

Karol Olejniczak

TEORIA I PRAKTYKA EWALUACJI W ŚWIETLE DOŚWIADCZEŃ EUROPEJSKICH

Celem artykułu jest przedstawienie ram teoretycznych (definicja, funkcje, typologie) i praktyki ewaluacji w Unii Europejskiej. Analiza praktyki europejskiej przebiega na trzech poziomach: ewaluacji wykorzystywanej przez administracje krajowe do krajowych programów, ewaluacji na poziomie instytucji europejskich (głównie Komisji Europejskiej) i wreszcie ewaluacji programów europejskich na poziomie krajowym. Akcent został postawiony na politykę regionalną Unii Europejskiej. Autor konkluduje, że mimo swoistego boomu ewaluacyjnego, który nastąpił w UE głównie za sprawą wymogów polityki regionalnej, rzeczywisty wpływ ewaluacji na zarządzanie jest wciąż niedostatecznie oszacowany. Istnieje znacząca luka między praktycznym wykorzystaniem ewaluacji (liczbą i skalą prowadzonych ewaluacji) a europejską teorią oraz badaniami empirycznymi, które określałyby efektywność ewaluacji i jej wpływ na zarządzanie rozbudowanymi programami Unii Europejskiej.

1. Wstęp

Termin „ewaluacja” pojawia się coraz powszechniej w praktyce polityki regionalnej, a szerzej zarządzania w sektorze publicznym. Stanowi obecnie jeden z wymogów zarządzania funduszami strukturalnymi, jest też promowany jako dobra praktyka, wręcz standard zarządzania programami publicznymi. Rosnąca lawinowo liczba i skala badań ewaluacyjnych zachęca do przyjrzenia się przyczynom i formom tego procesu. Niniejsza analiza została ujęta w dwóch płaszczyznach. W pierwszej przedstawiono ramy teoretyczne – definicję ewaluacji, jej funkcje w sektorze publicznym oraz teorie związane z użyciem ewaluacji. Prowadzą one do części drugiej prezentującej praktykę ewaluacyjną w Unii Europejskiej opisywaną na trzech poziomach: ewaluacji wykorzystywanej przez administracje krajowe do krajowych programów, ewaluacji w Komisji Europejskiej i wreszcie ewaluacji programów europejskich na poziomie krajowym. Na ten właśnie ostatni element – praktykę ewaluacji i jej perspektywy w ramach funduszy strukturalnych – położono szczególny akcent.

2. Teoria ewaluacji

2.1. Definicja ewaluacji

Semantycznie ewaluacja oznacza „oszacowanie wartości” (Worthen 1990; *Oxford English Dictionary* 2000). W takim przypadku termin „ewaluacja” zasadniczo odpowiadałby polskiemu słowu „ocena”. Jednak definicje ewaluacji programów pochodzące z literatury tematu, a także praktyka, podkreślają jej odmienność od oceny¹. Najszerzej ujmując, ewaluacja to użycie metod badań społeczno-ekonomicznych do systematycznego zbierania danych, analizy, interpretacji, oceny i informowania o skuteczności programów publicznych (Rossi et al. 1999; Mathison 2004, s. 139).

Trzy cechy odróżniają ewaluację od innych badań i metod dyscyplinowania sektora publicznego takich jak na przykład kontrola, audyt, analiza. Po pierwsze, ewaluacja łączy w sobie element empiryczny (bada program i zjawiska pojawiające się jako konsekwencje programów) z charakterem normatywnym (ocenia strukturę działań, ich wewnętrzną i zewnętrzną logikę, a przede wszystkim ich wartość) (Mathison 2004, s. 139). Po drugie, rdzeniem ewaluacji jest jej utylitaryzm – ma wspomagać konkretnych odbiorców, w konkretnej sytuacji (Stufflebeam 2001), a jej celem jest wywołanie pozytywnej zmiany. Wreszcie po trzecie, co jest coraz częściej podkreślane, „ewaluacja” jest procesem uczenia się, w którym kluczowe są interakcje i negocjacje (Greene 2003, s. 592). W dokumentach Unii Europejskiej ewaluacja jest definiowana nieco bardziej technicznie – jako ocena interwencji według jej rezultatów, oddziaływania oraz potrzeb, które ma (miała) spełnić (KE 2000a; KE 2004a).

2.2. Rola ewaluacji w sektorze publicznym

Najszerzej określaną, podstawową rolą ewaluacji jest podnoszenie jakości i skuteczności programów publicznych (WE 1999, art. 40; KE 2003; Kierzkowski 2002). W sektorze prywatnym skuteczność działań jest w prosty sposób weryfikowana przez mechanizm rynkowy. Działania sektora publicznego, choć często prowadzone w warunkach rynkowych, nie mogą być oceniane wyłącznie

¹ W polskich dokumentach na określenie ewaluacji stosowany jest najczęściej, choć nie zawsze konsekwentnie, termin „ocena”. Moim zdaniem użycie angielskojęzycznego terminu jest bardziej wskazane z kilku powodów. „Ewaluacja” jest procesem, ocena, czy raczej osąd, jest zaś tylko jednym z elementów tego procesu. Według niektórych podejść osąd nie zawsze jest kluczowy w procesie ewaluacji, a może być nawet pominięty (akcent jest wtedy postawiony na interakcje i animacje negocjacji pomiędzy uczestnikami programu). Użycie terminu „ocena” po pierwsze koncentruje uwagę wyłącznie na jednym elemencie i przypisuje mu cechy całego procesu. Po drugie, wyłącza cały szereg alternatywnych podejść. Wreszcie po trzecie „ewaluacja” to także nie tylko pojedyncze działanie, ale dziedzina wiedzy. Jak zauważa Kierzkowski (2002, s. 10), ze względów semantycznych trudno byłoby więc określać terminem „ocena” samodzielną dziedzinę naukową. W związku z powyższymi argumentami w niniejszym artykule stosowany jest termin „ewaluacja”.

z perspektywy osiągniętego zysku (Stewart, Ranson 1994). Kryteria osądu wartości programów publicznych są zróżnicowane i zależne od wielu czynników, w tym dominujących przekonań czy aktualnych potrzeb społeczności. Ewaluacja dostarcza kryteriów, metod i środków do oceny racjonalności działań publicznych, odgrywając w sektorze publicznym rolę analogiczną do mechanizmu rynkowego w sektorze prywatnym (Shadish et al. 1991, s. 19).

Realizację tak szeroko zakreślonej roli ewaluacji zapewniają liczne funkcje (zwykle określane mianem celów ewaluacji). W literaturze tematu przedstawiane są różne kategoryzacje (np. Chelimsky 1997; Bachtler 2000; KE 2003), ale wyróżnić można pięć głównych funkcji ewaluacji:

- poprawianie planowania – gwarancja uzasadnienia dla polityki/programu oraz wydajnej i racjonalnej alokacji zasobów,
- poprawa wdrażania – podnoszenie jakości struktur zarządzania programem oraz samego procesu implementacji działań,
- wzmocnienie odpowiedzialności – informowanie społeczeństwa, w jakim stopniu cele programu zostały osiągnięte i jak efektywnie użyto przypisanych zasobów publicznych,
- poszerzanie wiedzy i wspieranie procesów uczenia się – wyjaśnianie czynników i procesów warunkujących sukces programu lub jego porażkę, weryfikowanie teorii leżących u podstaw danych polityk, promowanie dobrych wzorców,
- wzmacnianie instytucji – podnoszenie umiejętności i zasobów uczestników programu, w tym wzmocnianie powiązań i relacji między nimi.

Do tej listy warto dodać jeszcze funkcję szóstą – minimalizowanie negatywnych efektów ubocznych zarządzania opartego na wynikach (*performance management*). Jak zostanie to szerzej omówione w części poświęconej praktyce europejskiej, ewaluacja jest silnie złączona z nowym podejściem do zarządzania w sektorze publicznym (czerpiącym wzorce z zarządzania biznesowego). Podejście to, mimo iż dynamizuje i dyscyplinuje sektor publiczny, może prowadzić do pewnych zniekształceń. Umiejętnie stosowana ewaluacja może minimalizować ryzyko ich wystąpienia. W tab. 1 przedstawiono kilka dysfunkcji mogących wystąpić przy zarządzaniu opartym na wynikach (Smith 1995; Leeuw 2002) oraz sposób, w jaki mogą być one neutralizowane przez ewaluację.

Oczywiście poszczególne funkcje ewaluacji w praktyce nie są rozdzielne. Pojedyncze badanie ewaluacyjne może, a nawet powinno spełniać kilka funkcji.

