

Dr Ryszard Szarfenberg
Instytut Polityki Społecznej
Uniwersytet Warszawski
www.ips.uw.edu.pl/rszarf
r.szarfenberg@uw.edu.pl

Ewaluacja i analiza polityki społecznej (22.11.06)

Trzy rodzaje systematycznych dociekań

Systematyczne dociekania nad rzeczywistością społeczną można podzielić na trzy podstawowe rodzaje, które odróżnimy od siebie na podstawie ich ogólnych celów. Badanie ma na celu odkrywanie i opisywanie faktów empirycznych, testowanie hipotez stwierdzających tego rodzaju fakty lub powiązania między nimi. Ewaluacja, to systematyczne dociekania, które mają nam dać ocenę jakości i wartości przedmiotu dociekań. Ostatni rodzaj to analiza, nastawiona na sformułowanie rekomendacji i zaleceń dotyczących tego, co zostało przeanalizowane.

Ewaluacja wykracza ponad badania, ponieważ uwzględnia również wartościowanie (ocenie) tego, co jest przedmiotem dociekań. Ocenianie nie należy do standardowych procedur badawczych, od badaczy nie wymaga się ocen tego, co badają. Z kolei analiza wykracza poza badanie i ewaluację, gdyż uwzględnia również praktyczny wymiar tego, co badane i ewaluowane, tzn. nie tylko badamy pewien obszar rzeczywistości społecznej i oceniamy go, ale również formułujemy zalecenia, czy i jak go zmienić. Analiza zakłada aktywistyczną postawę wobec świata społecznego, możemy go nie tylko badać i oceniać, ale też świadomie kształtować według takiego lub innego planu.

Ewaluacji potrzebne są wyniki badań, gdyż ocenianie czegoś, o czym się niewiele wie nie ma nic wspólnego z systematycznymi dociekaniami. Można bez wyników badań i ewaluacji formułować różne recepty na społeczne bolączki, ale trudno je nazwać wynikami analizy. Badania nad problemem oraz ewaluacja sposobów jego rozwiązywania są konieczne dla dokonania profesjonalnej analizy i wyciągnięcia z niej wniosków praktycznych.

W nauce o polityce społecznej ważne są wszystkie rodzaje systematycznych dociekań, powinniśmy jednak pamiętać o tym, że różnią się one pod względem m.in. celów, metod, jak i warunków instytucjonalnych w jakich są dokonywane, np. badania najczęściej prowadzone są w instytucjach naukowych z silnym etosem obiektywizmu, a analiza – w instytucjach doradczych, które zwykle są zaangażowane w tworzenie rozwiązań w sferze polityki.

Ocenianie jest jedną z naturalnych czynności w życiu codziennym. Oceniamy mniej lub bardziej złożone rzeczy czy obiekty (np. na ten film warto iść, ten obraz jest piękny – ocena estetyczna), ludzi (np. jego zachowanie było niewłaściwe – ocena moralna) i instytucje (np. ten system podatkowy jest niesprawiedliwy – ocena polityczna). Są to oceny nieformalne, niezaplanowane, niemethodyczne, często nie wiemy dlaczego oceniamy dane rzeczy czy zjawiska tak, a nie inaczej i trudno nam je uzasadnić. Ludzie różnią się w swoich ocenach tych samych rzeczy, zachowań i instytucji, stąd częste dyskusje na temat ocen i próby ich uzasadniania oraz podważenia uzasadnienia tych, które są sprzeczne z naszymi. W polityce (społecznej) oceny są równie powszechne jak w innych dziedzinach życia. Różnice między oceną potoczną a ewaluacją można oddać w uproszczony sposób następująco.

