et al. 2001). Tworzenie i wdrażanie metodologii ewaluacji powinno cechować się rygorem i być skrupulatnie dokumentowane. Podnosi to przejrzystość oraz zaufanie do wyników i zaleceń ewaluacji.

Wymiana i uczenie się od innych

Wiele państw UE i OECD ma długoletnie doświadczenie w prowadzeniu polityki innowacyjnej i ewaluacji jej efektów. Poza studiami ewaluacyjnymi na poziomie narodowym prowadzone są również międzynarodowe programy badawcze dotyczące polityki innowacyjnej oraz ewaluacji w obszarze polityki innowacyjnej[4]. Warto w sposób mądry korzystać z tych zasobów i wiedzy eksperckiej oraz brać czynny udział w wymianach międzynarodowych, jednak analizując istniejące praktyki, należy zawsze mieć na uwadze specyfikę polskiego systemu innowacji i polityki innowacyjnej. Proces ewaluacji musi być „zakorzeniony" i dostosowany do lokalnego kontekstu. Proste kopiowanie podejść zastosowanych w innych realiach mija sie z celem.

5. PODSUMOWANIE

Ewaluacja tematyki innowacyjności wymaga wiedzy na temat specyfiki procesu innowacji i funkcjonowania całego systemu innowacji, znajomości polityki innowacyjnej oraz szerokiego rozeznania w różnych metodach badawczych i analitycznych stosowanych w ewaluacji. Należy podkreślić, że tak jak w przypadku jakichkolwiek pogłębionych badań, głębokie zrozumienie kontekstu ewaluowanych procesów jest wymogiem podstawowym i niezbędnym.

---

[4] Przykładem godnym przytoczenia jest inicjatywa Komisji Europejskiej *PRO-INNO Europe*, włączająca między innymi projekty *INNO-appraisal* (ewaluacja polityki innowacyjnej) oraz *INNO-policy trendchart* (analiza zmian i trendów w polityce innowacyjnej państw członkowskich UE). Wszystkie materiały dostępne są na http://www.proinno-europe.eu

Wybrane wyzwania ewaluacji polityki innowacyjnej w skrócie

Wyzwania strategiczne

- ewaluacja powinna być zintegrowana z planowaniem i programowaniem polityki, co nadaje jej funkcję strategiczną w procesie ustalania priorytetów i poprawy wdrażania działań publicznych;
- nadanie ewaluacji funkcji strategicznej wymaga wysokiej jakości procesu ewaluacji oraz użycia wniosków z ewalkuacji w szerokiej publicznej debaty debacie na temat efektów i kierunków polityki państwa;
- ewaluacja powinna dostarczać informacji i wniosków użytecznych nie tylko dla administracji publicznej, ale też dla beneficjentów i grup celowych polityki.

Wyzwania analityczne

- strategiczny proces ewaluacji powinien równoważyć ewaluacje na poziomie pojedynczych instrumentów (np. programow), podsystemów (ewaluacje tematyczne) oraz całego systemu (ewaluacja polityki);
- dostrzeżenie szerszych efektów społeczno-gospodarczych interwencji publicznej wymaga oceny na wyższym poziomie agregacji;
- podejście systemowe pozwala lepiej zrozumieć kontekst ewaluowanych działań oraz odpowiednio dobrać poziom i zakres ewaluacji;
- ewaluacja powinna dążyć do zrozumienia znaczenia interwencji publicznej z punktu widzenia beneficjenta (grupy celowej) oraz w kontekście jej udziału w całym procesie innowacji (nacisk na dodatkowość behawioralną);
- ewaluacja powinna objąć analizą zarówno (pozytywne i negatywne) zamierzone, jak i niezamierzone efekty interwencji publicznych;
- analiza powinna wziąć pod uwagę zróżnicowanie efektów interwencji publicznej w zależności od profilu grupy celowej, sektora działalności oraz regionu.

Wyzwania metodologiczne

- metodologia ewaluacji powinna odpowiadać celom strategicznym, zakresowi i poziomowi analizy;
- podstawy metodologii i sposób zbierania danych powinny zostać uzgodnione w tym samym czasie, kiedy planowana jest interwencja publiczna;
- metodologia powina opierać się na zarówno ilościowych, jak i jakościowych, uzupełniających się metodach badawczych.

Zadaniem ewaluacji jest przeanalizowanie i wyciągnięcie wniosków na temat wpływu interwencji publicznej na innowacyjność przedsiębiorstw i całej gospodarki. Wiedza na temat specyfiki procesu innowacji pozwala umieścić analizowane działania w szerszym kontekście. Oczywiście, ewaluacja relatywnie małego projektu nie może ocenić jego wpływu na cały system innowacji, lecz może dążyć do przedstawienia jego działań w kontekście specyficznego podsystemu (np. sektora, grona itd.). Z kolei ewaluacja całego działania lub programu, w ramach którego wdrożono szereg podobnych projektów, może już aspirować do oceny szerszych efektów. Wreszcie ewaluacja całej polityki powinna odnosić się do wpływu szeroko pojętego *portfolio* działań publicznych na system innowacji i funkcjonowanie całej gospodarki.

Ewaluacja polityki innowacyjnej powinna łączyć zarówno funkcje konkluzywne, jak i formatywne. Systemowe (czy też kontekstowe) podejście do analizowania procesu innowacji pozwala nadać ewaluacji strategiczną rolę w procesie planowania strategicznego i programowania. W tym kontekście rozwijanie ewaluacji powinno być postrzegane jako dlugookresowa inwestycja w podnoszenie jakości interwencji publicznej. Praktyka ewaluacji potwierdza pojawiającą się tendencję do postrzegania ewaluacji jako części procesu tworzenia strategicznej wiedzy (*strategic intelligence*) wykorzystywanej przez rządy do lepszego planowania i priorytetyzowania swoich działań (Fahrenkog et al. 2002).

Niezwykle ważnym wyzwaniem w tym kontekście, zresztą istotnym dla wszystkich skomplikowanych obszarów tematycznych, jest unikanie prostego reprodukowania istniejącej wiedzy i podejść do ewaluacji. Procesy innowacji i dynamika systemu innowacji podlegają ciągłym zmianom. Jeśli ewaluacja ma stać się procesem uczenia się, istotnym dla planowania i programowania polityki państwa, to oryginalność podejścia i przemyślany dobór metod stają się warunkami jej użyteczności.

BIBLIOGRAFIA

Andersson T., Appelquist J., Schwaag Serger S. (2004), *Public research and innovation Policy for the good of society: how to assess the way forward?*, artykuł na seminarium IKED i agencji VINNOVA, listopad 2004, Sztokholm.

Arnold E., Kuhlman S. (2001), *RCN in the Norwegian Research and Innovation System, Background Report No. 12 in the Evaluation of the Research Council of Norway*, Oslo: Royal Norwegian Ministry for Education, Research and Church Affairs.

Arnold E. (2004), Evaluating research and innovation policy: a systems world needs systems evaluations, *Research Evaluation* **1**, 3–17.

Arnold E. (2007), *Governing the Knowledge Infrastructure in an Innovation Systems World, artykul dla Inter-American Development Bank*, http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=976013 (15.05. 2008)

Asheim B., Gertler M. (2005), *The Geography of Innovation: Regional Innovation Systems*, w: J. Fagerberg, D. Movery, E. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford University Press.

Baumann B., Bührer S., Greer H., Hochreiter H., Mayer S., Simmonds P., Stampfer M., Sturn D. (2004), *Road MAP – Good practices for the management of Multi Actors and Multi Measures Programmes (MAPs) in RTDI policy*, raport dla Komisji Europejskiej, STRATA.

Boekholt P., Lankhuizen M., Arnold E., Clarke J., Kuusisto J., de Laat B., Simmonds P., Cozzens S., Kingsley G., Johnston R. (2001), *An international review of methods to measure relative effectiveness of technology policy instruments*, Haga Min, EZ.

Boekholt P. (ed.) (2002), *Innovation policy and sustainable development: can innovation incentives make a difference?, Contributions to a Six Countries Pogramme conference*, February 28 – March 1, IWT-ObservatoryBrussels, IWT-Studies 40.

Cooke P., Heidenreich M. and Braczyk H-J (eds.) (1998*), Regional Innovation Systems: the Role of Governances in a Globalized World*, London, Routledge.

Dosi G., Malerba F., Ramello G. (2006), Information, appropriability, and the generation of innovative knowledge four decades after Arrow and Nelson: an introduction, *Industrial and Corporate Change* 15(6), 891–901.

Fahrenkrog G., W Polt, J Rojo, A Tübke and K Zinöcker (2002), *RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies*, IPTS Technical Report Series, STRATA.

Falk F. (2006), *Measuring the Effects of Public Support Schemes on Firms' Innovation Activities, Survey Evidence from Austria*, WIFO Working Papers (267) January.

Freeman C. (1987), *Technology Policy and Economic Performance: Lessons from Japan*, Londyn, Pinter.

Georghiou L. (1998), Issues in the Evaluation of Innovation and Technology Policy, *Evaluation* **4(1)**, 37–51.

Georghiou L. (2000), Evaluating technology programs: tools and methods, *Research Policy* **29**, 657–678.

Georghiou L. (2002), *Impact and additionality of innovation policy*, w: P. Boekholt (ed.), *Innovation policy and sustainable development: can innovation incentives make a difference?*, Contributions to a Six Countries Pogramme conference, February 28 – March 1, IWT-Observatory, IWT-Studies 40, Brussels.

Georghiou L., Rigby J., Cameron H. (eds.) (2002), *Assessing the Socio-economic Impacts of the Framework Programme*, projekt ASIF, raport dla Komisji Europejskiej.

Georghiou L. (2007), What lies beneath: avoiding the risk of under-evaluation, *Science and Public Policy* **34(10)**, 743–752.

LL&A, PREST, ANRT, Reidev Ltd. (2006), *SMART Innovation: A practical guide to evaluating innovation programs*, ftp://ftp.cordis.lu/pub/innovation/docs/sar1_smartinnovation_master2.pdf (15.05.2008)

Lundvall B.-A. (1993), *National Innovation Systems: Towards a Theory of Innovation and Interactive Learning*, London, Pinter.

Lundvall B.-A. (2005), *Science, Technology and Innovation Policy*, w: J. Fagerberg, D. Movery, E. Nelson, *The Oxford Handbook of Innovation*, Oxford, Oxford University Press.

Malerba F. (2005), *Sectoral Systems: How and Why Innovation Differs Across Sectors*, w: J. Fagerberg, D. Movery, E. Nelson, *The Oxford Handbook of Innovation*, Oxford, Oxford University Press.

Miedziński M. (2001), *Koordynacja procesów innowacji na przykładzie polskiego województwa. Aspekty instytucjonalne regionalnych systemow innowacji*, w: A. Kuklinski (red.) *Gosporka oparta na wiedzy*, KBN, Warszawa 2001.

Nelson R. (ed.) (1993), *National Innovation Systems: a Comparative Analysis*, Oxford,Oxford University Press

OECD (2006), *Government R&D Funding and Company Behaviour*, *Measuring Behavioural Additionality*, Paris.

OECD (2005a), Governance of Innovation Systems, Vol. 1: Synthesis Report, Paris

OECD (2005b), *Oslo Manual*, Paris.

Perrin B. (2002), How to – and how not to – evaluate innovation, *Evaluation* **8(1)**.

Reid A., Miedzinski M. (2008), *Eco-Innovation, Final sectoral report, Sectoral Innovation Systems in Europe: Monitoring, Analysing Trends and Identifying Challenges* (SYSTEMATIC project, contract number: 022534).

Reid A., Peter V. (2008), *Sectoral Innovation Systems: the Policy Landscape in the EU25, Final report, Sectoral Innovation Systems in Europe: Monitoring, Analysing Trends and Identifying Challenges* (SYSTEMATIC project, contract number: 022534)

Smith K. (2000), Innovation as a Systemic Pheonmenon: Rethinking the Role of Policy, *Enterprise and Innovation Management Studies* **1(1)**, 73–102.

Soete L. (2007), From Industrial to Innovation Policy, *J Ind Compet Trade* **7**, 273–284.

ROZDZIAŁ 6.4.

Agnieszka Sternik

EWALUACJA INTERWENCJI W OBSZARZE ROZWOJU ZASOBÓW LUDZKICH

Nie wszystko, co jest ważne, jest policzalne, Nie wszystko, co jest policzalne, jest ważne.

Albert Einstein

1. WPROWADZENIE

Z roku na rok wzrasta rola i liczba badań, w tym ewaluacji w obszarze polityk publicznych. Decydenci coraz częściej stawiają pytania nie tylko o to, co zostało wdrożone i w jakim zakresie, ale również w jaki sposób i jakie przyniosło rezultaty (Lion 2006). Od badaczy wymaga się zatem dostarczenia wiedzy jednocześnie na temat rezultatów programów i wytłumaczenia dlaczego i jak doszło do ich osiągnięcia. Wyzwania stojące przez ewaluacją polityk i programów publicznych w dużej mierze zależą jednak od obszaru polityki, który poddany jest analizie.

Niniejszy rozdział podejmuje próbę zdefiniowania głównych charakterystyk ewaluacji interwencji prowadzonych w obszarze rozwoju zasobów ludzkich. W tekście dyskutowane są główne aspekty wyróżniające tę ewaluację od oceny innych polityk. Podjęta jest też próba zaprezentowania metodologii badań, która umożliwia analizę, z jednej strony zmian zachodzących w ogólnej sytuacji społeczno-ekonomicznej na skutek realizacji interwencji publicznych w obszarze rozwoju zasobów ludzkich, z drugiej zaś wpływu programów na sytuację osób objętych tymi interwencjami. Ta dwoistość perspektywy makro i mikro odzwierciedla

złożoność polityk skierowanych na rozwój kapitału ludzkiego oraz samej ewaluacji, od której wymaga się wyjścia poza ocenę dokonywaną za pomocą metod ilościowych, agregowanych na stosunkowo wysokim poziomie, umożliwiając w ten sposób uchwycenie szerszego kontekstu, w którym realizowany był program bądź projekt.

W związku z tym, że katalog możliwych działań w tym obszarze jest bardzo szeroki, zakres niniejszego rozdziału zawężony został przede wszystkim do zagadnień związanych z ewaluacją interwencji finansowanych ze środków Europejskiego Funduszu Społecznego (EFS).

2. DEFINICJA OBSZARU BADAWCZEGO – JAK ROZUMIEĆ ROZWÓJ ZASOBÓW LUDZKICH

Rozważania na temat specyfiki ewaluacji interwencji w obszarze rozwoju zasobów ludzkich należy rozpocząć od próby zdefiniowania jego zakresu. Obejmuje on interwencje publiczne podejmowane w ramach polityk, m.in. zatrudnienia i rynku pracy, edukacji oraz integracji społecznej. Dodatkowo, pojęcie „zasoby ludzkie" w niektórych kontekstach określa się jako „kapitał ludzki". W klasycznej definicji, stosowanej m.in. w celu zbadania indywidualnych strategii życiowych przez Department for International Development (DFID) w Wielkiej Brytanii, pojęcie kapitału ludzkiego występuje obok czterech innych „kapitałów". Kategoryzacja ta, wywodząca się po części z metodologii *sustainable livelihood analysis*, identyfikuje, poza kapitałem ludzkim, kapitał fizyczny, rozumiany jako posiadanie przez obywatela dostępu do podstawowej infrastruktury; kapitał finansowy wyrażony w posiadanych zasobach i dostępie do środków finansowych przez jednostki; kapitał naturalny, definiowany jako możliwość korzystania z zasobów naturalnych oraz kapitał społeczny – znajomości, powiązania, przynależność do grup. Ostatnie dwie kategorie z definicji DFID można przyjąć jako wyznacznik interwencji w obszarze zasobów ludzkich. Kapitał ludzki, rozumiany jest jako umiejętności, wiedza, ale również zdrowie, które umożliwiają jednostce wybór preferowanej przez siebie strategii życiowej (DFID, 2005: 38). Budowa kapitału ludzkiego odnosi się przede wszystkim do nabywania kognitywnych i pozakognitywnych umiejętności człowieka zarówno w domu rodzinnym, jak i przez proces formalnej i nieformalnej edukacji trwającej przez całe życie. Umiejętności te umożliwiają jednostkom uczestniczenie w życiu społecznym i gospodarczym oraz współuczestniczenie w rozwoju wiedzy (Fuente 2006).