Tab. 1. Dysfunkcje zarządzania opartego na wynikach

Nazwa i opis dysfunkcji	Przeciwdziałanie dzięki ewaluacji
Nacisk na mierzenie celów, pomiar sukcesu zamiast na rzeczywiste efekty (ang. <i>measure fixation</i>).	Analiza rzeczywistych zarówno planowanych, jak i ubocznych efektów (pozytywnych i negatywnych).
Zniechęcenie do innowacji poprzez sztywny system wskaźników i pomiarów – wszystko, co nie pasuje do struktury, nie jest brane pod uwagę (ang. <i>ossification</i>).	Szersze spojrzenie, użycie różnych metod. Krytyczna analiza samego systemu wskaźników (szerzej: monitoringu) – ich zasadności, praktycznego zastosowania.
Koncentrowanie uwagi na kwantyfikowalnych efektach ilościowych i ignorowanie trudno mierzalnych, jakościowych efektów programu (ang. <i>tunnel vision</i>).	Badanie efektów jakościowych, miękkich, relacji pomiędzy interesariuszami przy użyciu narzędzi analizy społ.-ekon. (wywiady, ankiety, obserwacja uczestnicząca).
Koncentrowanie się na wąskich celach technicznych lub celach krótkookresowych kosztem celów strategicznych całego programu czy polityki (ang. <i>suboptimization & myopia</i>).	Weryfikowanie, w jakim stopniu projekt (pozytywnie zakończony od strony formalnej) przyczynił się do szerszej strategii, wpłynął na cele polityki.
Błędne interpretowanie danych przede wszystkim kontekstowych i w rezultacie budowanie błędnych założeń programu i ocen (ang. <i>misinterpretation</i>).	Poprawne użycie metod badawczych, spojrzenie z kilku perspektyw, używanie kilku rodzajów narzędzi do oceny jednej kwestii (tzw. triangulacja).
Manipulowanie danymi, tak by pokazać na papierze efekty programu różne od rzeczywistości (ang. <i>misrepresentation</i>).	Weryfikowanie efektów na podstawie różnych źródeł. Utrzymanie dystansu do zlecającego dające niezależność i obiektywizm sądów (ang. <i>arm-length distance</i>) i przestrzeganie standardów zawodowych.

Źródło: opracowanie własne na podstawie: Smith 1995; Leeuw 2002.

2.3. Typy ewaluacji

W literaturze tematu spotyka się szereg typologii ewaluacji (np. KE 2003, s. 22; Hermans et al. 1987; Greene 2003; Owen, Rogers 1999). To zróżnicowanie wynika po pierwsze z odmiennych inspiracji teoretycznych i podstaw metodycznych, po drugie zaś z przyjmowania różnych kryteriów podziału ewaluacji (ze względu na czas wykonania ewaluacji, cel, użyte metody etc.). Upraszczając, możemy jednak wyróżnić cztery podstawowe typologie ewaluacji: ze względu na kryterium przedmiotu ewaluacji, cel ewaluacji, moment podjęcia badania, i wreszcie administracyjną organizację badania. Ich zestawienie prezentuje tab. 2.

Jeśli chodzi o relacje między poszczególnymi typami (które nie są rozłączne), to metaewaluacje prowadzone są głównie dla ewaluacji końcowych, rzadziej ewaluacji średniookresowych. Ewaluacje tematyczne są najczęściej elementem ewaluacji średniookresowej lub jeszcze częściej bieżącej. Zwykle zlecane na zewnątrz są ewaluacje końcowe i średniookresowe. Ewaluacje wstępne są zorientowane na decyzje i mają charakter formatywny. Ewaluacje średniookresowe i bieżące mają charakter formatywny (mogą również zawierać elementy podsumowujące), powinny też pobudzać uczestnictwo interesariuszy (*stakeholders*) w programie. Ewaluacje końcowe mają charakter podsumowujący.

Tab. 2. Typy ewaluacji

Typ ewaluacji	Charakterystyka
Podział ze względu na przedmiot ewaluacji	
Ewaluacja projektów	Badanie koncentrujące się na analizie tylko jednej, drobnej interwencji. W praktyce UE takie ewaluacje są elementem większej całości (badania programu) i później wykorzystywane są jako studia przypadku lub też wykonywane są dla działań pilotażowych i mają pomoc w decyzji uruchomienia większej interwencji.
Ewaluacja programu	Najczęściej spotykana forma ewaluacji – dotyczy zarówno pojedynczych programów sektora publicznego, jak i całych pakietów interwencji – działań, priorytetów (np. Ramowy Program Wsparcia).
Ewaluacja polityki	Częściej nazywana analizą polityk (<i>policy analysis</i>) – składa się z szeregu badań i analiz (często na przestrzeni wielu lat). W administracjach państw członkowskich często praktykowana, na poziomie UE jest kwestią delikatną politycznie (np. próby analizy wspólnej polityki rolnej), niemniej niektóre działania UE doczekały się szczegółowych, regularnych i krytycznych ewaluacji (np. polityka spójności, pomoc zewnętrzna).
Ewaluacja tematyczna (ang. <i>thematic/horizontal evaluation</i>)	Przekrojowa (tzn. wykonywana dla kilkunastu programów lub projektów) ocena konkretnej kwestii zarządzania lub aspektu interwencji (np. zasady partnerstwa w funduszach strukturalnych czy wsparcia innowacyjności MŚP w ramach kilku celów polityki regionalnej UE).
Metaewaluacja (ang. <i>meta-evaluation</i>)	Analiza porównawcza, podsumowująca obserwacje, wnioski i rekomendacje kilkunastu ewaluacji z wielu analogicznych programów, często także oceniająca zastosowane w poszczególnych badaniach podejścia i techniki (a więc jakość samej ewaluacji).
Podział ze względu na główny cel ewaluacji	
Ewaluacja formatywna (<i>formative evaluation</i>)	Jej celem jest przede wszystkim ocena formy i procesu 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 programu i podnieść jego ostateczną jakość.
Ewaluacja podsumowująca (<i>summative evaluation</i>)	Ma dokonać cabościowej oceny danej interwencji (najczęściej już po jej zakończeniu), to jest określić efekty (zarówno zakładane, jak i uboczne). Ma charakter opisowy i wyjaśniający.
Podział ze względu na moment prowadzenia ewaluacji	
Ewaluacja wstępna (ang. <i>ex ante evaluation</i>)	Prowadzona przed rozpoczęciem interwencji. Jej celem jest stworzenie podstaw przygotowania i uzasadnienia programu. Ma zweryfikować spójność interwencji (zabozania przyczynowo-skutkowe leżące u podstaw programu), jej przystawalność do zidentyfikowanych problemów (kontekstu), ocenę tego, czy oczekiwane oddziaływanie jest realistyczne, oszacować rezultaty, ustalić kwantyfikowalne cele (co daje podstawy do monitoringu i dalszej ewaluacji).
Ewaluacja śródkresowa/średniookresowa (ang. <i>mid-term evaluation</i>)	Przeprowadzana w połowie trwania danego programu. Kładzie nacisk na ocenę dotychczasowego sposobu realizacji interwencji (głównie proces zarządzania), weryfikację jej kierunku oraz aktualności założeń, zidentyfikowanie pierwszych efektów interwencji oraz zaproponowanie ewentualnych bieżących korekt.

Cd. tab. 2

Ewaluacja ciągła/bieżąca (ang. <i>on-going evaluation</i>)	Wykonywana w trakcie cyklu implementacji programu – może być badaniem punktowym albo składać się z szeregu badań. Ma zapewnić wsparcie dla zarządzania programem. Jej celem jest dogłębna analiza pytań, kwestii i problemów, które pojawiają się w trakcie realizacji programu, a także zaproponowanie konkretnych rozwiązań (może koncentrować się na jednym, wybranych aspekcie). Ma też wzmacniać relacje między operatorami programu, sprzyjając ścisłej współpracy i akumulacji doświadczeń.
Ewaluacja <i>ex post</i> /końcowa /pełna (ang. <i>ex post evaluation</i>)	Jest wykonywana 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 zidentyfikowaniu ewentualnych efektów ubocznych (tak pozytywnych, jak i negatywnych). Tym samym ożywia dyskusję na temat racjonalności przyszytych analogicznych programów.
Podział ze względu na organizację badania	
Ewaluacja wewnętrzna	Prowadzona przez osoby z danej instytucji odpowiedzialnej za wdrażanie analizowanej interwencji (najczęściej przez odpowiedni departament lub komórkę ewaluacyjną funkcjonującą w ministerstwie kierującym danym programem).
Ewaluacja zewnętrzna	Zlecana poza instytucję prowadzącą projekt czy program – wykonywana przez zespoły niezależnych ekspertów wyłonione w ramach postępowania przetargowego (najczęściej taką ewaluację wykonują instytuty naukowe lub firmy konsultingowe).

Źródło: opracowanie własne na podstawie: KE 2003; Patton 1990; Hermans et al. 1987.

2.4. Badania faktycznej roli ewaluacji

Kwestie badania rzeczywistej roli, jaką odgrywa ewaluacja w sektorze publicznym (inaczej ujmując – wypełnienia funkcji w praktyce), są w literaturze tematu określane jako „wykorzystanie ewaluacji” (*evaluation utilization*). Koncepcje te czerpią głównie z podejść do procesów uczenia się organizacji (Scriven 1991, s. 369).

Wyróżniane są dwa podstawowe rodzaje wykorzystania ewaluacji (Alkin za: Mathison 2004, s. 436). Pierwsze to użycie instrumentalne (ang. *instrumental use*) – czyli proste do udokumentowania przełożenie wniosków z ewaluacji na działania odbiorców ewaluacji (np. wpływ rekomendacji z ewaluacji na podjęcie decyzji co do zmiany kierunku programu). Drugie to użycie koncepcyjne (ang. *conceptual use*) – czyli wpływ ewaluacji na zmianę sposobu myślenia odbiorców ewaluacji o istocie realizowanych programów czy procesów społecznych. Dodatkową osią analizy wykorzystania ewaluacji jest perspektywa czasowa. Efekty ewaluacji – zarówno instrumentalne, jak i koncepcyjne – mogą nastąpić bezpośrednio po zakończeniu ewaluacji i wpłynąć na badany program lub też pojawić się z czasem – oddziałując na kolejne edycje danego programu lub przyszłe inicjatywy o analogicznym charakterze.