Wymiary porównania	Ocenianie potoczne	Ewaluacja
--------------------	--------------------	-----------

1. Kryteria oceny	Nieustalone, niezdefiniowane, przypadkowe, niejasne, zmienne	Ustalone, zdefiniowane, określone przed przystąpieniem do oceniania i wydaniem oceny
2. Informacje o przedmiocie oceny	Wyrywkowe, zasłyszane, subiektywne	Rzetelne, obiektywne, zebrane metodami naukowymi według kryteriów oceny
3. Sposób oceny	Intuicyjny, momentalny, przypadkowy	Metodyczny, przemyślany, czasochłonny, według ustalonych procedur porównywania informacji o przedmiocie oceny z kryteriami oceny i wyciągania wniosków z takich porównań
4. Kto dokonuje oceny	Każdy	Profesjonalista, profesjonalny ewaluator

Widzimy tu powiązanie ewaluacji z badaniem, które ma dostarczać obiektywnych informacji o przedmiocie ewaluacji (nazywanym czasem „ewaluandum”), a także konieczność przemyślanego wyboru kryteriów ewaluacji oraz opracowania metody porównywania obiektywnych informacji z kryteriami oraz wyciągania stąd wniosków dotyczących oceny ewaluandum.

Związki analizy z ewaluacją są bardzo ścisłe, ponieważ rekomendacja, zalecenie lub rada zwykle związane są z problemem, który wymaga rozwiązania. Zamawiający analizę chce otrzymać radę w sytuacji, gdy nie wie, jak rozwiązać problem lub chce się upewnić, że rozwiązanie, które znalazł jest dobre. W przypadku polityki społecznej mamy do czynienia z problemami i kwestiami społecznymi, a ich rozwiązania to regulacje i programy społeczne. Uzasadnienie rekomendacji wymaga co najmniej ewaluacji alternatywnych rozwiązań i porównania ich wyników ze sobą. Wymaga to m.in. przewidywania skutków ich realizacji na podstawie dociekań badawczych. Wnioski z analizy tego rodzaju to wskazanie na mocne i słabe strony poszczególnych programów i pokazanie ich względnej pozycji od najlepszych do najgorszych. Daje to możliwość dokonania świadomego wyboru przez tego, kto zamówił analizę.

Polityka, strategia, program, projekt

Przedmiotem badań, ewaluacji i analizy, jaki będzie nas tu interesował jest oczywiście polityka społeczna. Ewaluacji i analizie częściej podlegają jednak strategie, programy lub projekty, a nie sama polityka.

Wszystkie te pojęcia (polityka, strategia, program, projekt) mogą występować w trzech ogólniejszych znaczeniach: 1) zamierzeń, które zostały bezpośrednio wyrażone i skonkretyzowane, np. zamiar pomocy niepełnosprawnym dzieciom w nauce zapisany w jakimś dokumencie wraz z uzasadnieniem i nakreśleniem kierunków działań pomocowych (plan), 2) realizacji tych zamierzeń, np. udzielanie pomocy niepełnosprawnym dzieciom w nauce, 3) zarówno zamierzeń, jak i ich realizacji, np. strategia czy projekt pomocy dzieciom niepełnosprawnym w nauce oznacza nie tylko treść dokumentu, w którym skonkretyzowano takie zamierzenia i zaplanowano ich realizację, ale również działania, które zostały podjęte w celu zrealizowania tej strategii, czyli samo udzielanie pomocy.

Relacje między tymi pojęciami można uporządkować w taki sposób: 1) strategia określa kierunki działań i ogólne cele, jakie chcielibyśmy osiągnąć (w tym znaczeniu używa się też słowa „misja”), 2) program (operacyjny) przekłada ogólne cele strategii na cele bardziej szczegółowe i konkretne (cele operacyjne) oraz określa ogólne sposoby ich osiągnięcia w danych warunkach (nazywane czasem „działaniami”), 3) projekt to już szczegółowe zadania do wykonania i szczegółowo określony plan ich realizacji oraz konieczne do tego zasoby (m.in. personel, harmonogram, budżet). Całość można nazwać polityką.

Inny sposób odróżnienia polityki od programu lub projektu polega na wskazaniu, że ta pierwsza jest stałą działalnością bez wyznaczonego momentu zakończenia, np. polityka bezpieczeństwa socjalnego w postaci ubezpieczenia społecznego, a w drugim przypadku

mamy do czynienia z działalnością w konkretnym przedziale czasu, np. projekt pomocy małym podmiotom gospodarczym trwający od stycznia 2006 do marca 2007. W skrócie, polityka podobnie jak program czy projekt ma swój początek w czasie, ale w odróżnieniu od nich nie ma określonego z góry końca.