W Polsce Program Operacyjny Kapitał Ludzki (PO KL), który współfinansowany jest ze środków Europejskiego Funduszu Społecznego (EFS), definiuje kapitał ludzki jako pojęcie oznaczające „zasób wiedzy, umiejętności oraz potencjał zawarty w każdym człowieku i w społeczeństwie

jako całości, określający zdolności do pracy, adaptacji do zmian w otoczeniu oraz możliwości tworzenia nowych rozwiązań" (MRR, 2007(1): 5). Program odnosi się także do pojęcia „kapitału społecznego", który w rozumieniu autorów dokumentu oznacza „zasoby umiejętności, informacji, kultury, wiedzy i kreatywności jednostek oraz związki pomiędzy ludźmi i organizacjami" (ibidem).

Mówiąc o ewaluacji interwencji z obszaru rozwoju zasobów ludzkich, mamy więc na myśli badania tych programów i działań, których celem jest sprzyjanie budowie potencjału adaptacyjnego, kreatywności i wiedzy, zarówno na poziomie jednostki, jak i społeczeństwa, dzięki którym możliwe jest pełniejsze wykorzystanie zasobów pracy oraz wsparcie wzrostu konkurencyjnego gospodarki.

3. PRIORYTETY INTERWENCJI W OBSZARZE ROZWOJU ZASOBÓW LUDZKICH W POLSCE

Interwencje w obszarze Rozwój Zasobów Ludzkich stanowią istotny element polityki spójności prowadzonej przez Komisję Europejską (KE) i kraje członkowskie Unii Europejskiej (UE). Pozostają one również w zgodzie z głównym celem odnowionej Strategii Lizbońskiej, zakładającej uczynienie z Europy najbardziej konkurencyjnej gospodarki, a tym samym bardziej atrakcyjnego miejsca do lokowania inwestycji i podejmowania pracy. Jednym z kluczowych czynników realizacji Strategii jest wzrost zatrudnienia, osiągany przez rozwój kapitału ludzkiego i społecznego. W przypadku Polski, założenia do prowadzenia interwencji w obszarze polityki zasobów ludzkich zostały określone w Narodowych Strategicznych Ramach Odniesienia (NSRO). NSRO, odpowiadając celom wspólnotowym, identyfikuje jako kluczowy problem „pełne i dobre wykorzystanie jakości kapitału ludzkiego i zasobów pracy, w taki sposób, by służyło to wzrostowi gospodarczemu, konkurencyjności i innowacyjności polskiej gospodarki, zmniejszeniu dysproporcji w rozwoju pomiędzy Polską a UE, oraz poszczególnymi regionami" (MRR, 2007(2): 49). NSRO zakłada, iż całość interwencji w obszarze rozwoju zasobów ludzkich w latach 2007–2013 będzie realizowana przez Program Operacyjny Kapitał Ludzki (PO KL). Konsekwentnie, cel główny PO KL został zdefiniowany jako „umożliwienie pełnego wykorzystania potencjału zasobów ludzkich przez wzrost zatrudnienia i potencjału adaptacyjnego przedsiębiorstw, podnoszenie kwalifikacji ich pracowników, podniesienie wykształcenia społeczeństwa, zmniejszanie obszarów wykluczenia społecznego oraz wsparcie budowy struktur administracyjnych państwa" (MRR, 2007(1): 5).

Analizując wspomniane dokumenty programowe, widzimy wyraźnie, że interwencje w obszarze rozwoju zasobów ludzkich rozumiane są

raczej szeroko i obejmują swym zakresem działania mające, między innymi, na celu wsparcie:

- wzrostu poziomu edukacji społeczeństwa oraz poprawy jakości kształcenia;
- prowadzenia aktywnej polityka rynku pracy;
- tworzenia warunków sprzyjających przedsiębiorczości;
- przeciwdziałania ubóstwu i zapobiegania wykluczeniu społecznemu;
- wzmocnienia potencjału zdrowotnego kapitału ludzkiego.

W Polsce, podobnie jak w części krajów UE, zwłaszcza tzw. nowych krajach członkowskich (UE12), interwencje w obszarze rozwoju zasobów ludzkich finansowane są w przeważającej mierze ze środków EFS. Fundusz ten jest najstarszym i jednym z najważniejszych instrumentów finansowych dostępnym dla krajów UE. Wsparcie finansowane z EFS koncentruje się na projektach z obszaru rynku pracy, służących poprawie zdolności do zdobycia i utrzymania zatrudnienia, promowaniu przedsiębiorczości i równych szans. Obecnie EFS wspiera ponad 9 milionów osób rocznie w skali całej UE, dysponując budżetem wielkości 76,2 mld EUR (EC, 2007(1): 2).

4. PRZEWIDYWANE REZULTATY INTERWENCJI W OBSZARZE ZASOBÓW LUDZKICH

Szeroki zakres interwencji w obszarze rozwoju zasobów ludzkich sprzyja definiowaniu bardzo bogatego katalogu oczekiwanych rezultatów. W przypadku interwencji finansowanych ze środków EFS, jest on jednakże ograniczony do efektów bezpośrednio i pośrednio związanych z rynkiem pracy. W skali mikro, do rezultatów interwencji w obszarze zasobów ludzkich należą m.in.: poprawa sytuacji osób uczestniczących w projektach (wsparcie prowadzi do integracji społeczno-zawodowej wykluczonych); uczestnicy są lepiej przygotowani na zmiany i nowe wyzwania na rynku pracy. W skali makro natomiast, dąży się do osiągnięcia większej adaptacyjności rynku pracy i osób do zachodzących zmian, większej integracji społecznej oraz podniesienia konkurencyjności zasobów ludzkich.

W przypadku działań w obszarze polityki zatrudnienia i rynku pracy bezpośrednie efekty programów definiowane są głównie przez pryzmat ich uczestników. Do najczęściej rejestrowanych rezultatów należą: podjęcie pracy lub kontynuowanie szkolenia przez beneficjentów oraz uzyskanie dyplomów i certyfikatów. Określenie wpływu, a zwłaszcza wpływu netto interwencji, wydaje się bardziej skomplikowane. W niektórych przypadkach, badacze podejmują próbę wyceny kosztów i korzyści wsparcia jednego beneficjenta w podjęciu pracy, np. w ocenie *mid-term* angielskich programów finansowanych z EFS (KE, 2006: 16). Także

rezultaty działań w obszarze integracji społecznej określane są najczęściej w kategoriach podjęcia oraz rodzaju zatrudnienia. W przypadku działań z obszaru polityki edukacji, głównymi rezultatami są zdobycie lub podniesienie posiadanych kwalifikacji, ale także formalne uznanie umiejętności, które nie są jednoznacznie sklasyfikowane.

Ewaluacja tych efektów nie jest jednakże zawsze jednoznaczna. Analizując wpływ programu z punktu widzenia mikro, uczestnicy programu, na przykład, mają mniej czasu na szukanie pracy, co może spowodować ich mniejsze szanse na rynku pracy po zakończeniu wsparcia; także sama perspektywa udziału w projekcie może wpływać demotywująco do szukania zatrudnienia. Przyjmując perspektywę makro, nie można zapominać, iż udział osób bezrobotnych w programie nie zmniejsza automatycznie ogólnej liczby bezrobotnych, choć często zdarza się, iż beneficjenci wsparcia pomijani są w statystykach publicznych (Pierre 1999). Należy również zauważyć, że to, w jakich kategoriach definiowane są zakładane rezultaty programu, ma zasadniczy wpływ na proces oceny oraz koszt samej ewaluacji.

5. ROZWÓJ EWALUACJI INTERWENCJI W OBSZARZE ROZWOJU ZASOBÓW LUDZKICH

Mimo że EFS wdrażany był w krajach ówczesnej Wspólnoty Europejskiej od samego jej początku w 1957 roku, wymóg systematycznego monitorowania oraz ewaluacji, początkowo jedynie *ex ante*, programów finansowanych z tych środków, wprowadzony został dopiero w ramach pierwszej reformy funduszy strukturalnych w okresie 1989–1993. John Bachtler (2007) zauważa, że reforma ta była znaczącym krokiem rozpoczynającym rozwój ewaluacji w UE, w tym również rozwój oceny interwencji w obszarze zasobów ludzkich. Kolejna reforma funduszu, w 1993 roku, wzmocniła rolę monitorowania przebiegu interwencji oraz zainicjowała koncepcję ewaluacji *mid-term*, prowadzonej w połowie okresu wdrażania programów. Duże znaczenie, nie tylko dla ewaluacji, ale także dla samej realizacji działań w obszarze rozwoju zasobów ludzkich, miało wprowadzenie wymogu definiowania mierzalnych celów już na etapie planowania interwencji. We wcześniejszym okresie, a zwłaszcza przed pierwszą reformą funduszy, przyznawanie pomocy odbywało się bowiem na podstawie indywidualnych aplikacji, bez ustrukturalizowanego planowania działań ani określania priorytetów w obszarze rozwoju zasobów ludzkich, które podzielane byłyby w ramach danego regionu lub całej Wspólnoty.

W latach 2000–2006 ewaluacja została wprowadzona jako wymóg formalny, zarówno przed, w trakcie, jak i po zakończeniu działań finansowanych z funduszy strukturalnych. Odgórne określenie przez

KE harmonogramu i zakresu ewaluacji *mid-term* spowodowało jednak ograniczenie jej przydatności i jakości przez niedostosowanie terminu i zakresu badania do rzeczywistych potrzeb poszczególnych krajów członkowskich. Na podstawie tego doświadczenia, ewaluacja, która jest prowadzona w latach 2007–2013, ma mieć charakter *on-going*, odpowiadając w sposób bardziej elastyczny na potrzeby informacyjne osób zarządzających programami i decydentów. W sposób znaczący zmieniło się również podejście do realizowania wsparcia w obszarze rozwoju zasobów ludzkich. Wielokrotnie podkreślano w czasie negocjacji programów operacyjnych konieczność bardziej strategicznego planowania interwencji w tym obszarze, tak aby przez ich realizację możliwe było osiąganie nadrzędnych celów, czy to na poziomie regionu, kraju, czy też całej Wspólnoty. Także sposób prowadzenia ewaluacji będzie musiał ulec stopniowym modyfikacjom, tak aby umożliwić dostarczanie wiedzy na temat poziomu osiągniętych celów strategicznych interwencji.

6. GŁÓWNE WYZWANIA EWALUACJI INTERWENCJI W OBSZARZE ROZWOJU ZASOBÓW LUDZKICH

Kolejne reformy i zmiany wymagań wobec ewaluacji działań finansowanych z EFS nie rozwiązały jednak głównych problemów, które napotyka przede wszystkim KE, ale także kraje członkowskie przy próbie oceny efektów interwencji w obszarze rozwoju zasobów ludzkich. W przypadku poziomu europejskiego, największym wyzwaniem jest konieczność syntezy i agregacji wyników z różnych poziomów ewaluacji EFS; zaczynając od pojedynczych projektów, grup projektów, programów krajowych, kończąc na perspektywie europejskiej. Agregacja taka wydaje się niezbędna, zwłaszcza w przypadku interwencji finansowanych przez instrumenty wspólnotowe, które wyznaczają określone cele do osiągnięcia. Tymczasem cele zdefiniowane w poszczególnych krajowych programach operacyjnych różnią się zdecydowanie, co wynika z różnorodności problemów, z jakimi spotykają się poszczególne państwa.

Każda synteza, niezależnie od poziomu, na którym jest prowadzona, wymaga dostępu do wiarygodnych danych, opierając się głównie na systemie monitorowania programów. W przypadku interwencji w obszarze zasobów ludzkich ten warunek nie jest i nie może być w pełni spełniony. Po pierwsze, mając na uwadze charakter interwencji, nie istnieje jeden, powszechnie stosowany katalog danych, które powinny być monitorowane przez kraje członkowskie w ramach wdrażania EFS. Należy przy tym podkreślić, iż decyzja o nietworzeniu takiego katalogu była podjęta w sposób świadomy, biorąc pod uwagę zróżnicowanie potrzeb i problemów w obszarze rozwoju zasobów ludzkich, z którymi spotykają się poszczególne kraje UE. Istnieją wprawdzie odrębne systemy wskaźników,

w trzech obszarach problemowych: zatrudnieniu[1], edukacji i szkoleniach[2] oraz integracji społecznej[3], ale mają one charakter wskaźników kontekstowych i służą raczej do mierzenia osiągania strategicznych celów Wspólnoty. Powoduje to, że porównywalność rezultatów interwencji pomiędzy różnymi krajami jest raczej ograniczona. Dodatkowo, brak jednolitych definicji oraz metodologii zbierania wskaźników, które niejednokrotnie formułowane w bardzo różny sposób w poszczególnych krajach UE, mogą prowadzić do skrzywionej analizy i, tym samym, błędnych wniosków.

Ograniczenia, z którymi spotyka się ewaluacja interwencji w obszarze rozwoju zasobów ludzkich, mogą wynikać również bezpośrednio ze sposobu funkcjonowania systemu monitorowania programów, stanowiącego jedno z podstawowych źródeł danych dla oceny. Typowymi danymi zbieranymi w procesie monitorowania są: liczba uczestników z określonych grup docelowych, status na rynku pracy po zakończeniu udziału w projekcie, kwalifikacje osiągnięte wskutek udziału w projekcie. Tymczasem błędy generowane przez system monitorowania zaczynają się praktycznie już na etapie wprowadzania danych przez osoby do tego zobowiązane, najczęściej projektodawców. Często wynikają one z nieprecyzyjnych definicji wskaźników, prowadzących do stosowania swobodnych interpretacji, co skutkuje posiadaniem nieporównywalnych wzajemnie danych. Ze względu na charakter zbieranych danych o beneficjentach, które, zgodnie z wymogami KE, obejmują również informacje prywatne, między innymi: wiek, wykształcenie, historię zawodową oraz przynależność do grup szczególnego ryzyka, w systemach monitorowania rejestrowany jest stosunkowo duży odsetek braku danych. Innym błędem, charakterystycznym dla interwencji w obszarze zasobów ludzkich jest wielokrotne rejestrowanie tych samych beneficjentów, którzy uczestniczą w więcej niż jednym projekcie, nazywane „podwójnym liczeniem".

Także współistnienie wielu instrumentów wdrażanych jednocześnie w obszarze rozwoju zasobów ludzkich powoduje, że trudno jest oddzielić wpływ poszczególnych instrumentów, na przykład interwencji finansowanych ze środków EFS, od efektów innych programów krajowych. Tym bardziej skomplikowane jest uchwycenie komplementarności pomiędzy poszczególnymi programami. Jednocześnie, zwłaszcza na poziomie wspólnotowym, decydenci oczekują od ewaluacji wykazania, nie tylko w jaki sposób przekazane środki zostały wykorzystane przez kraje członkowskie, realizując cele związane z polityką zatrudnienia, zdrowia

---

[1] Council Decision of 12 July 2005 on Guidelines for the Employment Policies of the Member States (2005/600/EC).

[2] Council Conclusions of 24 May 2005 on New Indicators In Education and Training (2005/C 141/04).

[3] European Commission, DG Employment (2006), *Social Affairs and Equal Opportunities*, Portfolio of overarching indicators and streamlined social inclusion, pensions and health portfolios.

i integracji społecznej, ale również jaka była wartość dodana tych interwencji. Dyskusja, w jaki sposób oceniana powinna być wartość dodana (*European Added Value*) toczy się już od wielu lat, na razie jednak proponowana metodologia ogranicza się do aspektów takich jak wpływ na sposób prowadzenia polityki, uczenie się, wymiana doświadczeń z innymi krajami, rozwój administracji (Maraite 2007).