Wskazuje się wiele czynników wpływających na efekty ewaluacji (Rossi et al. 1999, s. 434; Patton 1990). Można je pogrupować w trzech kategoriach. Do pierwszej należą czynniki kontekstowe programu. Za takie uznaje się w przypadku programów europejskich zróżnicowane kultury ewaluacji w administracjach poszczególnych krajów członkowskich UE (Bachtler 2001). Do grupy drugiej należą różnego rodzaju warunki wstępne badania: przedmiot badania², moment, w którym podjęte jest badanie, oczekiwania zlecających, które przybierają postać pytań ewaluacyjnych i koncentrują uwagę ewaluacji na wybranych funkcjach³. Ostatnią grupę czynników wpływających na efekty ewaluacji stanowi samo badanie ewaluacyjne. Chodzi po pierwsze o wybrany model, podejście do ewaluacji i używaną metodologię (wybór bądź też wyważenie proporcji pomiędzy metodami ilościowymi a jakościowymi). Po drugie wpływ ma sam proces badania – relacje między ewaluatorami a zlecającymi badanie oraz szerszą grupą użytkowników programu (Patton 1990; KE 2004b)⁴. Po trzecie, ważny jest produkt końcowy – raport ewaluacyjny – poprawność metodologiczna i wiarygodność wniosków, jakość rekomendacji, przejrzystość i komunikatywność raportu (Morris et al. 1987; Rossi

² Chodzi o charakter programu, głównie jego skalę (program lokalny vs. program o zasięgu krajowym czy ogólnoeuropejskim) i etap jego rozwoju (działanie pilotażowe – innowacyjne vs. ugruntowana interwencja).

³ W przypadku polityki regionalnej UE Komisja sugeruje pewien zakres pytań poprzez tzw. wymagane kryteria ewaluacji i przypisanie ich do momentów prowadzenia ewaluacji – szerzej o tym w dalszej części artykułu.

⁴ Na przykład metody partycypatywne (wywiady, panele dyskusyjne) będą warunkować funkcje informacyjne i wzmacniania odpowiedzialności, a także wzmacnianie instytucji i poszerzanie wiedzy.

et al. 1999). Tutaj uwzględniana jest również strategia użycia ewaluacji (*evaluation dissemination*) – przygotowana wspólnie przez ewaluatorów i zlecających badanie (KE 1999a, t. 1).

Kończąc kwestie teoretyczne, warto zaznaczyć, że zarówno teoria ewaluacji, jak i badania jej efektywności są rozwijane głównie w USA. Na przykład na rynku USA funkcjonuje kilkanaście naukowych periodyków specjalizujących się w tej tematyce, w Europie znaczący jest tylko jeden – kwartalnik *Evaluation*.

3. Praktyka ewaluacji w Europie

Praktykę ewaluacji na obszarze obecnej Unii Europejskiej można analizować na trzech płaszczyznach. Pierwszą są administracje krajowe i działania ewaluacyjne w odniesieniu do krajowych programów sektora publicznego. Płaszczyzną drugą jest poziom instytucji Unii Europejskiej – przede wszystkim użycie ewaluacji w samej Komisji Europejskiej. Wreszcie płaszczyznę trzecią – łączącą dwie poprzednie – stanowi praktyka ewaluacji w odniesieniu do programów unijnych realizowanych w administracjach państw członkowskich UE i razem z nimi – i tu chodzi przede wszystkim o fundusze strukturalne. Ten trójpodział został zastosowany w dalszej części artykułu.

3.1. Ewaluacja w tradycji administracji krajowych UE

Ewaluacja jest integralnym elementem dużo szerszych działań na rzecz podnoszenia efektywności zarządzania w sektorze publicznym. Momentem uznawanym za początek praktyki ewaluacji programów publicznych była połowa lat 60. Wtedy to amerykańska administracja prezydencka zaczęła realizować plan *Great Society* składający się z szeregu programów sektora publicznego. Ponieważ nastąpił znaczący wzrost wydatków administracji federalnej (szacuje się, że o ponad 200%), zaczęto poszukiwać metod racjonalizowania zarządzania w sektorze publicznym. Serią reform wprowadzono system zarządzania PPBS – *Planning, Programming, Budgeting System* (Planowanie, Programowanie, System Budżetowy) oraz MBO – *Management by Objectives* (Zarządzanie przez Cele). Opierały się one na trzech elementach (Heinrich 2003):

- strukturalizowaniu działań oraz funkcji instytucji publicznych w formie mierzalnych celów,
- wieloletnich planach budżetowych (konstruowanych na zasadzie zadaniowej),
- obiektywnej analizie wdrażanych programów opartej na badaniach i ewaluacji.

Ze Stanów Zjednoczonych PPBS (a wraz z nim ewaluację) szybko, wręcz z entuzjazmem importowały administracje krajów Europy Zachodniej (Wielka Brytania, Francji, Holandia, Niemcy) (Toulemonde 2000).

Od tego czasu ewaluacja, mimo iż zmieniająca formę i metody, była elementem kolejnych reform zarządzania w sektorze publicznym – począwszy od opisanego już PPBS i MBO oraz ich udoskonaleń, poprzez reformy Nowego Menedżeryzmu (ang. *New Managerialism*) w latach 80. (wymuszone kryzysami naftowymi), aż po dominujący obecnie nurt Nowego Zarządzania Sektorem Publicznym (ang. *New Public Management* – NPM – Heinrich 2003; Shafritz et al. 2003)⁵.

Analiza administracji państw UE pod względem podejścia do ewaluacji pozwala podzielić je na trzy grupy. Do pierwszej grupy należą państwa, które wprowadzały ewaluację systemowo, bazując na swojej kulturze administracyjnej (Bachtler et al. 2000, Toulemonde 2000). W tych krajach ewaluacja od początku była traktowana jako integralne i praktyczne narzędzie codziennej pracy sektora publicznego. Klasycznym przykładem jest Wielka Brytania; do grupy tej zaliczają się też Holandia, Niemcy, Francja, Szwecja, w mniejszym stopniu Finlandia. Część z tych państw od początku traktowała ewaluację jako unikatowe narzędzie; inne, takie jak na przykład Szwecja, mając silną tradycję audytu, traktowały ewaluację jako uzupełniający, bardziej miękki aspekt kontroli.

Drugą grupę (bardzo zróżnicowaną wewnątrz) stanowią natomiast kraje, w których wiedza i praktyka ewaluacji miały początkowo słabe bądź też żadne poparcie w kulturze administracyjnej, a były zaszczipiane z zewnątrz i rozwijane często oddolnie (ze szczebla lokalnego i regionalnego ku centralnemu). W przypadku tych krajów wymóg ewaluacji przyszedł wraz z realizowanymi tam programami Banku Światowego, a później funduszy strukturalnych. Ewaluacja była traktowana początkowo jako kolejna biurokratyczna powinność wobec zewnętrznych darczyńców. Jej użyteczność dla krajowych działań sektor publiczny odkrywał dopiero z czasem i stopniowo internalizował. Tempo odkrywania i adaptowania ewaluacji było różne. Administracje Irlandii, Belgii oraz Włoch szybko ją adaptowały i zaczęły skutecznie stosować do działań krajowych (przykład Irlandii jest tu wręcz wzorcem europejskim). W Hiszpanii, Portugalii, a przede wszystkim Grecji, ewaluacja, mimo iż wykonywana (acz słabo i pobieżnie), pozostawała papierową rzeczywistością tworzoną na potrzeby Brukseli, a ignorowaną przez urzędników krajowych (Bachtler 2000; Toulemonde 2000).

Ostatnią grupę stanowią nowe kraje członkowskie. Ich administracja dopiero od początku lat 90. przedstawiała się na nowoczesny sposób zarządzania (wieloletnie planowanie, budżety zadaniowe, strategie), i to głównie za sprawą pieniędzy pomocowych (unijnych, Banku Światowego, USAid). Wraz z programami pomocowymi szły wymagania procedur i specyficznego sposobu zarządzania. Wśród tych wymagań była również ewaluacja. Początkowo

⁵ Po raz pierwszy termin NPM został użyty przez Hood (1991). Obecnie funkcjonuje co najmniej w trzech kontekstach, co może prowadzić do pewnego zamieszania. Po pierwsze, oznacza doktrynę administracji publicznej, po drugie, cały zestaw technik zarządzania w sektorze publicznym, po trzecie wreszcie służy określaniu międzynarodowej tendencji, ruchu reformatorskiego w administracjach krajowych (głównie z kręgu kultury anglosaskiej).

stosowana tylko dla pieniędzy pomocowych, wykonywana głównie przez firmy zagraniczne i tylko na potrzeby darczyńców (przyjmująca ich perspektywę), stopniowo i selektywnie jest podpatrywana i przejmowana przez krajowe instytucje do krajowych programów. Analizując rozwój ewaluacji w nowych krajach członkowskich, należy jednak podkreślić dwie kwestie. Po pierwsze, proces internalizacji był dotychczas bardzo powolny (bo długotrwały jest szerszy proces – zmiany sposobu myślenia w sektorze publicznym), po drugie, przyspiesza on w ostatnich dwóch latach – czego głównym czynnikiem są fundusze strukturalne⁶.