Możemy prześledzić te relacje na przykładzie Strategii Przeciwdziałania Ubóstwu i Wykluczeniu Społecznemu UE z 2000 r. Komisja Europejska uzgodniła wówczas cele oraz mierniki ich realizacji (strategia), następnie państwa członkowskie przygotowały Krajowe Plany Działań (KPD)¹, które przedstawiały, jak zamierzają osiągać wskazane cele (były to więc programy realizacji strategii). W Polsce powstał w tym kontekście Fundusz Inicjatyw Obywatelskich (FIO)², z którego finansuje się konkretne projekty proponowane przez organizacje pozarządowe.

Przykłady te są typowe dla takiej polityki społecznej, którą przygotowują i ustalają władze publiczne (z mniejszym lub większym udziałem dialogu społecznego), ale realizują organizacje pozarządowe w całości lub w części za środki publiczne. Tradycyjnie jej realizatorami były i nadal są przede wszystkim organizacje publiczne podporządkowane administracji rządowej lub samorządowej, chociaż mniej lub bardziej samodzielne, np. ośrodki pomocy społecznej, urzędy pracy, publiczne szkoły czy zakłady opieki zdrowotnej. Prowadzą one stałą działalność, której treść i formę wyznaczają ustawy i akty prawne niższego rzędu. Czy w tym przypadku również mamy do czynienia ze strategiami, programami i projektami? Jeżeli na podstawie ustawodawstwa można takiej działalności przypisać jakieś cele, określić sposób ich realizacji, zasoby, które zużywa oraz zidentyfikować proces jej wykonywania oraz jakieś tego skutki, to niewątpliwie spełnia ona wszystkie warunki by uznać ją za strategię czy program społeczny. Każda polityka, aby w ogóle nią być, powinna spełniać te warunki.

Przyjmujemy tutaj, że polityką, strategią, programem czy projektem jest zawsze pewna całość, która składa się m.in. z celów, planów ich realizacji oraz ich wykonania. Można oczywiście badać, ewaluować i analizować same cele, metody ich realizacji lub same plany, ale najbardziej interesująca jest właśnie całość, wraz z realizacją i wykonaniem, która od planów może odbiegać i przynosić niekoniecznie i nie tylko to, co zamierzano osiągnąć. Ponadto każdy program i projekt ma swoją dynamikę, czyli jest procesem, na który składa się co najmniej przygotowanie programu oraz jego realizacja lub wykonanie. Efektem przygotowań jest m.in. plan realizacji, a efektem wykonania planu m.in. określona ilość i jakość produktu (np. dobra lub usługi udzielone klientom programu).

Struktura logiczna programu i jego teoria

Program lub projekt jest złożoną całością, która musi być przygotowana, uruchomiona, podtrzymywana w działaniu i zakończona, ma też różnego rodzaju skutki m.in. zamierzone i niezamierzone. Każdy program niezależnie od jego celów można schematycznie przedstawić w następujący sposób³:

Zasoby => działania => produkty => klienci/uczestnicy/odbiorcy programu => wyniki w krótkim okresie => wyniki w średnim okresie => wyniki w długim okresie

¹ Informacje o KPD na rzecz Integracji Społecznej można znaleźć na stronach Serwisu MPiPS Działania na rzecz Integracji Społecznej www.mps.gov.pl/integracja.

² Informacje o FIO i projektach realizowanych w jego ramach znajdują się na stronach Departamentu Pożytku Publicznego MPiPS www.pozYTEK.gov.pl.

³ Schemat za: J.S. Wholey i in. red. „Handbook of Practical Program Evaluation”, II wydanie, Jossey-Bass, San Francisco 2004, s.9.