Ostatnim ograniczeniem, o którym należy wspomnieć w tym kontekście, a które zostanie w sposób bardziej dokładny przedstawione w dalszych częściach rozdziału, jest trudność definicji relacji przyczynowo-skutkowej w przypadku interwencji o charakterze miękkim, czyli w programach z obszaru rozwoju zasobów ludzkich. Odgadnięcie, czy też raczej przeprowadzenie metodologicznie poprawnej analizy, która będzie w stanie wykazać, iż zarejestrowane efekty są wyłącznym rezultatem prowadzonej interwencji, wydaje się niezwykle trudne. W przypadku tego typu interwencji osiągnięte rezultaty są dużo bardziej podatne na wpływ zmian w otoczeniu społeczno-ekonomicznym niż na przykład inwestycje w infrastrukturę. Ewaluacja wymaga zatem dobrze dobranych wskaźników, które są „wrażliwe" na zmiany zachodzące pod wpływem programu. Jednocześnie te ze wskaźników, które mimo że rejestrują zachodzące zmiany, ale nie potrafią oddzielić ich od innych zmian, niebędących specyficznymi dla programu, nie powinny być wykorzystywane w monitorowaniu efektów interwencji w obszarze rozwoju zasobów ludzkich.

7. EWALUACJA WPŁYWU INTERWENCJI W OBSZARZE ROZWOJU ZASOBÓW LUDZKICH

Celem analizy wpływu interwencji w obszarze rozwoju zasobów ludzkich jest określenie, czy i w jaki sposób realizacja konkretnych programów i projektów przyczyniła się do osiągnięcia celów polityki (wpływ na poziomie makro) oraz czy objęcie wsparciem spowodowało polepszenie sytuacji osób biorących w nich udział (wpływ na poziomie mikro). Istnieje wiele typów efektów identyfikowanych jako rezultat interwencji w tym obszarze, wśród nich: bezpośredni wpływ na beneficjantów (efekty związane z zatrudnialnością i wysokością zarobków, poprawą sytuacji materialnej i zadowoleniem z życia), efekty społeczne (poprawa środowiska społecznego, np. przez zmniejszenie poziomu przestępczości), wpływ na rynek pracy (poprawa dopasowania wymogów rynku pracy do systemu kształcenia, zmiany w stopie bezrobocia, zmiany charakterystyk osób bezrobotnych, zwiększenie produktywności), finanse publiczne (sposób wydatkowania środków publicznych). Większość z obecnie stosowanych metod mierzenia wpływu ma jednak charakter częściowy i nie skupia się na analizie wszystkich możliwych efektów programów (Pierre 1999).

Główne wyzwania związane z mierzeniem efektów interwencji w obszarze rozwoju zasobów ludzkich zostały omówione w poprzednim podrozdziale, należy jednak podkreślić, że głównym problemem jest uchwycenie ich rzeczywistego wpływu. Jak wspomniano, wpływ programu jest bowiem wynikiem zarówno wewnętrznych, jak i zewnętrznych wobec programu czynników, często nieprzewidzianych na etapie planowania, oraz ich wzajemnych interakcji. Mierzenie efektów w postaci linearnej, za pomocą identyfikacji relacji przyczynowo-skutkowych, zdaje się mieć bardzo ograniczone zastosowania w przypadku tego typu interwencji.

W ostatnich latach obserwuje się coraz częstszy zwrot ku metodom, zwłaszcza quasi-eksperymentalnym, które umożliwiałyby analizę wpływu programów w obszarze polityki rynku pracy, polityki społecznej i edukacyjnej. Popularność tej metodologii tłumaczona może być specyfiką samych interwencji, które przez definiowanie konkretnej grupy odbiorców działań, umożliwiają konstruowanie porównań z grupami osób niepoddanych interwencji, tzw. grupami kontrolnymi. Ta sytuacja, z którą porównuje się wyniki osób poddanych interwencji, określana jest jako kontrfaktyczna. Zjawiska rejestrowane w ramach sytuacji kontrfaktycznej określa się jako potencjalne rezultaty, które mogłyby być osiągnięte, gdyby program nie został wdrożony (Purdon 2002: 5). Przyjęcie tego stosunkowo prostego rozumowania budzi jednak wiele wątpliwości u badaczy, którzy podkreślają niejednoznaczność sytuacji kontrfaktycznej w przypadku interwencji w obszarze rozwoju zasobów ludzkich. Nie jest bowiem wcale oczywiste, czy jest nią sytuacja, w której wdrażany byłby inny program, czy też raczej sytuacja, w której niewdrażany byłby żaden program. Równie trudne wydaje się określenie ram czasowych prowadzonej analizy. Co więcej, metody stosowane do ewaluacji wpływu zawężają analizę do właściwej programowi grupy odbiorców. Tymczasem, możliwym i częstym zjawiskiem w przypadku tych interwencji jest to, że programy oddziaływają również na inne grupy, nieobjęte bezpośrednio interwencją.

Aby zapewnić jak największą wiarygodność i jakość analizy, grupa kontrolna powinna posiadać identyczne charakterystyki jak osoby uczestniczące w projekcie. Zapewnienie tej identyczności nie jest jednak zadaniem prostym. Po pierwsze, beneficjenci programu nie uczestniczą w nim przez cały czas, a zatem możliwe są sytuacje, w których uprzedni beneficjenci stają się potencjalnymi członkami grupy kontrolnej. Po drugie, sami beneficjenci, poza udziałem w konkretnym, badanym programie mogą równolegle uczestniczyć w innych działaniach i projektach, nieznanych ewaluatorowi. Problem ten został zilustrowany przez Heckmana (Boer 2003: 13), który oszacował, iż pomiędzy 5 a 36% beneficjentów programów analizowanych w próbie zrezygnowało z udziału w projekcie w trakcie trwania badania, natomiast 3 do 55% z grupy kontrolnej uczestniczyło w programach bardzo podobnych do badanego.

Dobór grupy kontrolnej stanowi również problem ze względu na to, iż w przypadku programów z obszaru rozwoju zasobów ludzkich, na beneficjentów oddziaływają inne, niż tylko fakt uczestniczenia w programie, czynniki. Na przykład, rezultatem projektu, którego celem jest wsparcie osób bezrobotnych i umożliwienie im powrotu na rynek pracy, byłoby podjęcie pracy przez beneficjenta. Jednakże na to, czy ten cel został osiągnięty, ma wpływ wiele innych czynników osobistych beneficjenta, takich jak kwalifikacje, historia zatrudniania, warunki ekonomiczne, sytuacja rodzinna, motywacja do znalezienia pracy. Być może pewnym rozwiązaniem jest próba zdefiniowania sytuacji kontrfaktycznej przez samych beneficjentów, czyli wyobrażenia sobie, w jakiej sytuacji byliby w momencie badania, gdyby nie uczestniczyli w projekcie. To podejście jednak obarczone jest także dużym błędem, wynikającym z subiektywności analizy.

8. PRZYKŁADY METOD DO MIERZENIA WPŁYWU INTERWENCJI NA POZIOMIE MAKRO

Spośród wielu istniejących metod mierzenia wpływu, w tej sekcji przedstawione są te, które są najczęściej wykorzystywane w ewaluacji interwencji w obszarze rozwoju zasobów ludzkich.

Ewaluacja *ex post* interwencji

Mowa tutaj o typowej ewaluacji *ex post*, która znajduje swoje zastosowanie w przypadku większości programów w obszarze rozwoju zasobów ludzkich. Ocena wpływu interwencji dokonywana jest w sposób przede wszystkim jakościowy, na podstawie danych zebranych za pomocą klasycznych narzędzi badań społecznych, takich jak: studia przypadków, wywiady pogłębione, grupy fokusowe, analizy danych zastanych. Ten typ analizy jedynie w ograniczonym zakresie pozwala na ocenę specyficznych efektów interwencji, np. *deadweight*.

Analiza typu „przed – po"

W przypadku tej metody, ocena wpływu interwencji dokonywana jest przez porównanie sytuacji wyjściowej, mierzonej na początku wdrażania programu (tzw. *baseline*) z rezultatami zarejestrowanymi po zakończeniu realizacji interwencji. Pomimo większych możliwości analitycznych, analiza typu „przed-po" obciążona jest dużym błędem, spowodowanym brakiem możliwości odseparowania rezultatów będących bezpośrednią konsekwencją interwencji od zmian zachodzących w otoczeniu społeczno-ekonomicznym oraz sytuacji osobistej beneficjenta.

Metody quasi-eksperymentalne

Metody quasi-eksperymentalne opierają się na zastosowaniu analizy porównawczej dwóch populacji: beneficjentów oraz osób niebiorących udziału w projekcie. W tym celu do zbiorowości beneficjentów dobierana jest grupa osób, które nie brały udziału w programie, charakteryzująca się tymi samymi cechami w zakresie wybranych zmiennych, a jedyną różnicą pozostaje fakt brania udziału w programie. Jedną z metod doboru populacji, stosowaną w quasi-eksperymentach jest *prospentity score matching*. Zarejestrowana różnica pomiędzy osiągnięciami uczestników programu a grupy niebędącej beneficjentami programu daje obraz wpływu interwencji. Stosowanie metod quasi-eksperymentalnych wiąże się niejednokrotnie z prowadzeniem badań ilościowych na dużą skalę oraz zastosowaniem zaawansowanej analizy statystycznej. Zaletą tej metody jest to, iż dobór osób odbywa się niejednokrotnie *ex post*, a zatem badanie jako takie, nie ma wpływu na faktyczny udział lub jego brak w programie. Konieczny jest natomiast pełen rygoryzm w doborze grup, dokonywany nie tylko ze względu na cechy takie jak: płeć, wiek czy wykształcenie, ale również aspekty mniej uchwytne, jak aktywność w poszukiwaniu pracy.

Metody eksperymentalne (*randomised trial*)

Polegają na przeprowadzeniu analizy porównawczej dwóch populacji, beneficjentów objętych programem i grupy kontrolnej, pozostającej poza wpływem programu, z tą różnicą, że dobór do tych populacji na charakter losowy. Podejściem alternatywnym do losowego doboru beneficientów i nie-beneficjentów, jest losowy dobór obszarów (regionu, dzielnicy), które są desygnowane bądź jako objęte programem, bądź jako grupa kontrolna. Niezaprzeczalną przewagą, jaką dysponuje to podejście w stosunku do metod quasi-eksperymentalnymi, jest możliwość wykluczenia innych, niż fakt udziału w programie, czynników wpływających na sytuację badanych osób. Z drugiej strony, metody eksperymentalne wzbudzają wiele wątpliwości natury etycznej, wynikających z konieczności odmowy wsparcia pewnej grupie potencjalnych beneficjentów. Prowadzenie dużej skali eksperymentów pociąga za sobą również ograniczenia administracyjne i wysokie koszty.

Modele ekonomiczne

Użycie modeli ekonomicznych pozwala na zrekonstruowanie rzeczywistości i prowadzenie analiz wpływu interwencji o charakterze prognostycznym. Pomimo istnienia wielu typów modeli ekonomicznych, ich użycie i dokładność analizy są uzależnione od typu prowadzonej interwencji oraz dostępności danych statystycznych. Na przykład, model

makro-ekonometryczny QUEST, stosowany jest do analizy wpływów interwencji w obszarze zasobów ludzkich przez Komisję Europejską, tymczasem w Polsce większość badań prowadzona jest na podstawie modelu HERMIN (EC, 2007(3)).

Spośród omówionych metod analizy wpływu, za najbardziej skuteczne w ocenie dodatkowości interwencji w obszarze rozwoju zasobów ludzkich uważane są metody eksperymentalne. W literaturze określane są nawet jako „złote standardy" ewaluacji (Prudon, 2002: 20). Praktyka badawcza pokazuje jednak, że stosowanie metodologii eksperymentalnej wiąże się z wieloma ograniczeniami i w rezultacie tylko nieliczne programy rynku pracy i zasobów ludzkich stosują metodologię eksperymentalną. Analizując w tym kontekście sytuację w Wielkiej Brytanii, gdzie metody eksperymentalne są stosunkowo najszerzej wykorzystywane, Susan Purdon identyfikuje czynniki, z powodu których metody eksperymentalne nie są stosowane na szerszą skalę. Wiążą się one głównie z niejasnością, na ile takie badania są etyczne, wysokich kosztów prowadzenia eksperymentów wynikających z konieczności wprowadzenia odpowiednich procedur losowego doboru próby oraz uzyskania zgody na badanie od wszystkich beneficjentów i nie-beneficjentów. Nie należy również zapominać, że wprowadzając nawet najbardziej rygorystyczne założenia metodologiczne, nie da się uniknąć błędów wynikających z charakterystyki samych osób badanych, np. ich chęci i aktywności w poszukiwaniu pracy. Najważniejszym jednak czynnikiem, na który Purdon zdaje się zwracać uwagę, jest skuteczne wykorzystywanie przez badaczy metod quasi-eksperymentalnych (Purdon 2002).

Metody te nie udzielają jednak odpowiedzi na typowe pytania ewaluacyjne. Otrzymujemy wynik – wpływ netto programu, pozbawiony oceny głównych czynników zmiany oraz podziału na osoby, którym udział w programie pomógł i na te, którym przeszkodził. Interesującym pytaniem ewaluacyjnym, na które odpowiedzi nie udzieliłyby metody eksperymentalne, jest pytanie o jakość zatrudnienia, np. która z tych dwóch grup pozostaje dłużej w miejscu pracy.

9. PRZYKŁADY METOD MIERZENIA WPŁYWU NA POZIOMIE MIKRO – MIERZENIE MIĘKKICH REZULTATÓW PROJEKTÓW W OBSZARZE ROZWOJU ZASOBÓW LUDZKICH

Analiza na poziomie mikro dostarcza informacji o zmianach, jakich doświadczają bezpośrednio beneficjenci na skutek otrzymania wsparcia w ramach konkretnego programu. Interwencje realizowane w obszarze rozwoju zasobów ludzkich mogą prowadzić do zmian dwojakiego rodzaju, określanych w literaturze jako „twarde" i „miękkie" rezultaty. Rezultaty „twarde" są jednoznacznie definiowalne i policzalne. Wśród nich

znajdziemy zdobycie kwalifikacji, znalezienie pracy, znalezienie stałego miejsca zamieszkania, uczestnictwo w dalszym szkoleniu. Jednak, w przypadku tego typu interwencji zawężenie analizy jedynie do „twardych" rezultatów nie pozwala na pełną identyfikację wyników programu. Ograniczenie to związane jest przede wszystkim z rodzajem udzielanego wsparcia oraz charakterystyką beneficjentów. Osoby uczestniczące w programach z obszaru rozwoju zasobów ludzkich to niejednokrotnie osoby doświadczające pewnych trudności w pełnej integracji społecznej; wśród nich znajdujemy długotrwale bezrobotnych, niepełnosprawnych, młodzież o niskich kwalifikacjach.

W tym kontekście uzasadnione jest twierdzenie Jane Pitcher, że w przypadku analizy programów mających na celu budowę kapitału ludzkiego, prowadzenie ewaluacji jedynie za pomocą tradycyjnych, pozytywistycznych metod badawczych, w których wnioskowanie oparte jest na poszukiwaniu relacji przyczynowo-skutkowej, nie jest wystarczające, a prowadzić nawet może do błędnych wniosków. Metody ilościowe nie pozwalają na ujawnienie możliwego niedostosowania programu do potrzeb beneficjentów i zdolności beneficjentów do osiągnięcia oczekiwanych w programie rezultatów (Pitcher 2002: 486). Pitcher proponuje zwrot ku paradygmatowi ewaluacji, który oparty byłby na bliskiej komunikacji z beneficjentami i który odwoływał się do rezultatów programu widzianych przez pryzmat jego uczestników. W praktyce, zmiana metodologii oznacza wprowadzenie dodatkowych, jakościowych, wskaźników, które umożliwiałyby identyfikację mniej oczywistych rezultatów osiąganych przez beneficjentów na różnych etapach wdrażania programu.

W tym celu stworzona została kategoria rezultatów „miękkich". Należą do nich między innymi: wzrost samooceny, poprawa zdolności komunikacyjnych, rozwój umiejętności zarządzania czasem, a także wzrost pewności siebie. Już na podstawie wymienionych przykładów „miękkich rezultatów" można wnioskować, że zdefiniowanie mierzalnych wskaźników służących ich pomiarowi jest trudne, ponieważ dotyczą one postaw, umiejętności i innych cech, których istnienie stwierdzone może być jedynie w drodze specyficznych badań czy obserwacji (DWP & WEFO, 2003: 5). Wydaje się jednak, że rezygnacja z oceny wskaźników „miękkich" w programach z obszaru rozwoju zasobów ludzkich, pozbawia zarówno beneficjentów, jak i zarządzających programami możliwości analizy rzeczywistych rezultatów prowadzonych działań. W przypadku pewnych grup beneficjentów osiągnięcie rezultatów „miękkich" jest jedynym rejestrowanym sukcesem projektu, ponieważ nie są oni w stanie osiągnąć „twardych" rezultatów, takich jak podjęcie pracy po zakończeniu interwencji. Dlatego, w sytuacji braku mierzalnych wyników programu w osiąganiu celów, wskazane jest, aby zweryfikować cele oraz sposób ich określenia, czego konsekwencją może być wprowadzenie dodatkowych celów pośrednich, tzw. *milestones,* niezbędnych do osiągnięcia w dalszej perspektywie tego celu ostatecznego.