Podsumowując przyczyny upowszechniania się ewaluacji w Europie, można wymienić trzy czynniki. Pierwszym i podstawowym jest oczywiście konieczność racjonalizowania (a w praktyce najczęściej ograniczania) wydatków publicznych. Drugim czynnikiem jest wytworzenie się pewnych ogólnostanowionych wzorców zarządzania w sektorze publicznym krajów rozwiniętych – silnie promowanych przez OECD i Bank Światowy – ich elementem jest i ewaluacja⁷. Wreszcie trzeci czynnik stanowi standaryzacja narzucana w ramach UE – przez wymagania polityk i programów tej organizacji – przede wszystkim przez wymagania i procedury funduszy strukturalnych.

Obserwacja wszystkich administracji państw UE pozwala stwierdzić, że obecnie w sposób bardzo wyraźny ewaluacja staje się standardem nowoczesnej administracji publicznej. Oczywiście proces adaptacji przebiega odmiennie – bo inne są kultury administracyjne. Wyraźnie jednak widać, że ewaluacja jest internalizowana i rzeczywiście używana w administracjach północy kontynentu. Na południu wciąż jest często papierową rzeczywistością tworzoną na potrzeby funduszy strukturalnych. Nowe kraje członkowskie stoją na rozdrożu – między „północnym” a „południowym” podejściem do ewaluacji.

3.2. Ewaluacja w instytucjach UE

W instytucjach wspólnotowych ewaluacja pojawiła się na początku lat 80. Co ciekawe (aczkolwiek typowe dla procesów reformatorskich w UE), pojawiała się najpierw w politykach o stosunkowo najmniejszych budżetach i małej wadze politycznej. Pierwsze po ewaluację sięgnęły w Komisji Europejskiej Dyrekcje Generalne: DG-Stosunki Zewnętrzne (i późniejsze zintegrowane biuro Euro-

⁶ Szerzej o transferze praktyk ewaluacyjnych z programów unijnych do polityk krajowych na przykładzie polityki regionalnej w Polsce zob. Olejniczak 2002.

⁷ Obecny wzorem jest *New Public Management*. Opiera się on na czterech podstawowych cechach: 1) skoncentrowaniu się na rezultatach (ilościowe i jakościowe podejście) i zarządzaniu przez cele; 2) zorientowaniu działań administracji publicznej na klienta – obywatela; 3) wprowadzeniu rynkowego i *quasi*-biznesowego podejścia do sektora publicznego (zlecanie zadań na zewnątrz poza sektor publiczny – tzw. *contracting out*, wprowadzanie konkurencji do sektora publicznego, przekazanie zarządzania w administracji publicznej wykwalifikowanym menedżerom, zapożyczanie z sektora prywatnego technik zarządzania – np. SWOT, zarządzanie ryzykiem) oraz 4) decentralizacji zadań i funkcji (głównie przekazanie ich wyspecjalizowanym agencjom *quasi*-rządowym).

peAid) i DG-Badania. Szukając narzędzi bardziej efektywnego ukierunkowania prowadzonych przez siebie programów, zaczerpnęły z doświadczeń Banku Światowego i OECD. Symptomatyczne, iż w przypadku polityki rolnej ewaluacja pojawiła się dopiero w latach 90.

Wyróżnić można dwie główne przyczyny wykorzystania ewaluacji do polityk i programów unijnych. Pierwszą był skokowy wzrost wydatków oraz związana z nim konieczność kontroli, bardziej racjonalnego wykorzystania pieniędzy unijnych. Momentem przełomowym była tu reforma budżetowa Delorsa z 1987 r. (tzw. I Pakiet Delorsa), a głównym polem dyskusji – polityka regionalna⁸. Drugą przyczyną była konieczność podniesienia wiarygodności instytucji europejskich w oczach obywateli. Po spektakularnym kryzysie Komisji Santera w końcu lat 90. rozpoczęto reformę Komisji Europejskiej (głównymi jej architektami byli Anglosasi – opierali się więc na anglosaskim podejściu do zarządzania w sektorze publicznym, które włączali w typowo francuską strukturę Komisji). Celem reformy było podniesienie przejrzystości działań instytucji unijnych, zapewnienie ich racjonalności, jakości, sprawności, poszanowania wartości pieniądza europejskiego podatnika. Wymóg ewaluacji, obok wzmocnionych procedur audytu i monitoringu, stał się jednym z kluczowych narzędzi reformy⁹.

Od 2000 r. każde działanie, które ma wpływ na budżet UE, musi być poddane ewaluacji. Wszystkie dyrekcje zostały zobowiązane do działań ewaluacyjnych w swoich dziedzinach¹⁰. Nadzór i koordynację w Komisji Europejskiej przejęła DG-Budżet¹¹. Oprócz Komisji ewaluację wykorzystuje także (głównie jako „amunicję” do debat politycznych) Parlament Europejski. Trzeba również wspomnieć o badaniach prowadzonych przez Trybunał Audytorów. Mimo iż są to audyty, to jednak w przypadku audytów programów europejskich stawiają one również pytania z zakresu ewaluacji (o skuteczność działań, efektywność procedur etc.).

W praktyce instytucji UE (przede wszystkim Komisji Europejskiej) istnieją nieco odmienne podejścia do ewaluacji¹². Wyróżnić można trzy główne nurty.

⁸ Państwa płatnicy netto i Komisja Europejska szukały odpowiednich narzędzi, które zapewniłyby kontrolę zasadności tak znaczącej interwencji publicznej oraz wymusiły jej skuteczność. Sięgnięto więc po ewaluację i wpisano ją jako jeden z wymogów systemu zarządzania funduszami strukturalnymi. Ewaluacja po raz pierwszy pojawiła się w kontekście dużych europejskich pieniędzy.

⁹ Prace nad podnoszeniem jakości zarządzania, przede wszystkim finansowego, w administracji europejskiej rozpoczęto już w połowie lat 90. W 1995 r. Komisja uruchomiła program SEM 2000 *Sound and Efficient Management 2000* (Przejrzysty i skuteczny system zarządzania). SEM stał się jednak priorytetem dopiero po wspomnianym już kryzysie Komisji.

¹⁰ Por. KE 2000b.

¹¹ http://www.europa.eu.int/comm/budget/evaluation/index_en.htm.

¹² Różnice w podejściu dotyczą najczęściej terminów (co po części wynika z tradycji danej dyrekcji, a po części z niechęci do współpracy z innymi dyrekcjami), rozłożenia akcentów (wynikającego z odmiennej tematyki poddawanej badaniu), formy badania (specyfika poszczególnych polityk oczywiście narzuca formę ewaluacji) oraz przewidywanej roli raportów (np. raporty dla EuropeAid mają silne przełożenie praktyczne – program jest kontrolowany, może być wstrzymany, przy funduszach strukturalnych natomiast raporty dają raczej obraz sytuacji i są punktem wyjścia do dyskusji Komisja–administracja krajowa lub do dużych debat politycznych).

Pierwszym są praktyki DG-Badania i Nauka, w której ewaluacja oparta jest (z uwagi na specyfikę dziedziny) na tzw. *peer assessment* (czyli ocenie przez panel naukowców – ekspertów uczestników danego programu). Drugim jest podejście EuropeAid oparte na zarządzaniu cyklem projektu (*Project Cycle Management*). Jest stosowane przy takich programach jak TACIS, MEDA, ECHO, CARDS, a do niedawna również PHARE. Wreszcie trzecia, wiodąca metodologia to MEANS (*Evaluation Methods for Actions of the Structural Nature*), rozwijany od początku lat 90. przez DG-Regio. Jest on wykorzystywany przy funduszach strukturalnych, ale także w działaniach koordynowanych przez DG-Rolnictwo, DG-Zatrudnienie czy DG-Przedsiębiorczość. Oczywiście poszczególne dyrekcje wydają swoje poradniki na temat ewaluacji¹³. Należy zaznaczyć, że metody EuropeAid i DG-Regio są zbieżne, mogą być łączone, a często się przenikają. W ostatnich trzech latach wyraźnie też widać unifikację metodologii, głównie za sprawą działań koordynacyjnych DG-Budżet.

W praktyce programów UE ewaluacja jest badaniem społeczno-ekonomicznym składającym się z pięciu etapów (KE 1999a; KE 2003): strukturalizacji (sprecyzowania pytań badawczych oraz kryteriów, według których program będzie oceniany), obserwacji (zebrania danych wtórnych i pierwotnych), analizy, oceny i rozpowszechniania wyników (wniosków i rekomendacji). Rdzeniem ewaluacji jest oczywiście moment oceny. Ocena jest podejmowana pod kątem tzw. kryteriów ewaluacyjnych. W praktyce UE stosowanych jest najczęściej pięć kryteriów zamienianych następnie na szczegółowe pytania badawcze. Przedstawiono je w tab. 3. W zależności od momentu badania, potrzeb zlecającego oraz środków przeznaczonych na badanie wybiera się kilka kryteriów i pytań¹⁴.

Poszukiwania i odpowiedź na powyższe pytania, a więc wykonanie badania, w praktyce opiera się na użyciu szerokiego wachlarza narzędzi badań społeczno-ekonomicznych. Metody najczęściej stosowane w praktyce UE można podzielić na statyczne – a więc niewymagające „wyjścia z biurka”, oraz na interaktywne – wymagające bezpośredniego kontaktu z przedmiotem badania, aktorami programu. Ponadto różne narzędzia znajdują zastosowanie przy różnych typach ewaluacji, co więcej, ich użyteczność jest różna na poszczególnych etapach badania. Nieco uproszczoną mapę wykorzystania głównych narzędzi badawczych przedstawia tab. 4.