Zasoby traktuje się bardzo szeroko, jest to wszystko czego potrzebujemy, żeby prowadzić zaplanowane działania, a więc m.in. ludzie, pieniądze, informacje (np. o potrzebach klientów/uczestników), czas. Działania oznaczają wszystko to, co musi być zrobione, aby otrzymać produkt, który może przybrać postać m.in. usług, rzeczy, pieniędzy. Produkt jest oczywiście przeznaczony dla kogoś, czyli klientów, uczestników lub odbiorców programu (cele dotyczące produktu nazywane są w kontekście programów UE celami operacyjnymi). Bez klientów trudno myśleć o wynikach programu, czyli o zmianach i korzyściach, które dla nich lub dla kogoś innego wynikają z działań i produktów programu. Wyniki z kolei dzieli się na tym schemacie według czasu ich wystąpienia, ale są też inne podziały, np. wyniki bezpośrednie (nazywane też rezultatami, w UE dotyczą ich cele strategiczne ang. *specific*) i wyniki pośrednie (nazywane też oddziaływaniem lub wpływem, w UE dotyczą ich cele globalne).

Powyższy lub podobne schematy służą niekiedy do odróżniania audytu, który dotyczyć ma zasobów, działań i produktów od monitoringu, który obejmuje poza tymi elementami również wyniki krótkoterminowe lub rezultaty. Ewaluacja dotyczy zaś wszystkich elementów włącznie z wynikami średnio i długoterminowymi.

Przykładowo, przeszkolenie nauczyciela akademickiego w prowadzeniu warsztatów daje w krótkim okresie wynik w postaci przyrostu jego wiedzy i umiejętności, następnie nauczyciel stosuje tę wiedzę i umiejętności prowadząc warsztaty i są to wyniki w średnim okresie, w końcu korzystają na tym studenci uczestniczący w tych zajęciach (oddziaływanie programu szkoleń na studentów). Klientem programu był w tym przypadku nauczyciel, ale korzyści odniosą również odbiorcy usług edukacyjnych dostarczanych przez niego, ponieważ program ma podnosić ich jakość. Na tej podstawie można też odróżnić klientów bezpośrednich (w przykładzie - nauczyciele) od klientów pośrednich lub ostatecznych (czyli studentów). Audytem byłaby ocena wykorzystania i wystarczalności potrzebnych do jego realizacji zasobów oraz zgodności zadań wykonywanych przez personel programu z planem ilości i jakości usług. Monitoring obejmowałby dodatkowo obserwację poziomu wiedzy i umiejętności nauczycieli, co w przypadku typowych programów edukacyjnych polega na przeprowadzaniu różnego rodzaju sprawdzianów i analizie ich wyników, a także ankiet oceniających produkt programu wśród klientów (np. zadowolenie nauczycieli ze szkolenia). Ewaluacja natomiast dotyczyłaby również tego, czy i jak nauczyciele wykorzystali nabyte w wyniku szkolenia kwalifikacje oraz, jaki to miało wpływ na studentów.

Konkretna struktura każdego programu lub projektu opiera się na teorii programu, czyli często nie wypowiedzianych wprost założeniach, że „odpowiednie zasoby przekształcone we właściwe działania dla odpowiednich ludzi doprowadzą do wyników, dla osiągnięcia których zaprojektowano program”. Określenie, jakie zasoby będą odpowiednie, jakie działania właściwe i kto ma być ich klientem programu, aby osiągnąć zamierzone wyniki opiera się na takich założeniach. Rozwój teorii programu przed jego realizacją służy do przewidywania, jaki wpływ możemy za jego pomocą osiągnąć, a po jego realizacji pomaga wyjaśniać, w jaki sposób program przyczynił się do osiągnięcia takich, a nie innych wyników. W przytoczonym przykładzie szkolenia nauczyciela akademickiego na teorię programu mogą się składać hipotezy w rodzaju: jeżeli program da nauczycielowi informacje z zakresu wiedzy i umiejętności warsztatowych, to podniesie się poziom jego wiedzy na temat prowadzenia zajęć; jeżeli podniesie się poziom wiedzy nauczyciela na ten temat, to poprawi się jakość jego usług; jeżeli program poprawi umiejętności nauczyciela, to poprawi się jakość usług, jakich dostarcza on studentom; jeżeli nauczyciel zastosuje informacje i umiejętności zdobyte podczas programu, to podniesie się poziom wiedzy i umiejętności studentów.