Analiza miękkich rezultatów zdaje się odgrywać coraz większą rolę w projektach z obszaru rozwoju zasobów ludzkich. Wybór właściwego narzędzia do mierzenia miękkich rezultatów powinien zależeć od celu danego projektu, specyfiki grupy docelowej, posiadanego zaplecza, a także od rodzaju wskaźników, które zostały wybrane do mierzenia. Badając rezultaty miękkie, należy również przyjąć pewien punkt wyjściowy, który będzie odniesieniem w momencie późniejszego badania osiągniętego przez beneficjenta pomocy postępu (WEFO, 2004). Trudnością jednak jest jednoznaczne ocenienie, w jakim stopniu program i konkretne działania, w których uczestniczyli beneficjenci, doprowadziły do osiągnięcia konkretnych wyników. Niezbędne jest także dostosowanie narzędzia badawczego do możliwości kognitywnych beneficjenta. Ponadto należy brać pod uwagę możliwość zachwiania wyników przez brak obiektywizmu w prowadzonej ocenie. Dodatkową zaletą wynikającą z mierzenia miękkich rezultatów jest możliwość sprawdzenia, czy podjęte w projekcie działania są odpowiednie dla danej grupy beneficjentów ostatecznych, a jeżeli nie, to jakie udoskonalenia należy wprowadzić do realizacji projektu. Ostatnim odbiorcą informacji na temat miękkich rezultatów jest instytucja wdrażająca, która ma możliwość dostrzeżenia, że realizowany przez beneficjenta projekt daje pozytywne efekty, nawet jeżeli rezultaty twarde, wyrażone we wniosku projektowym za pomocą wskaźników liczbowych, nie zostały jeszcze osiągnięte.

Z mierzeniem „miękkich rezultatów" wiąże się jednak również wiele wyzwań. Po pierwsze, jak było to wielokrotnie poruszane w tym rozdziale, ze względu na charakter interwencji oraz grupę odbiorców, niezwykle trudno jest jednoznacznie określić relację przyczynowo-skutkową pomiędzy wsparciem, które jest udzielone beneficjentowi, a osiągniętymi rezultatami. Jednym z wyjść wydaje się zdefiniowanie przez samego beneficjenta sytuacji kontrfaktycznej, czyli próba oceny, w jakiej sytuacji znalazłby się beneficjent, gdyby nie uczestniczył w programie. Po drugie, zdecydowanie częściej niż w przypadku innych typów analiz, istnieje niebezpieczeństwo formułowania subiektywnych ocen przez badaczy. Ograniczenie to jest konsekwencją stosowanej metodologii mierzenia „miękkich" rezultatów, która w większości przypadków opiera się w jakimś stopniu na obserwacji prowadzonej przez projektodawców oraz rozmowach z beneficjentami. Brak rygoryzmu metodologicznego, charakterystycznego dla narzędzi ilościowych, sprzyjać może wysnuwaniu mniej obiektywnych wniosków. Przede wszystkim jednak, mierzenie „miękkich" rezultatów jest procesem nie tylko czasochłonnym, ale również wymagającym zapewnienia odpowiednich zasobów ludzkich po stronie projektodawcy. Każdy beneficjent wymaga bowiem indywidualnego określenia punktu wyjścia (*baseline*), a następnie systematycznego rejestrowania swoich postępów.

Pomimo rosnącej roli tej metodologii, nie można również zapomnieć o tym, że badanie „miękkich" rezultatów ma charakter mikro, jest analizowane na poziomie każdego projektu z osobna i tylko w bardzo

Przykłady metodologii stosowanej do mierzenia miękkich rezultatów w Walii

W 2006 roku na zlecenie Welsh European Funding Office (WEFO) przeprowadzono badanie mające na celu zidentyfikowanie i skatalogowanie metod mierzenia „miękkich" rezultatów w projektach finansowanych z EFS. Analizą objęto ponad 30 projektów realizowanych w ramach dwóch Programów Operacyjnych: West Wales and the Valleys (cel 1) oraz East Wales (cel 3). Wyniki badania pokazały, że projektodawcy w zdecydowanej większości przywiązywali dużą wagę do pomiaru rezultatów „miękkich", chociaż zgłaszali brak doświadczenia w tym zakresie. Metody pomiaru, którymi się posługiwano, można podzielić na kilka kategorii:

- metody oparte na obserwacji beneficjentów, uzupełnione wywiadami z beneficjentami ostatecznymi, prowadzone regularnie przez trenerów, w sposób formalny bądź nieformalny;
- metody oparte na indywidualnych planów działań, które m.in. uwzględniały cele rozwoju osobowego, określały podstawowy poziom umiejętności, a przez regularne przeglądy przedstawiały stopień realizacji celów i osiągniętego sukcesu. Indywidualne plany działań uzupełniane były przeglądem dokonywanym przez beneficjentów ostatecznych w formie kwestionariuszy na różnych etapach trwania projektu
- metody oparte na testach psychometrycznych, służących zarówno diagnozowaniu poziomu początkowego, jak i ocenie postępu beneficjenta w zwiększaniu zdolności do bycia zatrudnionym. Do takich narzędzi należały: stosunkowo popularna Rickter Scale, narzędzie Catching Confidence stosowane przez NIACE oraz Spirit Level.

W projektach, które podjęły się oceny „miękkich" rezultatów, osiąganych przez swoich beneficjentów, powszechne było dokonywanie pomiaru przynajmniej dwukrotnie: na początku i na końcu projektu. Umożliwiło to porównanie wyników uzyskanych w kolejnych ocenach, a przez to lepsze zrozumienie zmian zaistniałych u beneficjentów. Kilkukrotna ocena w trakcie trwania projektu pozwoliła niektórym projektodawcom na bardziej systematyczną rejestrację osiąganego postępu w osiąganiu miękkich rezultatów, nazywanego również tzw. przebytą drogą. Te rejestrowane kroki podnoszą szansę tychże beneficjentów na integrację społeczną, uzyskanie stałego zatrudnienia lub podniesienie kwalifikacji, które stanowią ostateczny etap projektu.

ograniczonym zakresie możliwe wydaje się agregowanie tych efektów na poziomie wyższym, na przykład w skali całego programu (WEFO, 2003: 5). Z tego względu, pomiar rezultatów „miękkich" nie powinien być wykorzystywany do porównywania osiągnięć różnych projektów, a tym bardziej służyć jako podstawa do wyciągania wniosków na temat ich skuteczności i efektywności.

10. PODSUMOWANIE

Celem tego rozdziału było podjęcie dyskusji na temat specyfiki ewaluacji interwencji w obszarze rozwoju zasobów ludzkich. Jak zostało to przedstawione, aby mówić o całościowym procesie ewaluacji w tym przypadku, należy wziąć pod uwagę co najmniej trzy elementy. Po pierwsze, analizuje się rezultaty interwencji ma poziomie mikro, czyli ocenia się bezpośredni wpływ programu na sytuację beneficjentów. Głównym pytaniem stawianym na tym etapie ewaluacji jest to, w jakim stopniu zarejestrowane rezultaty były spowodowane uczestniczeniem beneficjenta w programie. Po drugie, ewaluacja odbywa się na poziomie makro, analizując pośredni wpływ interwencji na osiąganie celów zdefiniowanych dla całego obszaru rozwoju zasobów ludzkich. W tym przypadku szczególna uwaga poświęcona jest ocenie efektów netto, czyli sprawdzeniu, na ile osiągnięte rezultaty są wynikiem realizacji interwencji i nie miałyby miejsca, gdyby nie zrealizowano programu. Po trzecie, dopełniając oceny, konieczne jest zweryfikowanie, czy zidentyfikowane rezultaty są proporcjonalne do zainwestowanych zasobów.

Ewaluacja, a zwłaszcza ta dotycząca bardzo kompleksowych polityk, tak jak w przypadku interwencji w obszarze zasobów ludzkich, nie jest zadaniem jednoznacznym i prostym. Badanie ewaluacyjne jest procesem stopniowego dobierania i dostosowywania metodologii, która odpowiadałaby kompleksowym założeniom polityki i umożliwiałaby zrozumienie szerzej definiowanej odpowiedniości interwencji, co jest przeciwieństwem stosowania stosunkowo prostych metod w obszarze niektórych programów, na przykład infrastrukturalnych. Ta odpowiedniość rozumiana jest przez Sandersona (2000: 450) jako identyfikacja efektów lub konsekwencji interwencji w odniesieniu do poszczególnych aktorów, zrozumienia, o czym świadczą owe efekty i w końcu oceny, nadania wartości interwencji w kontekście szerszych celów i wartości polityki.

Obszar zasobów ludzkich jest jednym z bardziej kompleksowych, wieloplaszczyznowych i dynamicznych obszarów, co powoduje bardzo konkretne i wymierne skutki dla prowadzenia ewaluacji. Jak pisze Sanderson, aby umożliwić uczenie się na podstawie wyników badań, ewaluacja musi być bardziej poszukująca i wyjaśniająca. Dostępne modele liniowe, obrazujące sytuacje przyczynowo-skutkowe, nie mają w tym przypadku większego zastosowania, a ich ustalanie przed rozpoczęciem wdrożenia

programu prowadzi do niepełnych rezultatów. Niesłuszne jest też izolowanie pojedynczego programu od całej sytuacji zewnętrznej, a zwłaszcza zmieniającego się kontekstu społeczno-ekonomicznego, w którym realizowany jest program. Wyzwaniem, które wprawdzie nie jest wyłącznie problemem ewaluacji w obszarze zasobów ludzkich, jest zatem konieczność zmierzenia rzeczywistego wpływu interwencji. Innymi słowy, potrzebna jest metodologia, która umożliwi określenie, jakie byłyby efekty, jeśli nie prowadzono by programu (sytuacja kontrfaktyczna), które z rezultatów mogłyby być osiągnięte niezależnie od tego, czy program byłby realizowany, oraz zmierzenie bezpośrednich i pośrednich efektów samej interwencji.

BIBLIOGRAFIA

Bachtler J., Michie R. (2007), A new era in EU Regional Policy Evaluation? The apprisl of the Structural Funds, *Regional Studies* **29.8**, 745–752.

Boer M. (2003) *Estimating the impact of Employment Programmes on Participant's outcomes*, Centre for Social Research and Evaluation.

Department for International Development (DFID) (2005), *A sourcebook Poverty and Social Impacts Analysis (PSIA)*, Vol. 1: *Tools for Institutional. Political and Social Analysis (TIPS)*, London.

European Commission (EC) (2007(1)), Directorate General Employment and Social Affairs, Information note – The European Social Fund – needs and performance.

European Commission (EC) (2007(2)), Directorate General Employment and Social Affairs, Note to the file on ESF and macro economic models used by the Commission.

European Commission (EC) (2007(3)), Directorate General Regional Policy, Note to the file on impact evaluation in cohesion policy programmes.

European Commission (EC) (2006), Directorate General Employment and Social Affairs, *Overview of the final evaluations of the ESF co-funded programmes.*

European Commission (EC) (1999), Directorate General Employment and Social Affairs, *The Guidelines for Systems of Monitoring and Evaluation of ESF Assistance in the Period 2000–2006*, http://www.europa.eu.int/comm/employment_social/evaluation/esf08_en.html

Lion C. (2006), Evaluating the implementation process: A contribution within the framework of the European Social Fund (ESF) Programme, *Evaluation* **12(3)**, 313–329.

Mairaite A. (2006), The added value of the European Union Cohesion Policy, *Regional Studies* **40.2**, 167–177.

Ministerstwo Rozwoju Regionalnego (MRR) (2007(1)), *Program Operacyjny Kapitał Ludzki 2007–2013*, Warszawa.

Ministerstwo Rozwoju Regionalnego (MRR) (2007(2)), *Narodowe Strategiczne Ramy Odniesienia (NSRO) 2007–2013* i *Narodowa Strategia Spójności*, Warszawa.

Pierre G. (1999), *A framework for active labour market policy evaluation*, University of Warwick.

Pitcher J. (2002), Policies and Programmes to Address Disadvantage among Young People: Issues for Evaluation, *Evaluation* **8(4)**, 474–495.

Purdon S. (2002), *Research Methods for Policy Evaluation*, Department for Work and Pension (DWP), Research Working Paper No. 2.

Purdon S. (2003), *Estimating the impact of labour market programmes*, Department for Work and Pension (DWP), Research Working Paper No. 3.

Sanderson I. (2000), Evaluation in Complex Policy Systems, *Evaluation* **6(4)**, 433–454.

Welsh European Funding Office (WEFO) and Department for Work and Pension (DWP) (2003), *A Practical Guide to Measuring Soft Outcomes and Distance Travelled* – Guidance Document.

ROZDZIAŁ 6.5.

Grzegorz Gorzelak
Dominika Wojtowicz

OCENA WPŁYWU NPR I NSRO NA ROZWÓJ REGIONALNY I PRZESTRZENNY ORAZ ZWIĘKSZENIE SPÓJNOŚCI REGIONALNEJ I PRZESTRZENNEJ

Jednym ze strategicznych obszarów tematycznych, uwzględnionych w ramach systemu ewaluacji Narodowego Planu Rozwoju (NPR) oraz Narodowych Strategicznych Ram Odniesienia (NSRO), jest rozwój regionalny i przestrzenny oraz zwiększenie spójności w obu tych wymiarach. Odpowiednie zaplanowanie i przeprowadzenie badań ewaluacyjnych w tym obszarze wymaga poznania szerszego kontekstu zagadnień z zakresu rozwoju regionalnego i przestrzennego oraz zapoznania się z problematyką różnych teorii interwencji publicznych, mających na celu wspieranie rozwoju społeczno-gospodarczego regionów. Próba oceny działań podejmowanych w ramach polityki regionalnej przy współudziale środków unijnych, a także stopnia osiągnięcia przez nich pożądanych rezultatów, może się powieść jedynie w przy umiejętnym zdiagnozowaniu problemów i logiki skutkowo-przyczynowej zjawisk współcześnie wpływających na charakter i tempo rozwoju w wymiarze regionalnym i przestrzennym naszego kraju. Wiąże się to ze zrozumieniem specyfiki polskiej przestrzeni i czynników ją kształtujących w warunkach

współczesnej gospodarki, która charakteryzuje się zaostrzającą się konkurencją. Jest to szczególnie istotne dla podjęcia właściwego toku myślenia i właściwej interpretacji zjawisk badanych w ramach procesu ewaluacji. Ponadto należy zaznaczyć, że dobór wskaźników rozwoju, metod i narzędzi badawczych ma charakter wtórny w stosunku do tez i pytań badawczych wysuwanych przez ewaluatorów. Niniejszy artykuł koncentruje się więc na zagadnieniach związanych z pożądanym i możliwym do realizacji modelem rozwoju regionalnego i przestrzennego. W pierwszej części zdefiniowane zostaną podstawowe pojęcia, których zrozumienie warunkuje prawidłowe podejście do ich ewaluacji. Dalej przedstawione zostaną obszary, które zgodnie z Planem Ewaluacji NPR i NSRO, powinny ulegać pozytywnym zmianom w wyniku realizacji interwencji publicznych. W ostatniej części podjęta zostanie próba wskazania kwestii kluczowych i potencjalnych ograniczeń istotnych dla prawidłowej realizacji procesu ewaluacji.

1. ROZWÓJ REGIONALNY I PRZESTRZENNY A SPÓJNOŚĆ REGIONALNA I PRZESTRZENNA

Przedmiotem ewaluacji w ramach omawianego obszaru jest terytorialny wymiar wdrażania Narodowego Planu Rozwoju i Narodowych Strategicznych Ram Odniesienia, a konkretnie ocena wpływu realizacji założeń obu dokumentów na rozwój regionalny i przestrzenny kraju oraz zwiększanie spójności regionalnej i przestrzennej kraju. Tak postawione cele wymagają zdefiniowania użytych pojęć, a więc zastanowienia się, czym jest rozwój regionalny i przestrzenny, a także na czym polega zwiększanie spójności w obu tych wymiarach.