Z uwagi na wprowadzony w 2000 r. wymóg ewaluacji liczba badań prowadzonych lub zleczanych przez Komisję jest znacząca – sięga ponad 1000

¹³ Podręczniki EuropeAid to: KE 2001a; KE 2001b; KE 2004c. Podręczniki DG-Regio to cytowany już MEANS (KE 1999a; KE 2003). Materiał DG-Budżet to KE 2004a oraz tłumaczony ostatnio na język polski podręcznik KE 1997.

¹⁴ Trafność obowiązuje tylko przy ewaluacji wstępnej i średniookresowej – nigdy *ex post*. Użyteczność (będąca lustrzanym odbiciem trafności) i trwałość stosowane są do ewaluacji *ex post*, nigdy *ex ante* (z prostej przyczyny, że nie można odpowiedzieć na te pytania przed rozpoczęciem programu).

Tabela 3. Kryteria ewaluacyjne i pytania badawcze¹⁵

Kryterium ewaluacyjne	Opis	Stawiane pytania
Trafność (<i>relevance</i>)	Bada adekwatność celów i zastosowanych metod programu do problemów i kwestii społeczno-ekonomicznych, które program miał poruszyć na danym terytorium.	Czy cele programu odpowiadają zidentyfikowanemu potrzebom? Czy program prawidłowo identyfikuje główne siły i słabości danego układu? Czy obrana strategia jest spójna? To znaczy, czy strategia jest spójna wewnętrznie (cele-działania), czy strategia jest zbieżna z celami szerszych działań, polityk? Czy logika interwencji jest poprawna (zależności przyczynowo-skutkowe)?
Skuteczność (<i>effectiveness</i>)	Obejmuje pytania badające stopień realizacji zakładanych celów, skuteczność użytych metod oraz wpływ czynników zewnętrznych na ostateczne efekty.	Czy cele programu zostały osiągnięte? Jak silny jest/był wpływ czynników instrumentów? Jakie są/były sukcesy programu? Jakie problemy napotyka/napotkał program? Czy wybrane instrumenty i rozwiązania są/były odpowiednie? Czy system zarządzania i wdrażania jest/był skuteczny? Czy instytucje poprawnie wypełniały przypisane im funkcje?
Wydajność (<i>efficiency</i>)	Bada relacje między nakładami, kosztami, zasobami (finansowymi, ludzkimi, administracyjnymi) a osiągniętymi efektami danej interwencji.	Czy podobne efekty można osiągnąć przy wykorzystaniu innych instrumentów? Czy podobne efekty można osiągnąć przy wykorzystaniu niższych nakładów finansowych?
Użyteczność (<i>utility</i>)	Jest swoistym powtórzeniem kryterium trafności. Stawia bowiem analogiczne pytania, tyle że w innym momencie (w trakcie realizacji projektu lub po jego zakończeniu).	Czy program spełnił oczekiwania adresatów? Czy przyczynił się do rozwiązania zidentyfikowanych problemów? Czy efekty programu są korzystne dla różnych grup odbiorców? Czy pojawiły się pozytywne efekty uboczne? A konkretnie: jakie środki publiczne zmobilizowane zostały przez środki pomocowe (efekt dodatkowości)? Czy pomoc zmobilizowała jakieś środki prywatne (efekt dźwigni)? Czy zrealizowana inicjatywa pobudziła kolejne inwestycje (efekty impulsu, mnożnikowe)? Czy inwestycja sprzęgła się z innymi działaniami publicznymi – wzmacniając je (efekty synergii)? Czy pojawiły się negatywne efekty uboczne? A konkretnie: czy wykonane inwestycje powstałyby i bez pomocy UE? Jeśli tak, to czy przesunięcie czasowe byłoby znaczące (efekt <i>deadweight</i> , tzw. biegu jałowego)? Czy osiągnięte pozytywne efekty nie zostały uzyskane kosztem pogorszenia sytuacji na sąsiednim obszarze (efekt przesunięcia)? Czy osiągnięte efekty nie zostały uzyskane kosztem pogorszenia sytuacji osób i instytucji nieobjętych programem (efekt substytucji)?
Trwałość (<i>sustainability</i>)	Obejmuje pytania o ciągłość efektów (przede wszystkim pozytywnych) danej interwencji w perspektywie średnio- i długookresowej.	Czy efekty działania programu są trwałe, długookresowe? Czy są odczuwalne po zakończeniu interwencji?

Źródło: opracowanie własne.

¹⁵ Oprócz podanych w tabeli kryteriów spotkać można również spójność (*consistency*, *coherence*), wybór najlepszy w danej sytuacji (*efficiency*), przejrzystość (*transparency*) czy oddziaływanie (*impact*).

Tab. 4. Wybrane metody stosowane w ewaluacji

METODA	Prospektywne (ewaluacja <i>ex ante</i>)		Retrospektywne (ewaluacja <i>mid-term</i> , <i>ex post</i>)		
	Strukturalizacja	Obserwacja	Analiza	Ocena	Ocena
Metody statyczne					
Użycie danych administracyjnych		+			+
Analiza SWOT	+			+	
Modele logiczne	+			+	
Mapy oddziaływania	+		+		
Analiza zmian udziałów			+		
Modele ekonometryczne			+		
Grupy porównawcze			+		
Analiza kosztów-korzyści				+	
Analiza kosztów-efektywności					+
<i>Benchmarking</i> (porównania)					+
Metody interaktywne					
Panele dyskusyjne, warsztaty	+			+	
Badania ankietowe		+			+
Wywiady		+			+
Studia przypadków		+			+
<i>Mindmapping</i> – mapy logiczne			+		+
Obserwacja uczestnicząca		+			+
Głosowanie kolorami	+			+	
Metoda delficka	+			+	

Źródło: opracowanie własne na podstawie: KE 1999a; KE 2003; Mathison 2004.

w ostatnich 10 latach (zob. tab. 5). Średni koszt jednego badania to 180 tys. euro (duże kontrakty dochodzą do 0,5 mln euro).

Tab. 5. Liczba ewaluacji zleczanych przez Komisję Europejską

Pole działań	1996- -1997	1998	1999	2000	2001	2002	2003	2004	2005*	SUMA
Rolnictwo	9	2	10	4	5	5	2	9	7	53
Fundusze strukturalne	4	11	6	8	2	5	6	15	4	61
Polityki wewnętrzne	61	26	39	33	36	36	63	140	104	538
Działania zewnętrzne	47	37	25	37	34	22	26	57	31	316
Administracja	0	1	1	0	1	2	7	14	4	30
Inne	0	0	0	0	10	7	13	9	bd	39
SUMA	121	77	81	82	88	77	117	244	150	1037

* Stan na lipiec 2005 r.

Źródło: Komisja Europejska – *Annual Evaluation Reviews 1998–2005*.

Patrząc z perspektywy badanego przedmiotu, można stwierdzić, że zasadniczo najczęstszym typem są ewaluacje programów bądź też poszczególnych działań polityki. Ewaluacje wstępne są wykorzystywane wewnątrz Komisji do szacowania zakładanych efektów i wyboru alternatywnych polityk (mają więc charakter formatywny). Przeważają ewaluacje średniookresowe i końcowe, ich celem jest podsumowanie i ocena danej polityki, programu i wyciągnięcie wniosków na przyszłość. One są też najczęściej publikowane. *Gros* badań jest wykonywanych przez niezależnych ekspertów lub konsorcja badawcze. Większość DG ma też swoje działy ewaluacji. Dotarcie do raportów znacznie ułatwiają publikowane od 1998 r. przez DG-Budżet *Annual Evaluation Reviews*. Przykłady ewaluacji z kilku polityk unijnych zaprezentowano w zestawieniu na następnej stronie.

Oczywiście przy takiej skali kluczowe pytanie brzmi: jak te analizy wykorzystywane są przez Komisję Europejską? Czy są one użyteczne, czy stanowią tylko część martwej biurokratycznej procedury? Dotychczas wykonano tylko jedno badanie sprawdzające wykorzystanie ewaluacji przez Komisję (Williams et al. 2002). Z badania wynika, iż ewaluacje są wykorzystywane w Komisji w bardzo zróżnicowany sposób. Bezpośrednie użycie – wykorzystanie rekomendacji technicznych – było najczęstsze. Dużą rolę odgrywał też proces ewaluacji (dyrekcje uczyły się z przebiegu samego badania), dużo mniejsze przełożenie na myślenie instytucjonalne i zmiany kierunków programów miały pojedyncze raporty (tu raczej następowała kumulacja doświadczeń – zmiany były widoczne dopiero po kilku raportach). Przykładem takiej kumulacji na poziomie polityki regionalnej jest Raport Kohezyjny – rdzeń debaty politycznej oparty na kilkudziesięciu wcześniejszych ewaluacjach. Najefektywniej wyników ewaluacji używały DG-Regio, Rolnictwo, Zatrudnienie, ECHO – a więc te o najdłuższej tradycji ewaluacji. Dużo słabiej wypadły DG-Przedsiębiorczość, Środowisko, Konkurencja.

Przykłady ewaluacji zleczanych przez Komisję Europejską

DG-Edukacja

Przedmiot badania: Program „Citizens for Europe” 2007–2013.

Typ ewaluacji: Wstępna, wewnętrzna.