Kryteria ewaluacji

Źródłem kryteriów oceny są wartości, czyli to, co jest cenione w danej kulturze lub cenione uniwersalnie przez ludzi, np. dobrobyt, bezpieczeństwo, wolność, godność, piękno, pokój, sprawiedliwość itd. Sądy wartościujące to inaczej wyrażanie oceny czegoś, np. X jest sprawiedliwe, bezpieczne, dobre dla pokoju, dobre dla dobrobytu. Można przyjąć, że wartości i oceny mają być dopiero odkryte w kontekście danego programu i jest to jedno z zadań ewaluacji. Nie zmienia to jednak faktu, że pewne wartości uznaje się za ewaluacyjnie uniwersalne, czyli istotne w każdej ewaluacji programu społecznego. Są one ważne niezależnie od tego, z jakim programem mamy do czynienia. W kontekście ewaluacji dokonywanych według wskazówek UE mowa o trafności (*relevance*), skuteczności (*effectiveness*), efektywności lub wydajności (*efficiency*), użyteczności (*utility*) i trwałości (*sustainability*).⁴

Nie pojawiły się one w przykładowym wyliczeniu wartości wyżej, gdyż nie mają znaczenia jako ostateczne ludzkie cele. Występują jednak jako cechy uniwersalnie pożądane każdego działania, za pomocą którego osiąga się takie cele. Tego rodzaju cechy nazywa się łącznie „sprawnością działania”, albo „wartościami prakseologicznymi” (prakseologia to nauka o sprawnym działaniu). Ze względu na to, że celów w życiu i społeczeństwie jest wiele, a zasoby umożliwiające ich osiągnięcie są ograniczone, cenimy działania skuteczne i efektywne, czyli takie, które pozwalają nam osiągać założone cele przy jak najmniejszym zużyciu zasobów. Skuteczność i efektywność to główne wartości prakseologiczne. Wielość celów powoduje, że ważny jest również ich wybór i hierarchizacja (uporządkowanie według ważności), dlatego cenimy trafność ich doboru, czyli to, jak odpowiadają zidentyfikowanym potrzebom. Osiągnięcie celów nietrafnie dobranych nie przyczyni się do ich zaspokojenia. Niezależnie od celów ważna jest więc zgodność wyników programu z potrzebami, czyli to na ile przyczynił się do ich trwałego zaspokojenia (użyteczność i trwałość).

Potrzeby definiuje się niekiedy jako różnicę między tym, co jest, a tym, co być powinno. Ich identyfikacja wymaga więc nie tylko zobrazowania rzeczywistej sytuacji (do czego sprowadza się często diagnozę), ale również określenia stanów pożądanych, a do tego służą nam wartości, czyli ostateczne cele ludzkiego życia lub wielostronnie ujmowane dobro człowieka. Biorąc to pod uwagę ważne jest nie tylko to, czy cele odpowiadają potrzebom, ale również to, czy właściwie zidentyfikowano potrzeby. Zależy to z kolei od poprawnego określenia stanów pożądanych, czyli wartości, stanów faktycznych oraz porównania ich ze sobą. Pominięcie wartości może prowadzić do tzw. „racjonalnej głupoty”. Możemy mieć doskonale sprawne programy, które zaspokajają określone z góry potrzeby danej grupy uczestników, ale jednocześnie powodują niezaspokojenie innych potrzeb, a ich produkty są szkodliwe z punktu widzenia różnych wartości, np. sprawiedliwości, wolności, dobrobytu zwierząt itp.

Przynajmniej część kryteriów prakseologicznych zobaczymy, gdy zestawimy ze sobą poszczególne elementy programu z wartościami, potrzebami i celami.

	wartości	potrzeby	cele	zasoby	nakłady	działania	produkt	wynik	wpływ
wartości		Tf			A	L	A		
potrzeby	Tf		Tf					U	U, Tw
cele		Tf						S	S
zasoby					O, W				
nakłady	A			O, W		W, O	E		
działania	L, A				W, O				
produkt	A				E				
wynik		U	S						
wpływ		U, Tw	S						

⁴ „Evaluating EU Expenditure Programmes: A Guide”, pierwsze wydanie, XIX Dyktoriat Generalny, Komisja Europejska, styczeń 1997, s.18.