W większości współczesnych koncepcji rozwój regionalny definiowany jest jako proces prowadzący do trwałych zmian o charakterze zarówno ilościowym, jak i jakościowym (w tym takich jak zmiany w strukturze zatrudnienia ludności, poprawa jakości warunków życia mieszkańców, większa dostępność do dóbr kulturowych, nauki, rekreacji i sportu, poprawa warunków środowiskowych i ogólnego bezpieczeństwa), zachodzący na terenie danego regionu. Z terminem rozwoju regionalnego często utożsamiane jest pojęcie wzrostu gospodarczego (który dotyczy zmian w ilościowo mierzalnych wartościach, z których najczęściej wykorzystywanym jest PKB), w rzeczywistości stanowiące jedynie część składową całości zjawiska (por. Przygodzki 2007). Ekonomika rozwoju regionalnego zależy od wielu czynników, jednak im większy jest udział czynników o znaczeniu egzogenicznym, tym większy jest potencjał decydujący o zdolności do efektywnego konkurowania. Można przyjąć, że rozwój regionalny związany jest z trwałym wzrostem potencjału gospodarczego i społecznego regionów oraz poprawą ich konkurencyjności.

Rozwój przestrzenny dotyczy natomiast pozytywnych zmian w układzie funkcjonalno-przestrzennym, w tym w jego poszczególnych elementach, takich jak system osadniczy – a jego ramach układy metropolitalne – czy sieci transportowe i komunikacyjne.

Więcej niejasności rodzi próba zdefiniowania terminu „spójność regionalna i przestrzenna". Najczęstsza interpretacja „spójności" polega na utożsamianiu jej z „konwergencją", czyli zbliżaniem się do siebie poziomów różnych kategorii w krajach czy regionach – w tym głównie poziomu PKB na mieszkańca dochodów ludności, wskaźników odnoszących się do rynku pracy, poziomu zaawansowania technologicznego i zdolności do wytwarzania innowacji itp. „Konwergencja" jest zapisana w najważniejszych traktatach określających integrację europejską. Nawet jeżeli ten termin w nich nie występuje, to mówi się w nich o dążeniach do wyrównywania poziomów rozwoju między regionami Unii (wcześniej Wspólnoty) Europejskiej przez pomaganie regionom zapóźnionym i przeprowadzającym restrukturyzację gospodarki. Przy przyjęciu takiego rozumienia „spójności" naturalnym – i dobrze uzasadnionym przez ugruntowaną wiedzę o rozwoju regionalnym – wnioskiem jest teza o sprzeczności między dążeniem do konkurencyjności a uzyskiwaniem spójności. Sprzeczność ta jest pochodną znanego w literaturze dylematu „równość czy wydajność", wskazującego, że efektywność inwestowania na obszarach o wyższym poziomie rozwoju jest z reguły wyższa niż na obszarach relatywnie gorzej rozwiniętych.

Utożsamianie spójności z konwergencją jest jednak niewłaściwe. O ile konwergencję możemy rozumieć jako proces zmierzający do zbliżania się stanów statycznych (ujmowanych w określonych punktach czasowych), o tyle spójność powinniśmy traktować jako cechę procesów, zachodzących w odcinkach czasu. Są to więc terminy odwołujące się do innych kategorii opisu procesów gospodarczych, społecznych i przestrzennych. Dylemat „równość czy wydajność" powinno więc odnosić się do sprzeczności między wzrostem a konwergencją. Relacje między rozwojem a spójnością są znacznie bardziej złożone.

Spójność można przeciwstawić entropii. Spójność jest niezbędną cechą organizmów i organizacji – bez jej uzyskania i zachowania organizmy te i organizacje przestałyby istnieć. Spójność ma w sobie wymiar harmonii i współpracy, braku destruktywnych napięć i nierozwiązywalnych konfliktów, możliwości współistnienia i współdziałania poszczególnych elementów całości. W odniesieniu do takich układów jak region, spójność tak rozumiana ma kilka wymiarów. Najczęściej wymienia się trzy rodzaje spójności: gospodarczą, społeczną i przestrzenną.

Wymiar gospodarczy spójności odnosi się do możliwości efektywnej współpracy między poszczególnymi elementami systemu ekonomicznego, do obniżania kosztów transakcyjnych, do harmonizowania relacji między podmiotami gospodarczymi i ich instytucjonalnym otoczeniem. Spójność gospodarcza oznacza harmonizację funkcjonowania całego

złożonego układu ekonomicznego i możliwie pełne wykorzystywanie potencjału tkwiącego w jego częściach składowych. Osiąganie spójności regionalnej w wymiarze gospodarczym polegać więc będzie na wspieraniu systemu, w którym funkcjonują przedsiębiorstwa (w tym instytucji okołobiznesowych), a mierzone może być m.in. zwiększeniem udziału eksportu w PKB regionu, napływem kapitału zagranicznego, powstaniem skupisk i gron, zwiększeniem zakresu kooperacji przedsiębiorstw i ich otoczenia, zmniejszeniem kosztów transakcyjnych.

Istotą spójności społecznej jest dążenie do eliminowania barier w mobilności poziomej i pionowej, polegającego na zapewnieniu możliwości przezwyciężania różnic w poziomie wykształcenia, awansu zawodowego i materialnego, łatwości poruszania się po obszarze poszczególnych regionów. Osiąganie spójności społecznej jest wspomagane przez solidaryzm społeczny i redystrybucyjne funkcje władz publicznych różnych szczebli, oraz realizację projektów ukierunkowanych na podnoszenie jakości kapitału ludzkiego i społecznego regionów. Miernikami spójności społecznej mogą być np. podniesienie poziomu wykształcenia młodego pokolenia, poprawa na rynku pracy, zmniejszenie skali ubóstwa, zmniejszenie przestępczości.

Osiąganie spójności przestrzennej (zarówno w wymiarze regionalnym, jak i krajowym) należy natomiast rozumieć jako eliminowanie barier i ograniczeń wynikających z zagospodarowania przestrzennego, ograniczających możliwości uzyskiwania spójności gospodarczej i społecznej. Spójność przestrzenną można więc osiągać przez tworzenie sprawnych i niezawodnych systemów transportowych, funkcjonujących na różnych szczeblach terytorialnych (międzynarodowym, krajowym, regionalnym i lokalnym), a także przez rozwijanie powszechnie dostępnych sieciowych powiązań telekomunikacyjnych. Gęstość wyposażenia danego terenu w infrastrukturę transportową i jej parametry powinna być oczywiście dostosowana do rzeczywistych potrzeb gospodarczych i społecznych, co pozwoli na uzyskanie racjonalności nakładów ponoszonych na jej budowę i eksploatację. Wskaźnikami odzwierciedlającym osiąganie spójności przestrzennej mogą być np.: poprawa powiązań transportowych między najważniejszymi ośrodkami miejskimi kraju oraz między nimi a ich zapleczem regionalnym; upowszechnienie dostępu do sieci telekomunikacyjnych.

Należy jednak pamiętać, że nie wszystkie inwestycje, których celem jest zwiększanie spójności przestrzennej są korzystne z punktu widzenia gospodarek lokalnych danego regionu. Budowanie dróg szybkiego ruchu (autostrady i drogi ekspresowe) i szybkiej kolei może mieć efekty dwojakiego rodzaju: z jednej strony zbliża do siebie punkty, w których znajdują się przystanki kolejowe oraz zejścia z autostrady – z drugiej zaś relatywnie oddala od nich miejsca położone między tymi węzłami, czyniąc je względnie mniej dostępnymi (występuje tu tzw. efekt tunelu). Nie

zawsze więc szybka kolej czy droga szybkiego ruchu, przecinająca dany teren – ale nieobsługująca go bezpośrednio – jest dla niego czynnikiem korzystnym i przyczyniającym się do jego rozwoju. Dążenie do uzyskania spójności na poziomie międzynarodowym czy krajowym (międzyregionalnym) może zmniejszyć szanse rozwojowe obszarów położonych między węzłami, a tym samym przyczynić się do zmniejszenia spójności w skali intraregionalnej.

Zgodnie z zapisami Narodowego Planu Rozwoju na lata 2007–2013 wyróżnia się trzy wymiary polityki spójności:

- **spójność społeczną** – rozumianą jako zmniejszenie zróżnicowań w wykorzystaniu zasobów ludzkich między poszczególnymi regionami, i mierzoną za pomocą stopy zatrudnienia;
- **spójność gospodarczą –** rozumianą jako zmniejszenie zróżnicowań w poziomie rozwoju gospodarczego pomiędzy regionami bogatymi i biednymi, i mierzoną za pomocą produktu krajowego brutto na mieszkańca, przy uwzględnieniu parytetu siły nabywczej;
- **spójność przestrzenną –** rozumianą jako eliminowanie barier dostępności regionów peryferyjnych przez ich lepsze powiązanie z obszarami centralnymi, mierzoną czasem przejazdu do danego obszaru komunikacją lotniczą, drogową i kolejową.

Źródło: http://www.funduszestrukturalne.gov.pl/informator/npr2/npr/NPR_3.doc

2. WPŁYW INTERWENCJI PUBLICZNYCH PODEJMOWANYCH W RAMACH NPR I NSRO NA WYBRANE OBSZARY ROZWOJU REGIONALNEGO I PRZESTRZENNEGO

Rozpoczynając rozważania nt. ewaluacji działań podejmowanych na rzecz wpierania rozwoju regionalnego i przestrzennego, należy zastanowić się nad pożądanym modelem rozwoju gospodarczo-społecznego, możliwym do zrealizowania w naszym kraju. Główne kierunki rozwoju, wspierane w ramach realizacji polityki regionalnej i przestrzennej, znalazły odzwierciedlenie w celach i ujętych w ich ramach strategicznych obszarach NPR oraz NSRO.

Zgodnie z wytycznymi Ministerstwa Rozwoju Regionalnego, dotyczącymi prowadzenia ewaluacji, ocenie poddawana będzie realizacja założeń wraz z oddziaływaniem na szersze otoczenie obydwu wieloletnich planów strategicznych w zakresie: wspomagania udziału w procesach

rozwojowych i modernizacyjnych wszystkich regionów i grup społecznych w Polsce (Cel 5 NPR), wzrostu konkurencyjności polskich regionów i przeciwdziałania ich marginalizacji społecznej, gospodarczej i przestrzennej (Cel 5 NSRO) oraz wyrównywania szans rozwojowych i wspomagania zmian strukturalnych na obszarach wiejskich (Cel 6 NSRO). Plan ewaluacji NSRO na lata 2007–2013 opracowany przez Krajową Jednostkę Oceny Departamentu Koordynacji Polityki Strukturalnej Ministerstwa Rozwoju Regionalnego, doprecyzowuje zakres oceny wpływu NSRO o konkretne obszary interwencji, do których zaliczono:

- zmniejszanie marginalizacji i peryferyzacji regionów – ewaluacji podlegać będą procesy zmniejszania dysproporcji rozwojowych polskich regionów na poziomie kraju oraz w stosunku do innych regionów UE, ze szczególnym uwzględnieniem Polski wschodniej oraz obszarów problemowych,
- spójność wewnętrzną obszarów miejskich – ocenie w tym zakresie podlegać będą procesy ograniczające wysoką koncentrację problemów gospodarczych, ekologicznych oraz społecznych wewnątrz dużych ośrodków miejskich oraz procesy dotyczące wzmacniania odbudowy gospodarczego potencjału małych miast i miasteczek,
- zwiększenie potencjału endogenicznego ośrodków miejskich – ewaluacji w tym zakresie podlegać będą procesy rozwoju funkcji metropolitalnych (naukowych, edukacyjnych, kulturalnych i turystycznych),
- rozprzestrzenienie się impulsów rozwojowych z obszarów metropolitarnych i terenów zurbanizowanych na obszary przyległe – zarówno małe i średnie miasta, jak i obszary wiejskie – ewaluacji w tym zakresie podlegać będą procesy budowy powiązań komunikacyjnych i teleinformatycznych, a także gospodarczych, usługowych i kulturalnych,
- rozwój współpracy terytorialnej: transgranicznej, ponadnarodowej i międzyregionalnej,
- kształt przestrzennego zagospodarowania kraju i regionów,
- wyrównywanie szans rozwojowych i wspomaganie zmian strukturalnych na obszarach wiejskich,
- spójność terytorialną kraju w aspekcie wdrażania zasady zrównoważonego rozwoju (Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013, s. 14–15).

Takie ujęcie obszarów, które mają ulegać pozytywnym zmianom w wyniku działań podejmowanych w ramach realizacji polityki regionalnej i przestrzennej, wymaga kilku uwag – w trakcie prowadzenia procesu ich ewaluacji niezbędne jest uwzględnienie specyfiki niektórych z wyróżnionych obszarów. W dalszej części niniejszego rozdziału podjęte zostaną rozważania na temat ograniczeń oraz kwestii problematycznych i kluczowych dla prawidłowego przeprowadzenia procesu ewaluacji, które wynikają z zachodzących współcześnie zjawisk i ich tendencji.

3. WZROST KONKURENCYJNOŚCI POLSKICH REGIONÓW

Oceniając wpływ realizacji NSRO na wzrost konkurencyjności regionu, należy mieć na uwadze, że region jest konkurencyjny, jeżeli spełnia dwa dualne względem siebie warunki: jest w stanie przyciągnąć kapitał, szczególnie kapitał innowacyjny, oraz stwarza takie warunki zlokalizowanym na jego terenie przedsiębiorstwom, że są one w stanie wygrać konkurencję. Stopień spełnienia tych warunków zależy od zakresu, w jakim cechy regionu są zgodne z obowiązującymi w danym paradygmacie rozwoju kryteriami lokalizacji, a więc potrzebami firm względem ich otoczenia. Regiony oferujące podmiotom gospodarczym warunki, dzięki którym podmioty te mogą uzyskać przewagę konkurencyjną w skali, w jakiej funkcjonują (a więc także w skali globalnej), same stają się konkurencyjne w walce o kapitał, szczególnie o kapitał niosący inwestycje w segmencie innowacyjnym.

Konkurencyjność regionów jest więc stopniowalna, nie jest także niezmienna. Nie zależy wyłącznie od samego regionu, bowiem jest wypadkową jego dość stabilnych i wolno kształtujących się cech społeczno-gospodarczych i przestrzennych oraz zmieniających się (raczej skokowo, niż linearnie) kryteriów lokalizacji. Dla przykładu można wskazać, iż silny w gospodarce zasobochłonnej (przemysłowej) region śląski stał się jednym z regionów problemowych w gospodarce wiedzochłonnej, mimo że sam region się nie zmienił – zmieniły się natomiast uwarunkowania względem niego zewnętrzne.

Z powyższego wynika, iż poszczególne polskie regiony mają silnie zróżnicowane możliwości uzyskania wysokiego poziomu konkurencyjności w globalnej gospodarce, w której trwałą przewagę konkurencyjną uzyskuje się dzięki zdolności do tworzenia innowacji. Najwyższy poziom konkurencyjności mają niewątpliwe wielkie miasta – metropolia warszawska i uzyskujące cechy metropolii Wrocław, Poznań, Kraków i Trójmiasto, natomiast w najmniej korzystnej sytuacji są regiony Polski wschodniej.

4. REALNE MOŻLIWOŚCI OSIĄGANIA SPÓJNOŚCI I ZMNIEJSZANIA NIERÓWNOŚCI ROZWOJOWYCH POLSKICH REGIONÓW

Problematyczne może okazać się badanie wpływu interwencji publicznych na procesy zmniejszania dysproporcji rozwojowych polskich regionów na poziomie kraju oraz w stosunku do innych regionów UE, ze szczególnym uwzględnieniem Polski wschodniej oraz obszarów problemowych. Przystępując do badań ewaluacyjnych z tego zakresu, należy

dokładnie poznać obecne zróżnicowania polskiej przestrzeni i ich przyczyny. Według najnowszych badań zmniejszanie nierówności praktycznie nie możliwe, bo dysproporcje występują i będą występować nawet w sytuacji ponoszenia znacznych nakładów na działania prorozwojowe regionów o niższym poziomie rozwoju społeczno-gospodarczym (por. np. Boldrin, Canova 2000, Sapir et al. 2003, Rodriguez-Pose, Fratesi 2004). Można postawić hipotezę, że zróżnicowanie rozwoju na poziomie regionalnym jest warunkiem do osiągnięcia wyższego poziomu rozwoju w kontekście całego kraju. Dlatego tak postawiony cel NSRO można uznać za nieosiągalny i sprzeczny z dążeniem do przyspieszenia rozwoju gospodarczego Polski. Różnice międzyregionalne nie tylko nie zmniejszą się – przeciwnie, będą wzrosnąć. Istotne jest to, by regiony najsłabsze osiągały pozytywne tempo wzrostu, by mogły wyrwać się ze stagnacji, a nawet z przedłużającej się recesji.