Pytania: Jak wyglądałaby sytuacja bez podjęcia programu (*no action/do nothing situation*)? Jakie mogą być alternatywne działania? Czy wyznaczone cele programu odpowiadają potrzebom? Czy wskaźniki są poprawne? Jaka będzie wartość dodana działań unijnych? Czy program będzie wydajny (efektywny kosztowo)?

Główne metody: Modele logiczne, konsultacja społeczna, analiza kosztowa.

Wnioski i efekt: Program po korektach jest w trakcie zatwierdzania.

DG-Polityka Regionalna (przy współpracy z DG-Badania i Rozwój)

Przedmiot badania: Inicjatywa Pilotażowa RIS (Regionalne Strategie Innowacji).

Typ ewaluacji: Średniookresowa, zewnętrzna.

Pytania: Do jakiego stopnia RIS były efektywne w budowaniu konsensusu w regionie? Do jakiego stopnia RIS były efektywne w oszacowaniu popytu i podaży na wsparcie innowacji w regionie? Do jakiego stopnia RIS były efektywne w przygotowywaniu ram strategicznych wspierania innowacji? Czy były dobrze zarządzane i wydajne? Jakie lekcje płyną z tych pilotażowych projektów?

Główne metody: Kwestionariusz wypełniany przez instytucje wdrażające RIS, studia przypadku, wywiady na miejscu – w lokalizacjach projektów.

Wnioski i efekt: RIS okazała się efektywnym narzędziem pobudzania innowacji. Inicjatywa przeszła ze stadium pilotażu w pełny program, wprowadzono poprawki i uproszczenia nowej edycji, uwzględniono sugestie przesunięcia akcentu na sieci współpracy, stworzono internetową platformę wymiany doświadczeń między regionami RIS.

DG-Rozszerzenie

Przedmiot badania: PHARE–CBC Polska–Niemcy 1994–1999.

Typ ewaluacji: *Ex post*, zewnętrzna.

Pytania: Badanie objęło wszystkie 5 kryteriów ewaluacyjnych, zamienionych na kilkanaście pytań, m.in.: Czy program korespondował z sytuacją i procesami zachodzącymi w danych regionach przygranicznych? Jakie były przyczyny nieuczestniczenia części gmin w programie? Czy program wpłynął na zmianę postaw lokalnej społeczności wobec społeczności po drugiej stronie granicy? Czy program spełnił oczekiwania beneficjentów, czy był przydatny? Czy wystąpiły efekty mnożnikowe? Czy pojawiły się kolejne projekty inwestycyjne, kontynuacje?

Główne metody: *Desk research* dokumentów programowych i dokumentacji projektów, wizyty terenowe projektów, analiza danych z monitoringu, analiza danych regionalnych, badania ankietowe beneficjentów i potencjalnych beneficjentów, wywiady w instytucjach wdrażających program, badania ankietowe ludności z obszaru objętego programem.

Wnioski i efekt: Program miał pozytywny wpływ na rozwój obszaru przygranicznego oraz na zacieśnianie współpracy ponadgranicznej; nieprzystawalność procedur INTERREG i PHARE–CBC oraz asymetria w poziomie rozwoju sprawiły, że większość projektów miała charakter przygraniczny zamiast ponadgranicznego; problemem okazał się też dobór właściwych wskaźników informujących o wymiernych efektach projektów. Wnioski z ewaluacji posłużyły do konstruowania nowego systemu monitoringu dla programów INTERREG w Polsce, były też podstawą w ewaluacjach *ex ante* nowych programów. Opisane w ewaluacji najlepsze doświadczenia posłużyły w przygotowywaniu regionów Polski wschodniej do INTERREG.

Źródło: opracowanie własne na podstawie: KE 2005; Diez 2002; Gorzelak et al. 2004, s. 83.

3.3. Ewaluacje programów europejskich w państwach UE

Ten poziom jest całkowicie zdominowany przez politykę regionalną. Wymóg ewaluacji funduszy strukturalnych pojawił się już w 1988 r. – w pierwszym rozporządzeniu dotyczącym polityki regionalnej na poziomie UE. Zapisy, nieco modyfikowane w kolejnych okresach programowania, zachowują jednak cztery najważniejsze kwestie:

- wszystkie działania funduszy strukturalnych muszą być poddane ewaluacji *ex ante*, *mid-term* i *ex post* (dodatkowo możliwa jest, acz nieobowiązkowa, ewaluacja *on-going*),
- odpowiedzialność za zlecenie ewaluacji spoczywa na administracji krajowej zarządzającej danym programem oraz Komisji Europejskiej¹⁶,
- ewaluacja jest elementem zarządzania zarówno operacyjnego, jak i strategicznego,
- wyniki ewaluacji są zawsze upublicznione.

Liczba dotychczasowych badań ewaluacyjnych dla funduszy strukturalnych jest olbrzymia. Dla lat 1994–2006 (dla pierwszego okresu brak statystyk) to ponad 2000 badań (tab. 6). Do tego należy dodać tzw. ewaluacje *ad hoc* zlecane w ministerstwach, dotyczące poszczególnych działań, priorytetów, kwestii tematycznych – robocze i często niepublikowane (w samej Irlandii dla Celu 1 tylko dla lat 1994–1999 było ich około 150). Jeśli dodać do tego nowe kraje członkowskie, to w najbliższych latach możemy się spodziewać podwojenia liczby badań! Koszty badań pokrywane są z budżetu pomocy technicznej programów.

Tab. 6. Polityka regionalna UE – ewaluacje w latach 1994–2006 (UE15)

	Cel 1*	Cel 2	Inicjatywy Wspólnotowe**	SUMA
Rodzaj ewaluacji	Okres programowania 1994–1999			
Ewaluacje <i>ex ante</i>	147	80	299	526
Ewaluacje <i>mid-term</i>	140	72	0	212
Ewaluacje <i>ex post</i>	18	25	4	47
SUMA dla lat 1994–1999	305	177	303	785
Rodzaj ewaluacji	Okres programowania 2000–2006			
Ewaluacje <i>ex ante</i>	100	95	232	427
Ewaluacje <i>mid-term</i>	100	95	232	427
Ewaluacje <i>ex post</i> ***	100	95	4	199
Ewaluacje <i>mid-term update</i> ***	100	95	232	427
SUMA dla lat 2000–2006	400	380	700	1480
OGÓŁEM dla lat 1994–2006	705	557	1003	2265

* W przypadku lat 1994–1999 również Cel 6.

** INTERREG, URBAN, LEADER, EQUAL.

*** Zaplanowane ewaluacje.

Źródło: obliczenia własne na podstawie dokumentów programowych.

Prowadzone badania są wykonywane według coraz powszechniejszego standardu (obejmującego strukturę i metody) – przedstawianego już w poprzedniej części artykułu (MEANS). W zależności od poziomu analizy oraz przedmiotu sięga się po różne narzędzia. I tak upraszczając, można powiedzieć, że głównymi narzędziami analizy makro (Narodowe Plany Rozwoju, cała polityka) są modele ekonometryczne (np. HERMIN), przy ewaluacjach średniookresowych i bie-

¹⁶ W projekcie rozporządzenia na lata 2007–2013 Komisja Europejska ma ustanawiać metody i standardy ewaluacji.

zących poszczególnych działań przeważają natomiast wywiady i analizy raportów z monitoringu. Ewaluacje końcowe, leżące głównie w gestii Komisji, to wywiady, wizyty studyjne, studia przypadku. Ewaluacje programów o tematyce transportowej oparte są głównie na analizach kosztów-korzyści i efektywności. Pomoc „miękka” (szkolenia, współpraca) to wywiady, *benchmarking*, studia przypadku. Przykłady badań ewaluacyjnych przedstawia poniższe zestawienie.

Przykłady ewaluacji funduszy strukturalnych

Przykład 1

Przedmiot badania: 4 programy INTERREG III (Austria, Czechy, Słowacja, Słowenia, Węgry).

Typ ewaluacji: Cykl badań obejmujący ewaluację średniookresową i bieżącą, zewnętrzne.

Pytania: Czy analiza SWOT i strategia są aktualne? Czy cele są dobrze skwantyfikowane? Jakie są postępy we wdrażaniu poszczególnych priorytetów i działań? Jak funkcjonują wspólne struktury tworzenia i wyboru projektów, podejmowania decyzji?

Główne metody: Analiza dokumentów, wywiady z interesariuszami programu i projektów, seminaria i warsztaty.

Wykorzystanie badania: Badanie umożliwiło rozpoznanie potrzeb i problemów uczestników z nowych krajów członkowskich i płynne włączenie ich do działających projektów INTERREG III.

Przykład 2

Przedmiot badania: Narodowy Plan Rozwoju Irlandia 2000–2006.

Typ ewaluacji: *Ex ante* (wstępna), mieszana, 2 części – zewnętrzna i wewnętrzna.

Pytania: Badanie składało się z dwóch części – w pierwszej oszacowano potrzeby i trendy gospodarki irlandzkiej, w drugiej analizowano spójność celów, priorytetów, poprawność wskaźników, SWOT, oczekiwane rezultaty, poprawność alokacji zasobów, przejrzystość systemu zarządzania.

Główne metody: Modele ekonometryczne, modele logiczne, analiza dokumentów, konsultacje.

Wykorzystanie badania: Wprowadzono korekty do NPR, zatwierdzono jego nową wersję i na jej podstawie zbudowano Wspólnotowe Ramy Wsparcia.

Źródło: opracowanie własne na podstawie: OAR 2003; Hegarty 1999.