Oznaczenia kryteriów ewaluacji w tabeli: Tf – trafność, U – użyteczność, Tw – trwałość, S – skuteczność, E – efektywność, W – wystarczalność, O – oszczędność, L – legalność, A – akceptowalność.

Wynik to oddziaływanie bezpośrednie programu w krótkim okresie, a wpływ – oddziaływanie pośrednie w dłuższym czasie. Odróżnienie zasobów od nakładów polega na tym, że te pierwsze są tym, czym dysponujemy, a te drugie są tą częścią zasobów do dyspozycji, która została zużyta na realizację programu.

Cztery dodatkowe kryteria, które zostały uwzględnione w tabeli to oszczędność, wystarczalność, legalność i akceptowalność. Im mniej zużyjemy zasobów do dyspozycji na realizację programu tym nasze działania będą bardziej oszczędne (gospodarne). Zbyt daleko posunięta oszczędność prowadzić może jednak do nieskuteczności w osiąganiu celów programu, gdyż nie pozwala na wykonanie wszystkich koniecznych działań. Porównanie zasobów z planowanymi nakładami może też prowadzić do wniosków o ich wystarczalności; jeżeli zasoby są zbyt małe w stosunku do planowanych nakładów mamy do czynienia z ich niewystarczalnością. Może to prowadzić do zaniechania realizacji programu lub do zmniejszenia jego skuteczności.

Przypomnijmy, że działania są tym wszystkim, co należy zrobić, aby zasoby przekształcić w produkt. Licznych działań wymaga też przygotowanie programu lub projektu i zdobywanie zasobów na jego realizację. Wartości moralne określają natomiast względnie powszechne wzorce i normy działań (w większości uwzględniane w treści przepisów prawnych), których przekraczanie lub naruszanie jest mniej lub bardziej nielegalne. Stąd legalność jest ważnym kryterium oceny każdego programu społecznego. Ostatnim z dodatkowych kryteriów jest akceptowalność. Ponieważ zasoby, którymi dysponuje społeczeństwo są ograniczone, wielkość nakładów wymaga akceptacji jego reprezentantów (np. rządzących polityków). W debacie publicznej argument o tym, że za dużo wydajemy pieniędzy na programy społeczne często pada z ust ekonomicznych liberałów. Z innych względów produkty programu również powinny być akceptowalne nawet, gdy są one legalne, np. usługi polegające na zaspokajaniu potrzeb seksualnych osób niepełnosprawnych przebywających w instytucjach zamkniętych mogą być nie do zaakceptowania dla konserwatywnych polityków.

Ponieważ w tabeli nie było już miejsca nie uwzględniono tam planu programu i kryteriów ewaluacyjnych, które mogą się z tym wiązać. Zgodność zasobów, nakładów i działań z planem jest podstawą do wyróżnienia kryterium, które można nazwać planowością programu. Im bardziej te elementy programu odbiegają od wcześniej przyjętego planu, tym mniejszy poziom jego planowości. Częstym problemem może być w tym kontekście przekraczanie zaplanowanych nakładów, gdy okaże się, że zostały one nieoszacowane lub nie uwzględniono działań, które później okazały się konieczne. Ponadto plany mają swoje własne kryteria oceny, np. celowość, wykonalność, zgodność wewnętrzną, operatywność, racjonalność, elastyczność, określoność czasowa, kompletność, określoność czasowa, optymalność szczegółowa (cnoty planu T. Kotarbińskiego⁵).

Rodzaje ewaluacji

Najczęściej stosowany podział dociekań ewaluacyjnych opiera się na tym, po której fazie w procesie „od pomysłu do przemysłu” dokonuje się oceny programu. Proces ten można podzielić na dwa główne etapy: 1) przygotowanie zakończone dokumentami takimi, jak m.in.