Polska należy do państw o wysokim wewnętrznym zróżnicowaniu poziomu rozwoju regionów. Zróżnicowania te, wbrew powszechnej opinii, nie są jednak większe niż w innych krajach Unii Europejskiej. Stosunek PKB województw o najwyższym i najniższym poziomie rozwoju gospodarczego wynosi ok. 2:1 i jest zbliżony do poziomu zróżnicowań, jakie występują pomiędzy regionami we Francji czy w Hiszpanii. Cechą charakterystyczną tego wymiaru nierówności jest podział Wschód–Zachód, często nazywany podziałem na Polskę „A" i „B". Regiony położone na zachód od linii Wisły odnotowują wyższe tempo rozwoju, są lepiej przygotowane do korzystania z szans związanych z funkcjonowaniem w gospodarce globalnej. Natomiast wschodnie województwa, których potencjał w znacznej mierze znajduje się w fazie przedprzemysłowej i charakteryzujące się wysokim poziomem zatrudnienia w rolnictwie, nie są w stanie sprostać wyzwaniom współczesności, co wpływa na wolniejsze tempo (a w niektórych przypadkach nawet regres) ich rozwoju. Te regiony w większym stopniu odczuły negatywne skutki przemian ustrojowych zachodzących po 1989 roku. Warto zaznaczyć, że w opinii wielu regionalistów ten wymiar zróżnicowań jest silnie uwarunkowany długotrwałymi procesami historycznymi, mającymi swoje korzenie już w średniowieczu. Podział na Wschód–Zachód został utrwalony w czasie rozbiorów, kiedy poszczególne regiony znalazły się w granicach oddziaływania zupełnie odmiennych organizmów państwowych, ekonomicznych i kulturowych (por. np. Gorzelak, Jałowiecki 1998, Bartkowski 2003, Hryniewicz 2004).

Na problem ograniczeń w zmniejszaniu dysproporcji rozwojowych zwracają uwagę także autorzy Koncepcji Przestrzennego Zagospodarowania Kraju pisząc, że „za najważniejszą dyrektywę przestrzennego zagospodarowania kraju należy więc uznać utrzymanie i wzmocnienie policentryczności struktury polskiej przestrzeni (…)" (Zaktualizowana Koncepcja Przestrzennego Zagospodarowania Kraju, Rządowe Centrum

Studiów Strategicznych 2005, s. 68). Pomimo podejmowanych przez rząd i władze regionalne działań, polaryzacja regionalna będzie w dalszym ciągu postępowała, tak jak miało to miejsce w całym okresie wzrostu polskiej gospodarki po 1992 roku, i jak to ma miejsce we wszystkich krajach bloku postsocjalistycznego. Co ciekawe, podobne procesy zachodzą także w innych krajach Unii Europejskiej, które pomimo odrabiania dystansu względem państw lepiej rozwiniętych, doświadczają pogłębiania się różnic w poziomie rozwoju społeczno-gospodarczego na poziomie regionów.

Dlatego realizacja założeń NPR i NSRO negatywnym tendencjom – takim jak np. stagnacja Polski wschodniej – mogą przeciwstawić się w jedynie niewielkim stopniu. Zadaniem obu tych strategii nie może być „zawracanie" trendów uznanych za niekorzystne, ale ich łagodzenie oraz wzmacnianie tendencji uznanych z korzystne.

Należy mieć na względzie, że regiony Polski wschodniej w dalszym ciągu będą speryferalizowane tak w Unii Europejskiej, jak i w kraju, pozbawione korzystnych impulsów z zewnątrz, przy zbyt niskim własnym potencjale endogennym, uniemożliwiającym efektywne nadrabianie dystansu do jednostek lepiej rozwiniętych. W regionach tych nasilać się będą procesy wypłukiwania zasobów ludzkich, które w północnej części Polski wschodniej będą prowadzić do wyludniana się niektórych obszarów, a w części południowej – zmniejszania przeludnienia wsi.

5. WZMACNIANIE OŚRODKÓW METROPOLITARNYCH ORAZ SPRZYJANIE DYFUZJI IMPULSÓW ROZWOJOWYCH NA OBSZARY PRZYLEGŁE

Podstawowa rola obszarów metropolitalnych w procesach rozwoju kraju oraz w kształtowaniu jego struktury funkcjonalno-przestrzennej jest słusznie podkreślana zarówno w NPR, jak i NSRO. Metropolie są najsilniejszym terytorialnym ogniwem gospodarki globalnej – światowej sieci metropolitalnej, w skład której wchodzą duże miasta skupiające funkcje zarządcze i kontrolne, dostarczające podstawowej części innowacji oraz tworzące wzorce kulturowe o zasięgu globalnym. Miejsce Polski w światowej gospodarce jest i będzie w przyszłości w znacznym stopniu zdeterminowane rolą, jakie polskie największe miasta, w tym głównie Warszawa, będą odgrywać w tej sieci, bowiem polskie metropolie (ukształtowana warszawska i kształtujące się krakowska, poznańska i wrocławska) są najsilniejszymi „łącznikami" Polski z gospodarką światową, szczególnie zaś z jej segmentem innowacyjnym.

Jednak w Planie ewaluacji wśród obszarów, na które wpływ powinny mieć działania podejmowane w ramach realizacji NSRO, wskazuje się jednocześnie na zwiększanie dyfuzji korzystnych procesów rozwojowych z obszarów metropolitalnych do ich otoczenia. Należy brać pod uwagę, że, jak wskazują badania, współczesne polskie metropolie (z których jedynie Warszawę można uznać za metropolię już ukształtowaną) „odrywają się" od swojego zaplecza regionalnego, co osłabia procesy dyfuzji, a wzmacnia zjawiska „wypłukiwania" (*back-wash*) zasobów z regionu do jego centralnego ośrodka (Gorzelak, Smętkowski 2005).

Ponadto najskuteczniejsze wsparcie metropolii związane będzie z interwencjami, które rozwijają relacje tych ośrodków z ich międzynarodowym, krajowym i regionalnym otoczeniem. Mniej efektywne okazuje się wspieranie ich rozwoju „własnego", realizowanego przez władze miejskie (metropolie te są z reguły dość zamożne), w porównaniu z przedsięwzięciami rozwijającymi ich funkcje egzogenne (np. systemy transportu regionalnego, potencjał turystyczny, szczególnie w zakresie turystyki biznesowej – kongresy, wystawy, targi; także potencjał naukowo-badawczy i akademicki, lokowanie siedzib instytucji międzynarodowych itd.). Istotne jest więc zwiększanie potencjału engogenicznego, ale takiego, który przekłada się na wzmocnienie funkcji egzogenicznych. Dopiero tak ukierunkowane wsparcie obszarów metropolitalnych pomoże rozwijać ich funkcje metropolitalne, a także umożliwi wzmocnienie procesów dyfuzji rozwoju do terenów słabiej rozwiniętych.

6. SPÓJNOŚĆ TERYTORIALNA KRAJU W ASPEKCIE WDRAŻANIA ZASADY ZRÓWNOWAŻONEGO ROZWOJU

Podejmując się ewaluacji skuteczności interwencji publicznych na rozwój regionalny i przestrzenny, nie sposób nie odnieść ich do koncepcji zrównoważonego rozwoju, którą zgodnie z definicją przyjętą w *Koncepcji przestrzennego zagospodarowania kraju* można rozumieć jako ingerencję w układy przestrzenne, nawiązującą „do ich naturalnych predyspozycji, z zachowaniem trwałości podstawowych procesów przyrodniczych i związanymi z nimi stanami równowagi ekologicznej w celu zaspokajania potrzeb bieżącego pokolenia i gwarantowania tego zaspokajania przyszłym pokoleniom". Definicja ta sugeruje formę socjoekonomicznego rozwoju, który powoduje wzrost społecznego dobrobytu dla obecnego pokolenia i równocześnie nie ma negatywnego wpływu na zmniejszenie szans na osiągnięcie dobrobytu przez przyszłe pokolenia. Do oceny działań podejmowanych ramach realizacji założeń NPR oraz NSRO pod kątem zgodności z założeniami idei zrównoważonego rozwoju przydatna

może okazać się koncepcja badań oparta na modelu 4 kapitałów. Model ten zastosowany dla wyjaśnienia trwałego, konkurencyjnego rozwoju zakłada, że rozwój społeczno-gospodarczy opiera się na pozytywnych zmianach w zakresie czterech występujących na danym terytorium kapitałach: ludzkim, gospodarczym, społecznym oraz środowiskowym. Rozwój może być traktowany jako trwały tylko wtedy, gdy zasoby kapitałów pozostają stałe, są w równowadze oraz gdy podejmowane w ramach interwencji publicznych działania nie prowadzą do uszczuplenia jednego z kapitałów do tego stopnia, że jego stan staje się krytyczny – trudny do odbudowania i bardzo kosztowny (Etkins 1992).

7. WNIOSKI KOŃCOWE

Proces ewaluacji działań podejmowanych w ramach prowadzonej polityki regionalnej i przestrzennej musi uwzględniać specyfikę i trendy zjawisk zachodzących w polskiej przestrzeni. Ewaluatorzy na wszystkich etapach realizacji badań – od planowania do wyciągnięcia wniosków i formułowania rekomendacji – powinni mieć na względzie pożądany i możliwy do zrealizowania model rozwoju regionalnego i przestrzennego Polski, zapewniający przełamanie jej zapóźnienia cywilizacyjnego. Wzmacnianie (bo osiągnięcie w tak krótkiej perspektywie czasu jest praktycznie niewykonalne) spójności regionalnej i terytorialnej powinno mieć swoje odzwierciedlenie w trwałym, wysokim tempie wzrostu polskiej gospodarki, przy zapewnieniu poprawy stanu środowiska przyrodniczego i umożliwieniu obecnym i przyszłym mieszkańcom kraju równoprawnego dostępu do zasobów przyrody i dóbr kultury.

BIBLIOGRAFIA

Bartkowski J. (2003), *Tradycja i polityka: wpływ tradycji kulturowych polskich regionów na współczesne zachowania społeczne i polityczne*, Warszawa, Instytut Socjologii UW, Wyd. Akademickie Żak.

Canowa F., Boldrin M. (2001), *Europe's Regions. Income disparities and regional policies*, http://www.econ.umn.edu/~mboldrin/Papers/ep_regpol.pdf

Gorzelak G., Jałowiecki B. (1998), *Koniunktura gospodarcza i mobilizacja społeczna w gminach*, Warszawa, Wyd. EUROREG, Uniwersytet Warszawski.

Gorzelak G., Smętkowski M. (2005), *Metropolia i jej region w gospodarce informacyjnej*, Warszawa, Wydawnictwo Naukowe Scholar.

Hryniewicz J. (2004), *Polityczny i kulturowy kontekst rozwoju gospodarczego*, Warszawa, Wydawnictwo Naukowe Scholar.

Kins P. (1992), *A Four-Capital Model of Wealth Creation*, w: Ekins P., Max-Neef M. (eds.) *Real- Life Economics: Understanding Wealth Creation*, London/New York, Routledge.

Przygodzki Z. (2007), *Teoretyczne podstawy rozwoju regionalnego i lokalnego*, w: Chądzyński J., Nowakowska A., Przygodzki Z. (red.), *Region i jego rozwój w warunkach globalizacji*, Warszawa, CeDeWu.

Rodriguez-Pose A., Fratesi U. (2004), Między rozwojem a polityką społeczną – europejskie fundusze strukturalne w regionach Celu 1, *Studia Regionalne i Lokalne* **3(17)**.

Sapir A. et al. (2003), *An agenda for a growing Europe. Making the EU Economic System Deliver*, http://www.euractiv.com/ndbtext/innovation/sapirreport.pdf

PODSUMOWANIE

PERSPEKTYWY I DYLEMATY EWALUACJI

PODSUMOWANIE – PERSPEKTYWY I WYZWANIA EWALUACJI

W niniejszym, ostatnim rozdziale książki, postaramy się krótko zarysować wyzwania i perspektywy ewaluacji programów publiczych w Polsce na najbliższe lata. Kreśląc nasze wnioski, opieraliśmy się zarówno na refleksjach zawartych w rozdziałach książki, doświadczeniach z przedstawianych studiów przypadków, jak i na obserwacjach i doświadczeniach z praktyki autorów niniejszego podręcznika.

Niewątpliwie fundusze strukturalne Unii Europejskiej są dominującym i pozostaną w najbliższych latach głównym polem praktyki ewaluacyjnej w Polsce (Krajowa Jednostka Oceny, 2005, 2008). Jak dotąd zjawisko „rozlewania się" praktyk ewaluacyjnych z programów UE na inne pola polskiego sektora publicznego było znikome. Zapisy prawne sprzed kilku lat, dotyczące ewaluacji kontraktów regionalnych, nie przyjęły się i nie wpłynęły zasadniczo na zmianę tego stanu rzeczy. Nic nie wskazuje też na to, by sytuacja ta miała ulec jakiejkolwiek zmianie. Jest to oczywiście zrozumiałe, jeżeli weźmie się pod uwagę skalę unijnych inwestycji i ich kluczowe znaczenie jako instrumentu działań prorozwojowych. Pozostaje mieć jednak nadzieję, że wiedza zdobyta w ewaluacji programów UE będzie internalizowana przez polskie instytucje i będzie stawać się standardem polskiej administracji publicznej. Jeśli jest to dobre dla programów unijnych, dlaczego nie mogłoby być dobre dla ewaluacji naszych własnych, polskich działań?

Niewątpliwym wyzwaniem jest racjonalizacja zleceń ewaluacyjnych. Ewaluacja jest sztuką zadawania pytań, pytań ważkich – które będą użyteczne dla menedżerów i decydentów sektora publicznego, ale jednocześnie pytań racjonalnych – stawianych we właściwym momencie cyklu życia programu i wyważonych czasem oraz środkami przeznaczonymi

na konkretne badania. I pytań niezamykających się tylko w kręgu kwestii problemów technicznych, ale poszukujących inspiracji i lepszego rozumienia interakcji między aspektem społecznym, gospodarczym i przestrzennym rozwoju. Ta sztuka zadawania precyzyjnych pytań i przekonującego odpowiadania na nie, pozostaje – jak pokazują przytaczane studia przypadków – szerokim polem do doskonalenia się tak zlecających, jak i wykonawców.

Wyzwaniem systemowym nawiązującym do szerszego pola uczenia się organizacji i stabilności systemów jest przekazywanie doświadczeń z kolejnych lat, gromadzenie ich w użytecznej formie w instytucjach zajmujących się ewaluacją, tak by w kolejnych okresach programowania funduszy strukturalnych można – zamiast uczyć się od początku – budować na wcześniejszych doświadczeniach (Ferry, Olejniczak 2008, w druku).

Skoro już przywołaliśmy temat wiedzy, warto na chwilę zastanowić się, jaka wiedza jest zdobywana dzięki dotychczasowym badaniom, a jakiej wciąż brakuje?

Dominuje wiedza o bieżących procesach – techniczna, zorientowana na poprawę wdrażania funduszy. Niewątpliwie – jak pokazują przywoływane studia przypadków – taka wiedza jest coraz bardziej efektywnie zdobywana, zaś polscy uczestnicy ewaluacji (tak zlecający, jak i wykonawcy) są już świadomi, że najlepszym sposobem identyfikowania i rozwiązywania problemów wdrażania programów są dialog i interakcje. Na tym polu wyzwaniem pozostaje doskonalenie warsztatu ewaluatorów i otwartość na współpracę interesariuszy poszczególnych programów. Jest tu także dużo miejsca na rozwijanie funkcji ewaluacji blisko związanej z zarządzaniem – budowanie partnerstwa i współwłasności, uwzględniające także szersze grono stakeholderów. Dialog jak na razie toczy się często w wąskim gronie wykonawców i instytucji zlecającej kontrakt.