Ewaluacje z pierwszego okresu programowania (1988–1993) były nieskoordynowane, trudne do porównania, często były też ćwiczeniem czysto administracyjnym (spełnienie minimum wymogów Komisji). Wynikało to z różnych doświadczeń państw w zarządzaniu funduszami publicznymi, odmiennych podejść oraz nowości całego systemu zarządzania funduszami strukturalnymi.

W okresie drugim (1994–1999) jakość ewaluacji wyraźnie się polepszyła, a z tym wiązało się ich powszechniejsze użycie. Niebagatelną rolę odegrał MEANS (początkowo program wymiany doświadczeń ewaluacji polityk regionalnych, z czasem zamieniony w podręcznik DG-Regio). Ponadto na jakość ewaluacji wpłynęły też nowe systemy monitoringu, zapewniające lepszy materiał badawczy.

Obecnie krystalizują się pewne ogólnoeuropejskie standardy. Wyższa jest też jakość badań. To sprawia, że rozporządzenie na lata 2007–2013 zakłada dalszą decentralizację (z Komisji do administracji krajowych), ale też pewne wymuszenie ram – szczególnie na nowych krajach (wprowadzenie krajowych planów ewaluacji dla krajów Celu 1 – konwergencja). Coraz wyraźniej widać też, iż podaż nie nadąża za popytem. Dotychczasowy główny usługodawca – duże firmy konsultingowe – koncentruje się na największych przetargach

(powyżej 500 tys. euro) – *gros* ofert pozostawiając małym firmom i instytutom. Te zaś oferują coraz ciekawsze ewaluacje (nieszampowe, pogłębione, często o zacięciu naukowym).

Z perspektywy trzech okresów programowania można wyraźnie zaobserwować, iż praktyka ewaluacji poprzez fundusze strukturalne rozszerza się na polityki krajowe. Wynika to z trzech kwestii. Po pierwsze, z olbrzymiej liczby wymaganych badań (w którą włączone są coraz szersze rzesze urzędników i firm krajowych). Po drugie, polityka regionalna ma charakter horyzontalny (łączy różne sektory) i tym samym wymusza ewaluację w różnych polach tematycznych (od transportu przez zasoby ludzkie aż po innowacje i przedsiębiorczość). Wreszcie polityka opiera się na zasadzie dodatkowości (łączenie pieniędzy krajowych z wkładem unijnym), co sprawia, że do krajowych pieniędzy stosowane są wymogi unijne.

W przypadku funduszy strukturalnych również warto zadać pytanie o rzeczywisty wpływ ewaluacji. Jej efekty są trudno uchwytnie i bardzo zróżnicowane. Dotychczas prowadzono niewiele badań porównawczych na ten temat (koncentrowały się raczej na zróżnicowaniu podejść, systemie badań, a nie tym, co się dzieje z raportami po ich przyjęciu)¹⁷. Na podstawie wybiórczej obserwacji można stwierdzić, że przykładem najlepszej praktyki ewaluacji jako narzędzia podnoszenia skuteczności funduszy strukturalnych jest Irlandia (ewaluatorzy wymusili zmiany w strategii wykorzystania funduszy strukturalnych – z infrastruktury na zasoby ludzkie i innowacje). Pozytywnym przykładem jest również Szkocja (gdzie zbudowano spójny system monitoringu i ewaluacji, umożliwiając porównywanie poszczególnych programów i badań, a także obserwację ich postępów w czasie – ewaluacje stopniowo włączyły fundusze strukturalne w cały system polityki regionalnej), a także Austria (gdzie ewaluacje przyczyniły się do płynnego włączania nowych krajów członkowskich do działających już programów INTERREG). Wciąż negatywnym przykładem pozostaje Grecja.

Na rzeczywiste wykorzystanie ewaluacji przy funduszach strukturalnych wpływa kilka czynników. Po pierwsze, przesunięcia w cyklu zarządzania funduszami. Najszybsze przełożenie na praktykę (głównie zarządzania programem) mają ewaluacje średniokresowe. Najbardziej przydatne okazują się ewaluacje bieżące na poziomie działań. Ewaluacje średniokresowe całych programów operacyjnych są tak naprawdę wykorzystywane do projektowania nowego okresu programowania (ich wyniki publikowane są w momencie dyskusji nad nowymi programami). Wreszcie funkcja ewaluacji *ex post* ogranicza się do informowania społeczeństwa i pokazania *post factum* roli funduszy. Ewaluacje *ex post*, wbrew teorii, nie mają żadnego wpływu na kontynuowanie czy zaprzestanie danej interwencji, z tej prostej przyczyny, że są kończone w połowie trwania kolejnej edycji funduszy (np. raporty z ewaluacji *ex post* dla lat 1994–1999 były dostępne w 2004 r.).

¹⁷ Por.: Bachtler 2000; Bachtler et al. 2000; EPRC 2000.

Po drugie, co wiąże się z omawianymi już w poprzednim dziale sprawami, efekty ewaluacji zależą od kultury organizacyjnej zarówno instytucji, jak i kraju. Jak pokazuje przykład Irlandii, liczy się nie tyle tradycja ewaluacji, ile otwartość administracji – jej gotowość do innowacji i zmian. W skali instytucji najważniejsza jest ciągłość (kompetencji i osób) i związana z tym pamięć organizacyjna.

Wreszcie po trzecie, liczy się sama jakość ewaluacji. Lepsze raporty (z jasnymi rekomendacjami) są bardziej użyteczne i częściej wykorzystywane. Pod jakością kryje się tak forma badania (badania interaktywne, włączające różnych interesariuszy, przekładają się na skuteczne, praktyczne rekomendacje, programy badane zza biurka rzadko są zmieniane dzięki ewaluacjom), jak i forma prezentacji wyników (krótki klarowny raport z listą rekomendacji operacyjnych działa lepiej niż obszernie naukowe analizy). Co ciekawe, na jakość raportów wpływa nie tylko wiedza wykonujących badanie, ale i precyzyjne oczekiwania zlecających. Badania porównawcze ewaluacji INTERREG wykazały, iż słabo zarządzane programy implikują słabe ewaluacje (nawet wykonywane przez renomowane firmy) (Taylor et al. 2004).

Na zakończenie warto krótko odnieść opisywane procesy do sytuacji w Polsce. Świadomość roli ewaluacji budzi się bardzo powoli w naszej administracji krajowej – czy raczej jest budzona wymogami funduszy. Dotychczasowa praktyka jest słaba i ograniczona. Na tworzącym się rynku, dla administracji o słabo sprecyzowanych wymaganiach, nieliczne firmy często oferują produkt będący okrojoną kopią badań zachodnich. Sprawę pogarsza fakt, iż przy przetargach na ewaluację podstawowym kryterium wyboru ofert wciąż pozostaje cena, a nie zawartość merytoryczna. Niemniej widać załóżki podejścia systemowego (stworzenie jednostki oceny w Ministerstwie Gospodarki) oraz coraz silniejszy akcent na jakość stawiany przez agencje publiczne (np. szczegółowe, konkretne Specyfikacje Istotnych Warunków Zamówienia przygotowywane przez PARP). Wydaje się, że kluczowe w najbliższych latach będzie przekonanie wykonawców – firm i instytucji badawczych – o konieczności włączenia się w temat i uświadomienie im specyfiki ewaluacji (wciąż traktowanej jako zwykłe analizy lub mniej wymagająca forma audytu).

4. Wnioski

Ewaluacja staje się ogólnoeuropejskim standardem administracji publicznej. Kluczową rolę odgrywały tu i wciąż odgrywają wymagania polityki regionalnej Unii Europejskiej. Doświadczenia ostatnich 10 lat pokazują, że administracje krajowe w bardzo zróżnicowany sposób reagują na ten zewnętrzny impuls. Wyraźnie lepiej przyjmuje się on nie tylko w tych krajach, które mają wcześniejszą tradycję ewaluacji (Wlk. Brytania, Francja), ale i w tych, których administracja jest innowacyjna i otwarta na zmiany (Irlandia). Jak pokazują ich wskaźniki wykorzystania funduszy strukturalnych i wskaźniki rozwoju – jest to element szerszego sukcesu.

Patrząc na perspektywy ewaluacji w Unii Europejskiej, można liczyć na to, że na pewno liczba badań będzie rosła. Wraz z przyrostem liczby dokumentów z pewnością coraz częściej na poziomie UE zlecane będą metaewaluacje. Pozwalają one po pierwsze uchwycić wnioski i wspólne trendy w skali całej Europy, a po drugie wychwycić najlepsze i najgorsze praktyki wykonywania i użycia ewaluacji.

Na poziomie krajowym z pewnością intensyfikacji ulegnie proces interaktywnej ewaluacji w programach krajowych (ewaluacja poprzez fundusze strukturalne będzie się rozszerzać na polityki i programy krajowe). Jest to również nieuniknione w przypadku Polski. W naszym kraju, tak jak i pozostałych nowych państwach UE, postępować będzie też budowanie kultury i zaplecza instytucjonalnego ewaluacji, będące elementem szerszego procesu – wprowadzania nowoczesnego zarządzania w sektorze publicznym. W praktyce powinno być to zarówno jednoczesne doksztalcanie polskiej administracji (a więc strony zlecającej badania), jak i uświadamianie specyfiki i roli ewaluacji wśród potencjalnych wykonawców (firm i instytutów badawczych). Do transferu *know-how* ze starej UE mogą się przyczynić konsorcja polsko-zachodnich firm budowane na potrzeby dużych przetargów.