Dużo słabiej reprezentowana jest wiedza o szerszych procesach, efektach programów i ich długookresowym oddziaływaniu (oraz o adekwatności założeń polityk i programów do rzeczywistych, osiągniętych zmian) – a więc wiedza spełniająca funkcje poznawcze, rozliczania i wzmacniania odpowiedzialności. Stawianie jasnych hipotez i weryfikowanie relacji przyczynowo-skutkowych jest jak dotąd rzadką praktyką. Z tym idą często w parze wyraźne niedobory warsztatu metodologicznego – dominują metody jakościowe. Dyskusja i praktyka metodologii oceny efektów programów (a więc obserwacji wykraczającej poza produkty) są w zasadzie w stadium początkowym[1]. Podjęto już jednak pierwsze próby dyskusji metodologicznej (por. Haber 2007). Rośnie także zainteresowanie metodami ilościowymi, „pożyczanymi" z pola programów zasobów ludzkich. Być może wkrótce pojawią się w Polsce próby zastosowania

[1] Wstępne wyniki najnowszych badań pokazują, że polska praktyka nie odbiega specjalnie od praktyki innych krajów członkowskich, takich jak Niemcy, Szwecja, Wielka Brytania czy Irlandia (Olejniczak 2009, w przygotowaniu).

podejść quasi-eksperymentalnych. Brakuje jednak rzeczowej refleksji zarówno nad możliwościami racjonalnego oszacowania efektów netto, czy szerzej, oddziaływania programów, jak i wyzwaniami w odniesieniu do programów wielosektorowych i wielocelowych. Uważamy, że można i trzeba podjąć taką refleksję i rozwijać ją, wychodząc poza upraszczający podział (z połowy lat 80.) na przeciwstawne podejścia jakościowe i ilościowe. Byłoby to wskazane tym bardziej, że w literaturze tematu oraz praktyce ewaluacji spoza funduszy strukturalnych pojawiają się już kolejne pasjonujące dyskusje metodologiczne – np. próby analizy systemowej (por. Elsevier Publishers, 2008).

Obserwowane dotąd zachwianie proporcji między orientacją na wdrażanie a orientacją na efekty, można tłumaczyć prostą presją na uczenie się zarządzania według standardów UE, uruchamianiem programów i naciskiem na szybkie zagospodarowanie unijnych środków w danym czasie. Niemniej jednak zjawisko to może być odbiciem szerszego i bardziej niepokojącego problemu polskiej administracji – koncentrowania się na bieżących kwestiach kosztem wizji strategicznej. Hausner (2007) określa to jako koncentrowanie się na absorpcji technicznej (zdolność wydania środków na czas i zgodnie z procedurami) kosztem absorpcji strukturalnej (umiejętność zagospodarowania środków w sposób strategiczny i prorozwojowy). Tę pierwszą wspomagają ewaluacje procesów (wdrażania), tę drugą ewaluacje wpływu (oddziaływania; nazywane również *impact assessment*).

Przytaczane studia przypadków pokazują także, że szczególnie przy ewaluacjach *ex post* istnieje też poważny problem z jakością danych, który rzuca cień na wiarygodność analiz i siłę argumentów płynących z wniosków badań. W ewaluacji, z założenia, punktem wyjścia są dane z monitoringu (na tej podstawie dopiero rozwija i pogłębia się obserwacje) i przejrzysty, stabilny system statystyki publicznej. Brak godnych zaufania, uporządkowanych danych uniemożliwia budowanie spójnych systemów wskaźników, oceny bieżących postępów, a w dalszej perspektywie oceny efektów. Jak dotychczas, praktycznie każda z ewaluacji rozpoczyna się od czasochłonnego i kosztownego uzupełniania czy wręcz tworzenia od początku własnych pakietów danych. Bazy danych zarządzane przez wdrażających progamy winny być przedmiotem ich codziennej troski.

Ostatnim wyzwaniem są relacje między zlecającymi a wykonawcami oraz uporządkowanie standardów działania ewaluatorów – przede wszystkim pola delikatnej równowagi między współpracą a zachowaniem obiektywności badawczej i obowiązujących standardów pracy. Dotychczasowa praktyka – tak po stronie ministerstw, jak i firm czy instytutów badawczych – była mocno intuicyjna. Wyzwaniem byłoby więc stworzenie pewnego kanonu powszechnych zasad bez jednoczesnego organiczania innowacji i popadania w zbytnią schematyczność. To wyzwanie skutecznie podjęło ostatnio Polskie Towarzystwo Ewaluacyjne. Jego najnowsza publikacja oparta na szerokiej dyskusji środowiska

wprowadza Polskę do grona krajów o rozwiniętej kulturze ewaluacyjnej (PTE, 2008). Pozostaje mieć tylko nadzieję, że nakreślone szerokie zasady i kod postępowania upowszechnią się w praktyce.

Podsumowując, dyskutowane przez nas powyżej kwestie funkcji, ograniczeń warsztatu i ról uczestników ewaluacji wpływają wprost na siłę oddziaływania i zakres pozytywnych zmian, jakie ewaluacja może wywołać w rzeczywistości programów publicznych. Rozwijanie warsztatu metodologicznego, gromadzenie doświadczeń praktycznych oraz budowanie pamięci instytucjonalnej, mogą sprawić, że Polska nie tylko sprawniej i efektywniej będzie realizować programy związane z bieżącym okresem planowania (2007–2013), ale też zgromadzi w użytecznej formie doświadczenia, które staną się źródłem przewag konkurencyjnych w dłuższej perspektywie. Współcześnie rozwój jest determinowany przez zdolność do zdobywania, przetwarzania i używania informacji. Ewaluacja może być bezcennym źródłem wiedzy dla całej polityki rozwoju. Pod warunkiem, że – zgodnie z nadziejami łączącymi autorów tego tomu – ewaluacja stanie się codziennością zarządzania interwencją publiczną.

BIBLIOGRAFIA

Elsevier Publishers (2008), *Evaluation and Program Planning* 31(2).

Ferry M., Olejniczak K. (2008, w druku) *The use of evaluation in management of EU programmes – a comparative perspective*, Ernst & Young, Program „Sprawne Państwo", www.sprawnepanstwo.pl

Haber A. (red.) (2007), *Ewaluacja ex post. Teoria i praktyka badawcza*, Warszawa, Polska Agencja Rozwoju Przedsiębiorczości.

Hausner J. (2007), *Czy należy pomagać regionom słabo rozwiniętym?*, Sympozjum Regional Studies Association – Sekcja Polska, Warszawa, 18 kwietnia.

Krajowa Jednostka Oceny (2005), *Rynek ewaluacji w Polsce – diagnoza stanu i perspektyw rozwoju*, I konferencja ewaluacyjna – *Ewaluacja programów o charakterze społeczno-gospodarczym finansowanych z funduszy strukturalnych*. Warszawa, Ministerstwo Gospodarki i Pracy i Polska Agencja Rozwoju Przedsiębiorczości, 16.09.

Krajowa Jednostka Oceny (2008), *Proces ewaluacji polityki spójności w Polsce. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość*, Warszawa, Ministerstwo Rozwoju Regionalnego.

Olejniczak K. (2009, w przygotowaniu), *Praktyka ewaluacji efektów programów rozwoju regionalnego - studium porównawcze (The evaluation practice of effects assessments for regional development programmes – a comparative analysis)*, grant Ministerstwa Rozwoju Regionalnego, Polska, www.euroreg.uw.edu.pl

PTE (2008), *Standardy ewaluacji*, Warszawa, Polskie Towarzystwo Ewaluacji, http://www.pte.org.pl/forum/viewtopic.php?t=32

INFORMACJA O AUTORACH

REDAKTORZY PODRĘCZNIKA

Marek Kozak – doktor nauk społecznych; adiunkt w Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG) na Uniwersytecie Warszawskim. Od 1991 roku był dyrektorem w Urzędzie Pełnomocnika Rządu ds. Integracji Europejskiej i Pomocy Zagranicznej, następnie prezesem Polskiej Agencji Rozwoju Regionalnego. Ekspert w zakresie polityki spójności, rozwoju regionalnego i lokalnego, turystyki w rozwoju oraz ewaluacji. Autor kilkudziesięciu prac poświęconych problematyce polityki regionalnej, działaniu administracji publicznej, politykom publicznym. Współpracował z agendami rządowymi i samorządami lokalnymi i regionalnymi w Polsce oraz w wieloma instytucjami międzynarodowymi (takimi jak Bank Światowy, Komisja Europejska, UNDP). Członek Komitetu Przestrzennego Zagospodarowania Kraju PAN oraz członek Prezydium Państwowej Rady Gospodarki Przestrzennej.

Bartosz Ledzion – ekonomista, absolwent Akademii Leona Koźmińskiego, ze specjalizacją w zarządzaniu projektami finansowanymi ze środków UE. Ukończył studia podyplomowe z zakresu zarządzania projektami w Szkole Głównej Handlowej w Warszawie. Od 3 lat jest koordynatorem merytorycznym studiów podyplomowych w zakresie ewaluacji w Akademii Leona Koźmińskiego. Ma 6-letnie doświadczenie w zakresie badań ewaluacyjnych i audytów organizacyjnych w jednostkach administracji publicznej. Uczestniczył w realizacji kilkudziesięciu projektów doradczych i badawczych. Specjalizuje się w badaniach o charakterze procesowym, ewaluacjach potencjału instytucjonalnego oraz audytach organizacyjnych w jednostkach administracji publicznej. Jest współwłaścicielem firmy badawczej EGO S.C.

Karol Olejniczak – doktor nauk o zarządzaniu, ukończył studia magisterskie europejskiej administracji publicznej (European Institute of Public Administration, Holandia) oraz gospodarki przestrzennej (EUROREG – Uniwersytet Warszawski). Realizował wiele ewaluacji

programów funduszy strukturalnych i przedakcesyjnych. Badania z zakresu metod i technik ewaluacji prowadził w USA, Wielkiej Brytanii, Austrii i Irlandii. Jest autorem ekspertyz i publikacji na temat ewaluacji interwencji publicznych. Głównym polem jego zainteresowań naukowych są: nowoczesne zarządzanie w sektorze publicznym, rozwój regionalny oraz współpraca terytorialna. Pracuje jako adiunkt w Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG) na Uniwersytecie Warszawskim. Jest także dyrektorem ds. Strategii i Rozwoju w firmie ewaluacyjnej EGO S.C.

AUTORZY PODRĘCZNIKA

John Bachtler – profesor nauk ekonomicznych, kieruje European Policies Research Centre (EPRC) University of Strathclyde, Glasgow. Specjalizuje się w problematyce rozwoju regionalnego i lokalnego, obejmującej przede wszystkim kwestie polityki regionalnej na poziomie zarówno krajów członkowskich, jak i krajów kandydujących oraz polityki spójności w Unii Europejskiej. Ponadto pełni funkcję dyrektora programu badań porównawczych nad wdrażaniem funduszy strukturalnych w ramach Celu 1 i 2 (IQ – Net). Na jego dorobek składa się autorstwo i redakcja kilkudziesięciu książek oraz ponad kilkuset artykułów i rozdziałów w książkach poświęconych problematyce polityki regionalnej, polityki spójności i ewaluacji. Pełni także funkcję członka Komitetu Naukowego w Instytucie Federalnym Zagospodarowania Przestrzennego w Bonn-Bad, Niemcy.

Stanisław Bienias – absolwent Szkoły Głównej Handlowej w Warszawie oraz podyplomowych studiów z zakresu ewaluacji projektów na Uniwersytecie Warszawskim. Od czterech lat pracuje w Ministerstwie Rozwoju Regionalnego, od 2006 roku na stanowisku naczelnika Wydziału Ewaluacji (Krajowa Jednostka Oceny). Odpowiedzialny za realizację procesu ewaluacji na poziomie Narodowego Planu Rozwoju na lata 2004–2006 i Narodowych Strategicznych Ram Odniesienia na lata 2007–2013 oraz za koordynację procesu ewaluacji na poziomie programów operacyjnych realizowanych w latach 2004–2006 i 2007–2013. Brał udział w realizacji ponad 60 badań ewaluacyjnych, był organizatorem i prelegentem licznych konferencji i seminariów dotyczących budowy kultury ewaluacyjnej w naszym kraju. Reprezentuje Polskę w spotkaniach „Evaluation Network" w ramach Dyrekcji Generalnej do Spraw Regionalnych w Komisji Europejskiej, współpracuje z OECD i innymi instytucjami krajowymi i zagranicznymi. Jeden z ekspertów obecnie realizowanego twinningu Niemiecko-Polsko-Rumuńskiego. Prowadzi wykłady na studiach podyplomowych z zakresu ewaluacji, monitoringu i wskaźników. Współautor kilku publikacji w obszarze ewaluacji i funduszy strukturalnych.

Anna Domaradzka – asystent w Instytucie Studiów Społecznych Uniwersytetu Warszawskiego. Ma wieloletnie doświadczenie w realizacji kompleksowych badań społecznych, rynkowych oraz ewaluacji dla międzynarodowych firm badawczych i doradczych. W swojej działalności badawczej specjalizuje się w obszarze polityki równouprawnienia. Ponadto ma doświadczenie i umiejętności w przygotowaniu i wykorzystaniu metodologii badań jakościowych i ilościowych. Jest autorką licznych publikacji i recenzji naukowych w Polsce i za granicą.

Martin Ferry – pracownik naukowy European Policies Research Centre (EPRC) University of Strathclyde, Glasgow. W swojej działalności badawczej specjalizuje się w obszarze rozwoju ekonomicznego i regionalnego krajów Europy Środkowo-Wschodniej. Jego najnowsze projekty dotyczyły analiz polityki regionalnej oraz restrukturyzacji przemysłu w Republice Czeskiej i Polsce. Brał także udział w badaniu ewaluacyjnym dotyczącym wpływu pomocy przedakcesyjnej w nowych krajach członkowskich. Ostatnio prowadzone badania dotyczyły prac nad mechanizmami realizacji polityki regionalnej w krajach członkowskich Unii Europejskiej, zwłaszcza w Wielkiej Brytanii. Jest autorem kilkudziesięciu prac poświęconych problematyce polityki regionalnej, rozszerzeniu Unii Europejskiej.

Grzegorz Gorzelak – profesor nauk ekonomicznych, specjalizujący się w problematyce rozwoju regionalnego i lokalnego. Kieruje Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG) na Uniwersytecie Warszawskim, jest także profesorem Akademii Leona Koźmińskiego. Twórca i redaktor naczelny kwartalnika „Studia Regionalne i Lokalne" oraz przewodniczący zarządu stowarzyszenia Regional Studies Association – Sekcja Polska. Jest stypendystą Fulbrighta, wykładał też na wielu uniwersytetach zagranicznych, m.in. jako *visiting professor* na University of Strathclyde. Na jego dorobek składa się autorstwo i redakcja ponad 40 książek oraz ponad 200 artykułów i rozdziałów w książkach, w znacznej części opublikowanych za granicą. Współpracował z agendami rządowymi i samorządami lokalnymi i regionalnymi w Polsce i na Ukrainie oraz w wieloma instytucjami międzynarodowymi (takimi jak Bank Światowy, Komisja Europejska, OECD, Open Society Institute).

Jarosław Górniak – dr hab., prof. Uniwersytetu Jagiellońskiego, kierownik Zakładu Socjologii Gospodarki i Edukacji Instytutu Socjologii UJ, prodziekan Wydziału Filozoficznego UJ, prof. w Wyższej Szkole Europejskiej im. Ks. 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. Współpracuje, jako konsultant, także z Departamentem Programowania i Ewaluacji

PARP. Konsultant w zakresie programowania i ewaluacji, a także w zakresie badań rynku i systemów analiz w zarządzaniu. Od ponad 10 lat konsultant SPSS Polska w zakresie metodologii badań i analizy danych. Kierował i uczestniczył w wielu projektach badawczo-analitycznych i ewaluacyjnych, krajowych i międzynarodowych, m.in. jako członek sześcioosobowego zespołu ekspertów wykonywał ewaluację międzynarodowego programu UNDP Poverty Strategies Initiative, obejmującego ponad sto państw.

Agnieszka Haber – absolwentka Wydziału Socjologii oraz Wydziału Stosowanych Nauk Społecznych Uniwersytetu Warszawskiego. Od roku 2004 kieruje Sekcją Ewaluacji w Polskiej Agencji Rozwoju Przedsiębiorczości. Jest odpowiedzialna za realizację ewaluacji programów przedakcesyjnych oraz strukturalnych w obszarach rozwoju sektora MSP, Rozwój Zasobów Ludzkich oraz projektów infrastrukturalnych wdrażanych przez Agencję. Problematyką badań społecznych zajmuje się od ponad 9 lat, zaś obszarem ewaluacji i zarządzania interwencjami publicznymi od 6 lat. Od roku 2005 jest organizatorem międzynarodowej cyklicznej Konferencji Ewaluacyjnej odbywającej się pod auspicjami PARP i MRR. Od roku 2005 jest uczestnikiem grup ds. ewaluacji kolejnych programów: EFS, PO IG, PO KL, PO RPW. Od 4 lat prowadzi wykłady i warsztaty z ewaluacji programów o charakterze społeczno-gospodarczym. Jest autorką wielu publikacji i opracowań z dziedziny ewaluacji programów sektora publicznego.

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. Od 2006 roku prowadzi zajęcia z szeroko rozumianego obszaru badań społecznych (metodologii badań, analizy danych, ewaluacji oraz wykorzystania specjalistycznych programów w procesie badawczym). Brała udział w wielu projektach badawczych, m.in. w pilotażowej na skalę kraju ewaluacji śródokresowej Działania 1.2 i 1.3 Sektorowego Programu Operacyjnego Rozwoju Zasobów Ludzkich 2004–2006 oraz w projekcie *Małopolski system analiz i programowania polityk rynku pracy*, poświęconym budowie systemu wskaźników wykluczenia społecznego i rynku pracy dla Małopolski, który realizowany był w ramach Zintegrowanego Programu Rozwoju Regionalnego. Jest współautorką publikacji wieńczących oba wspominanie projekty. Ponadto, w latach 2005–2007, uczestniczyła w trzech edycjach międzynarodowego kursu: *Social Exclusion in Europe* (w trzeciej edycji jako Socrates-Erasmus *exchange lecturer*). Współpracuje z Małopolską Szkołą Administracji Publicznej, Miejskim Ośrodkiem Pomocy Społecznej w Krakowie (MOPS) oraz Wojewódzkim Urzędem Pracy.

Paweł Kościelecki – absolwent Wydziału Historii Uniwersytetu Warszawskiego, doktorat uzyskał na Uniwersytecie Kardynała Stefana Wyszyńskiego. Ekspert w zakresie sektora kultury; specjalizuje się w dziedzinie zarządzania i ewaluacji programami pomocowymi. W latach 2002–2004 współpracował przy przygotowywaniu ZPORR, SPO-R, programów Inicjatywy INTERREG III A, a także programów Mechanizmu Finansowego EOG i Norweskiego Instrumentu Finansowego. Ekspert kluczowy przy ewaluacji *ex ante* Programu Operacyjnego Infrastruktura i Środowisko 2007–2013. W latach 2004–2007 pracownik Wspólnego Sekretariatu Technicznego Programu Sąsiedztwa Polska – Białoruś – Ukraina INTERREG IIIA/Tacis CBC, odpowiedzialny za programowanie i monitoring. Od 2004 roku prowadzi badania ewaluacyjne na poziomie projektów i programów, finansowanych ze środków UE. Właściciel firmy Grupa Trojden.

Elżbieta Kozłowska – ukończyła Międzywydziałowe Studia Ochrony Środowiska (MSOŚ) na Uniwersytecie Warszawskim (mgr ochrony środowiska) oraz Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG) także na Uniwersytecie Warszawskim (mgr gospodarki przestrzennej). Uczestniczyła w międzynarodowym programie EASY-ECO (*Evaluation of Sustainability: European Conferences and Training Courses*), w ramach którego ukończyła intensywny kurs ewaluacji trwałego rozwoju. Ma kilkuletnie doświadczenie w zakresie ewaluacji projektów i programów współfinansowanych ze środków UE ze specjalizacją w ewaluacji sektora środowisko oraz ocenach oddziaływania na środowisko przedsięwzięć inwestycyjnych.

Magdalena Kuśmierz – socjolog i ewaluator, absolwentka Wydziału Socjologii UW, oraz studiów podyplomowych *Ewaluacja programów finansowanych ze środków EU* w Akademii Leona Koźmińskiego; doktorantka Szkoły Nauk Społecznych przy IFiS PAN. Od ponad 12 lat zajmuje się badaniami społecznymi i marketingowymi, współpracując z firmami badawczymi w Polsce oraz z jednostkami naukowymi i badawczymi w kraju i za granicą (Instytut Stosowanych Nauk Społecznych UW, ALK w Warszawie, Amsterdam School for Social Science Research, Rutgers University). Od kilku lat realizuje ewaluacje projektów i programów finansowanych ze środków UE w obszarze: rozwoju zasobów ludzkich, zarządzania interwencjami publicznymi, analizy instytucjonalnej. Prowadzi zajęcia dydaktyczne i warsztatowe z procedur badań ewaluacyjnych, m.in. w ALK w Warszawie.

Sławomir Mandes – adiunkt w Instytucie Socjologii Uniwersytetu Warszawskiego. Od ponad 10 lat związany z badaniami ewaluacyjnymi. Zaczynał pracę jeszcze w okresie działania w Polsce Amerykańskiej Agencji do Spraw Pomocy (USAID), przy programach której zdobywał

pierwsze doświadczenie w zakresie ewaluacji. Specjalizuje się w ewaluacji programów rozwoju zasobów ludzkich. Uczestniczył w ewaluacji kilku ogólnopolskich programów finansowanych w ramach środków pomocowych UE. Prowadzi badania zarówno jakościowe, jak i ilościowe. Pracę ewaluatora uzupełnia badaniami naukowymi nad problemami przedsiębiorczości, kultury ekonomicznej oraz wykluczenia społecznego. Jest autorem wielu publikacji naukowych i ekspertyz z dziedziny polityki społecznej i procesów społecznych.

Michal Miedziński – konsultant w brukselskim biurze Technopolis Group. Ma wieloletnie doświadczenie w przeprowadzaniu analiz polityki innowacyjnej oraz polityki rozwoju regionalnego. Zrealizował wiele studiów i ewaluacji dla Komisji Europejskiej (głównie Dyrekcji Generalnej ds. Polityki Regionalnej oraz Dyrekcji Generalnej ds. Przedsiębiorstw i Polityki Przemysłowej), Parlamentu Europejskiego oraz Komitetu Regionów. Ma też bogate doświadczenie w pracy z władzami regionalnymi w różnych państwach europejskich (np. Włochy, Francja, Wielka Brytania). Pracował również jako ekspert dla europejskich sieci regionalnych skupiających sie na tematyce innowacji (IRE, ERIK). Jest absolwentem gospodarki przestrzennej (Centrum Europejskich Studiów Regionalnych i Lokalnych – EUROREG na Uniwersytecie Warszawskim) oraz studiów Europejskiej Administracji Publicznej (European Institute of Public Administration – EIPA i Uniwersytet w Maastricht, Holandia). Obecnie pracuje nad doktoratem na Uniwersytecie w Manchesterze (Manchester Institute of Innovation Research, znanym jako PREST, oraz School of Environment and Development – SED).

Wojciech Pander – absolwent studiów regionalnych (Centrum Europejskich Studiów Regionalnych i Lokalnych EUROREG na Uniwersytecie Warszawskim) oraz ekonomicznych (Wydział Ekonomii, Uniwersytet Opolski). Ukończył z wyróżnieniem studia podyplomowe z zakresu ewaluacji w Akademii Leona Koźmińskiego w Warszawie (Ewaluacja programów finansowanych ze środków Unii Europejskiej). Ma 4-letnie doświadczenie w zakresie prowadzenia badań ewaluacyjnych, związanych tematycznie z rozwojem regionalnym, innowacyjnością i rozwojem społeczno-gospodarczym. Wiceprezes Zarządu CASE – Doradcy Sp. z o.o.

Krzysztof Piech – pracownik Szkoły Głównej Handlowej w Warszawie (Katedra Polityki Gospodarczej), członek Centre for the Study of Economic and Social Change in Europe (w University College London – trzeciej, co do wielkości, uczelni w Wielkiej Brytanii), prezes Instytutu Wiedzy i Innowacji. Autor, współautor lub redaktor ok. 20 książek, w tym 5 w języku angielskim (wydanych m.in. w Londynie i Nowym Jorku). Wykładał w Londynie, Sztokholmie, Wilnie. Prowadzi prace eksperckie dla Ministerstwa Rozwoju Regionalnego, Ministerstwa Nauki i Szkolnictwa Wyższego oraz Polskiej Agencji Rozwoju Przedsiębiorczości w zakresie

m.in. oddziaływania funduszy unijnych na polską gospodarkę, społeczeństwa informacyjnego, wskaźników innowacyjności, oceny projektów innowacyjnych.

Adam Płoszaj – asystent naukowy w Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG) na Uniwersytecie Warszawskim. Jest magistrem gospodarki przestrzennej. Naukowo zajmuje się regionalnymi systemami i strategiami innowacji, sieciami organizacji, społeczeństwem informacyjnym oraz rozwojem trwałym w ujęciu regionalnym. Brał udział w licznych projektach badawczych, także międzynarodowych (np. SRDTOOLS, ESPON 1.2.3), oraz badaniach ewaluacyjnych (np.: *Polskie regionalne strategie innowacji. Ocena iwnioski dla dalszych działań*, *Evaluation of Directive 85/337/EEC – Environmental Impact Assessment – and Associated Amendments*, *Ocena udziału polskich partnerów w projektach realizowanych w ramach programu INTERREG IIIB CADSES*).

Jacek Pokorski – absolwent Instytutu Politologii Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie. Od blisko pięciu lat związany z Sekcją Ewaluacji Polskiej Agencji Rozwoju Przedsiębiorczości. Specjalizuje się w ewaluacji programów wsparcia konkurencyjności i innowacyjności gospodarki, skierowanych przede wszystkim do sektora MŚP oraz otoczenia biznesu. Kierownik i współautor wielu projektów badawczych. Doświadczenie w ewaluacji programów współfinansowanych z funduszy przedakcesyjnych i strukturalnych Unii Europejskiej (przede wszystkim Phare ESC 2000–2003 oraz program Wzrost Konkurencyjności Przedsiębiorstw, lata 2004–2006). Członek Grupy Sterującej Ewaluacją Programu Operacyjnego Innowacyjna Gospodarka oraz Grupy roboczej ds. ewaluacji przy Komitecie Monitorującym PO IG. Obecnie zaangażowany w przygotowywanie kompleksowego systemu pomiaru efektów 13 działań PO IG wdrażanych przez PARP. Współautor i współredaktor publikacji poświęconych ewaluacji oraz efektom projektów współfinansowanych ze środków UE. Od kilku lat prowadzi szkolenia z ewaluacji dla przedstawicieli instytucji publicznych oraz organizuje konferencje poświęcone tej tematyce.

Agnieszka Sternik – pracownik jednostki ewaluacyjnej Dyrekcji Generalnej ds. Zatrudnienia, Polityki Społecznej i Równych Szans w Komisji Europejskiej. Od roku 2004 zajmuje się ewaluacją programów europejskich, przede wszystkim finansowanych z 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 roku 2004, jako pracownik Instytucji Zarządzającej ESF w Polsce, stworzyła

jednostkę ds. ewaluacji EFS oraz, współpracując z Krajową Jednostką Oceny, wspierała budowę potencjału ewaluacyjnego w Polsce. Od roku 2006 pracuje w Komisji Europejskiej, w Dyrekcji Generalnej ds. Zatrudnienia, Polityki Społecznej i Równych Szans dla Wszystkich. Przeprowadza ewaluacje inicjatyw Komisji Europejskiej w zakresie polityki zatrudnienia i integracji społecznej, a także współpracuje z krajami UE przy ewaluacji EFS. W ciągu ostatnich 3 lat prowadziła wiele szkoleń z ewaluacji polityk publicznych, głównie w nowych państwach członkowskich.

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

Łukasz Widła – ma ponad 9-letnie doświadczenie w realizacji badań ilościowych oraz analiz statystycznych. Od 8 lat pracuje w renomowanych firmach badawczych, takich jak SMG/KRC czy Ipsos. Początkowo jako analityk, potem jako analityk ds. zaawansowanych analiz statystycznych, obecnie zajmuje stanowisko Project Managera w Ipsos Polska. Oprócz pracy badawczej prowadzi również kursy z analiz statystycznych. Ponadto uczestniczył w realizacji projektów badawczych i doradczych na zlecenie krajowych i międzynarodowych instytucji publicznych, takich jak: Polska Agencja Rozwoju Przedsiębiorczości, m.st. Warszawa oraz Biuro Wysokiego Komisarza Narodów Zjednoczonych do Spraw Uchodźców (UNHCR).

Dominika Wojtowicz – adiunkt w Katedrze Europeistyki Akademii Leona Koźmińskiego. Jest specjalistą i praktykiem w zakresie projektów badawczych i ewaluacyjnych realizowanych w kontekście programów pomocowych UE. Absolwentka studiów zarządzania (ALK) oraz Wydziału Stosunków Międzynarodowych specjalność Europeistyka Uniwersytetu Warszawskiego (tytuł licencjata). W trakcie studiów odbyła stypendia naukowe na Wydziale Nauk Politycznych Università di Siena (Włochy) oraz na Wydziale Zarządzania Universidad Politecnica de Cartagena (Hiszpania). Ma także szerokie doświadczenie praktyczne – brała udział w wielu

projektach badawczych, w tym: ewaluacji bieżącej ZPORR w woj. mazowieckim, projekcie badawczym PARITEA koordynowanym przez European Movement International (Bruksela), badaniach dyskryminacji osób starszych na zlecenie Akademii Rozwoju Filantropii. Ponadto brała udział w konsultacjach społecznych dla planowanego Programu Operacyjnego Wykształcenie i Kompetencje w ramach NPR na lata 2007–2013. Była koonsultantką i osobą odpowiedzialną za opracowanie dokumantacji wnioskowej (w tym studium wykonalności) projektu dofinansowanego w ramach MF EOG dla Biura Stołecznego Konserwatora.

Michał Wolański – doktorant w Katedrze Transportu Szkoły Głównej Handlowej, absolwent SGH, studiował również na Fachhochschule Gelsenkirchen (Niemcy) i Wydziale Ekonomicznym Uniwersytetu Gdańskiego. Specjalizuje się w problemach transportu pasażerskiego, przede wszystkim miejskiego i kolejowego, mając w tych dziedzinach szerokie doświadczenie praktyczne. Współautor studiów wykonalności oraz projektów ewaluacyjnych, a także autor innych opracowań z zakresu nowoczesnej organizacji i finansowania transportu. Członek redakcji „Biuletynu Komunikacji Miejskiej" oraz stały współpracownik „Transportu Miejskiego i Regionalnego". W zakresie ewaluacji – ekspert w firmie EGO S.C.

Piotr Żuber – tytuł doktora ekonomii uzyskał w Szkole Głównej Handlowej w Warszawie. Absolwent Wydziału Geografii i Studiów Regionalnych na Uniwersytecie Warszawskim. Był współkierownikiem twinningów francuskiego i niemiecko-francuskiego oraz brytyjskiego, a obecnie jest współkierownikiem twinningu rumuńskiego. Pracuje na stanowisku dyrektora Departamentu Koordynacji Polityki Strukturalnej w Ministerstwie Rozwoju Regionalnego. Jest odpowiedzialny za Instytucję Strategiczną NSRO oraz za Krajową Jednostkę Oceny. Uczestniczy w licznych pracach dotyczących planowania strategicznego w zakresie polityki regionalnej i tworzenia dokumentów programowych. Autor wielu publikacji dotyczących polityki regionalnej w Polsce, MŚP w polityce regionalnej, polityki spójności.

Wydawnictwa Akademickie i Profesjonalne Spółka z o.o.
Wydanie pierwsze
Arkuszy drukarskich: 34
Skład i łamanie: Wydawnictwo KeyText Sp. z o.o., Warszawa
Druk i oprawa: Fabryka Druku Sp. z o.o., Warszawa