Na zakończenie trzeba stwierdzić, że istnieje wyraźna asymetria między praktyką a teorią i badaniami naukowymi poświęconymi rzeczywistemu wpływowi ewaluacji na zarządzanie i polityki UE. Wiedza o skuteczności tego narzędzia jest wciąż znikoma (często zakładana *a priori*). Ten niedobór daje jednak szerokie pole do studiów porównawczych w skali europejskiej. Narzucone standardy ewaluacji polityki regionalnej UE i odpowiedź administracji krajowych wydają się też obiecującym tematem studiów nad kulturami administracji.

Literatura

- Bachtler J., 2000, *The Evaluation of Regional Policy in the European Union – Main Report*, EPRC.
- Bachtler J., Poverari L., Taylor S., 2000, *Culture, Commitment and Capacity. A Comparative Review of Structural Funds Evaluation in the European Union*, paper for Edinburgh Conference on Evaluation, September 2000.
- Bachtler J., 2001, „Quod erat demonstrandum? Ewaluacja polityki regionalnej”, *Studia Regionalne i Lokalne*, nr 4(7).
- Chelimsky E., 1997, „The coming transformation in evaluation (w:) E. Chelimsky (red.), *Evaluation for the 21st Century. A Handbook*, Thousand Oaks: Sage.
- Colin I., 1998, *Evaluation Strategy for Scotland*, presentation for European Conference on Evaluation, Seville, http://europa.eu.int/comm/regional_policy/sources/docconf/seville/sevil_en.htm.
- CSF Evaluation Unit, 1998, *Review of Ongoing Evaluation Function in the Community Support Framework (CSF) for Ireland 1994–1999*, <http://www.csfinfo.com> (sprawdzone 7 grudnia 2004 r.).
- Diez M.A., 2002, „Evaluating new regional policies”, *Evaluation*, nr 8(3).

- European Policies Research Centre and Flaser of Allander Institute, 2000, *Methodologies Used in the Evaluation of the Effectiveness of European Structural Funds: A Comparative Assessment*, Final Report to the Scottish Executive.
- Gorzela G., Bachtler J., Kasprzyk M. (red.), 2004, *Współpraca transgraniczna Unii Europejskiej. Doświadczenia polsko-niemieckie*, Warszawa: Wydawnictwo Naukowe „Scholar”.
- Greene J.C., 2003, „Understanding social programs through evaluation” (w:) N. Denzin, Y.S. Lincoln (red.), *Collecting and Interpreting Qualitative Materials*, 2nd edition, London: Sage.
- Hegarty D., 1999, *Presentation on the Evaluation Conference*, NDP/CSF Evaluation Unit www.ndp.ie/evaluation.
- Heinrich C.J., 2003, „Measuring public sector performance” (w:) G. Peters, J. Pierre (red.), *Handbook of Public Administration*, London: Sage.
- Hermans J.L., Morris L.L., Fitz-Gibbon C.T. (red.), 1987, *Evaluator's Handbook*, Newbury Park: Sage.
- Hood Ch., 1991, „A public management for all seasons?”, *Public Administration*, nr 69(1), s. 3–19.
- Kierzkowski T., 2002, *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, Warszawa: PARP.
- Komisja Europejska, 1997, *Evaluating EU Expenditure Programmes: A Guide for Ex-post and Intermediate Evaluation*, Brussels: European Commission – DG Budget.
- Komisja Europejska, 1999a, *MEANS Collection: Evaluating Socio-economic programmes*, t. 1–6, Luxembourg: Office for Official Publications of the European Communities.
- Komisja Europejska, 1999b, *The New Programming Period 2000–2006 Methodological Working Papers*, Brussels: European Commission DG-Regio, http://www.europa.eu.int/comm/regional_policy/sources/docoffic/working/sf2000_en.htm (sprawdzone 5 lipca 2005 r.): WP2, WP 3, WP 7, WP 8, WP 8A, WP 9.
- Komisja Europejska, 2000a, *Focus on Results: Strengthening Evaluation of Commission Activities* (SEC2000/1051).
- Komisja Europejska, 2000b, *Putting the Strategic Evaluation into Practice within the Commission, Communication of the President with the Agreement of Mrs. Schreyer*, SEC (2000)1051–26/07/2000.
- Komisja Europejska, 2001a, *A Guide to the Evaluation Procedures and Structures in the Commission's External Relations Programmes*, Brussels: European Commission – EuropeAid.
- Komisja Europejska, 2001b, *Manual: Project Cycle Management*, Brussels: European Commission – EuropeAid.

- Komisja Europejska, 2003, *The Evaluation of Socio-Economic Development. MEANS Guide – Update*, European Commission – Tavistock Institute & Co.: www.evaled.info (sprawdzone 21 lipca 2005 r.).
- Komisja Europejska, 2004a, *Evaluating EU Activities. A Practical Guide for the Commission Services*, Brussels: European Commission – DG Budget.
- Komisja Europejska, 2004b, *The Mid-term Evaluations in Objective 1 & 2 Regions. Growing Evaluation Capacity*, Brussels: European Commission – DG-Regio.
- Komisja Europejska, 2004c, *Project Cycle Management Guidelines*, Brussels: European Commission – EuropeAid.
- Komisja Europejska, 2005, *New Programme for the Period 2007–2013 to Promote Active European Citizenship: „Citizens for Europe”*, SEC(2005) 442.
- Leeuw F.L., 2002, „Evaluation in Europe 2000. Challenges to the growth industry”, *Evaluation*, nr 8(1), s. 5–12.
- Mathison S. (red.), 2004, *Encyclopedia of Evaluation*, London: Sage.
- Morris L., Fith-Gibbon C.T., Freeman M., 1987, *How to Communicate Evaluation Findings*, Newbury Park: Sage.
- OAR, 2003, *Mid-term Evaluations Reports of INTERREG IIIA Austria-Czech Republic, Austria-Hungary, Austria-Slovenia, Austria-Slovakia*, Vienna: OAR-Regionalberatung GmbH.
- Olejniczak K., 2002, *Towards the Evaluation Culture in Poland – Experiences and Prospects*, referat na Konferencji Międzynarodowej Regional Studies Association „Evaluation and EU Regional Policy”, Aix-en-Provence, 31 maja-1 czerwca 2002; http://www.regional-studies-assoc.ac.uk/events/Aix2_Papers.html.
- Owen J.M., Rogers P.J., 1999, *Program Evaluation: Forms and Approaches*, London: Sage.
- Oxford English Dictionary*, 2000, Oxford Dictionary online edition.
- Patton M.Q., 1990, *Qualitative Evaluation and Research Methods*, Newbury Park: Sage.
- Rossi P., Freeman H., Lipsey M., 1999, *Evaluation – a Systemic Approach*, 6th edition, London: Sage.
- Scriven M., 1991, *Evaluation Thesaurus*, 4th edition, Newbury Park: Sage.
- Shadish W., Cook T., Leviton L., 1991, *Foundations of Program Evaluation. Theories and Practice*, Newbury Park: Sage.
- Shafritz J.M., Russell E.W., 2003, *Introducing Public Administration*, 3rd edition, New York: Longman.
- Smith P., 1995, „On the unintended consequences of publishing performance data in the public sector”, *International Journal of Public Administration*, nr 18(2–3), s. 277–310.
- Stewart J., Ranson S., 1994, „Management in the public domain” (w:) D. McKeivitt, A. Lawton (red.), *Public Sector Management. Theory, Practice and Critique*, London: Sage.

- Stufflebeam D., 2001, „Evaluation models”, *New Directions for Evaluation*, nr 89.
- Taylor S., Olejniczak K., Bachtler J., 2004, *A Study of the Mid-term Evaluations of INTERREG Programmes for the Programming Period 2000–2006*, Report for the INTERACT Secretariat, www.interact-eu.net.
- Toulemonde J., 2000, *Evaluation Culture(s) in Europe: Differences and Convergence between National Practices*, paper for *Vierteljahrshefte zur Wirtschaftsforschung*.
- Williams K., de Laat B., Stern E., 2002, *The Use of the Evaluation in the Commission Services – Final Report*, Technopolis France, www.europa.eu.int/comm/budget/evaluation/index_en.htm (sprawdzone 7 grudnia 2004 r.).
- Worthen B.R., 1990, „Program evaluation”, (w:) H.J. Walberg, G.D. Haertel (red.), *The International Encyclopedia of Educational Evaluation*, Oxford: Pergamon Press.
- Wspólnoty Europejskie, 1999, *Rozporządzenie Rady nr 1260/1999 ustanawiające przepisy ogólne w sprawie funduszy strukturalnych* (Dz. Urz., seria L 161/1).

THEORY AND PRACTICE OF EVALUATION IN THE EUROPEAN UNION

This article investigates the theoretical background (definition, function, types) as well as practice of evaluation activities in the European Union. The analysis has been led on three levels. The first one is the evaluation use by national administration for the purposes of national public programmes. Second level is the use of evaluation within the EU Institutions (namely European Commission). The last one, is the evaluation of European interventions at the country level. Accent has been put on this last level (the most developed one) that is the practice of structural funds. Author also concludes that despite the evaluation boom in Europe (that has taken place due to the requirements of EU regional policy), the real evaluation utilization and impact on the EU management are poorly recognized. There is a significant gap between number and scale of performed evaluations and the scientific attempts to recognize the real effects and the real utility of evaluation in improving the management of EU programmes and policies.