⁵ T. Kotarbiński „Sprawność i błąd”, Państwowe Zakłady Wydawnictw Szkolnych, Warszawa 1970, ss. 157-164.

plan programu; 2) realizację, podczas której za pomocą zasobów wytwarzany i dostarczany jest produkt. Ewaluacja pierwszego etapu może dotyczyć przebiegu samego procesu przygotowań oraz jego wyników, czyli treści dokumentów takich jak diagnozy, strategie, plany, harmonogramy, budżety itd. Proces dochodzenia do nich mógł być niedemokratyczny, przeprowadzony bez dyskusji i konsultacji, pełen konfliktów i napięć, a ich treści mogą być niejasne, niespójne, nielogiczne, nierealistyczne, ogólnikowe i posiadać wiele innych luk i braków. Ewaluacja może je wykryć i dać szansę na poprawienie zarówno procesu przygotowawczego, jak i samych jego wytworów. Tego rodzaju dociekania nazywa się ewaluacją programu *ex ante*, czyli przed przystąpieniem do jego realizacji.

Po przystąpieniu do realizacji programu, a przed zakończeniem jego działania przeprowadza się czasem ewaluację nazywaną *interim* lub *in media res*, czyli w trakcie realizacji (określenie „*mid-term*” dotyczy ewaluacji dokonanej w połowie okresu działania programu, a *on-going* – ewaluacji bieżącej lub okresowej). Służy ona ocenie programu jeszcze przed jego zakończeniem i może być przydatna do poprawienia jego działania lub wcześniejszego zakończenia w razie wystąpienia istotnych problemów realizacyjnych. Gdy działanie programu zostanie zakończone przeprowadza się ewaluację nazywaną *ex post*. Oczywiście im później zostanie ona przeprowadzona tym większe szansa na uwzględnienie długookresowych wyników programu. Stąd też istotną rolę odgrywają dociekania ewaluacyjne, które wykonuje się periodycznie przez kilka lat od zakończenia programu.

Niekiedy ewaluację jako taką błędnie utożsamia się wyłącznie z ewaluacją *ex post*. Związane jest to z przekonaniem, że ludzkie działania należy oceniać w świetle ich skutków i wpływu, jaki miały, a nie intencji, jakie im przyświecały, doskonałości planu, według którego je prowadzono lub tego, jak dalece się go trzymano. Nie da się jednak zaprzeczyć, że intencje i zamierzenia wpływają na plany, a ich realizacja na wyniki programu, tzn. ogólnikowe intencje tak samo jak nierealistyczne plany i niedokładne ich wykonanie z dużym prawdopodobieństwem nie doprowadzą również do osiągnięcia tego, co zamierzono. Ewaluacja całościowa (*ex ante*, *interim* i *ex post*) jest ważna z perspektywy troski o to, aby programy społeczne działały sprawnie i zaspokajały uzasadnione potrzeby.

Istotny podział ewaluacji ma u podstaw różne cele, jakie się jej przypisuje.⁶ Celem ewaluacji podsumowującej jest oszacowanie jakości i wartości programu społecznego. Jakość dotyczy samego programu i jego charakterystyk, np. jakość dostarczanych usług może być wysoka lub niska, natomiast wartość ma on z perspektywy zewnętrznej, np. wartość dla klienta, dla społeczności lokalnej. W trakcie tego rodzaju ewaluacji ocenia się m.in. wyniki programu (zamierzone i niezamierzone, pozytywne i negatywne) oraz inne jego charakterystyki.

Z kolei ewaluacja usprawniająca nastawiona jest na poprawianie programu i jego realizacji. Celem jest uzyskanie informacji, które pomogą zmodyfikować i ulepszyć działanie programu. Ważne są w tym kontekście aktualne informacje o produktach i o procesie realizacji służące do zidentyfikowania elementów programu, które nie spełniają standardów i formułowanie zaleceń naprawczych lub usprawniających. Wykorzystuje się w tego rodzaju ewaluacji m.in. porównywanie działań programu do najlepszych praktyk (podobne programy, które zostały uznane za sukces i wzór do naśladowania). Ewaluacja tego rodzaju ma zbliżone funkcje do analizy, chociaż zalecenia nie dotyczą całego programu, a tylko jego elementów. Wyróżnia się też kilka innych celów ewaluacji, np. kontrola zgodności programu z różnego rodzaju przepisami (ewaluacja kontrolna), wzbogacenie wiedzy o programie (ewaluacja poznawcza).

W zależności od tego, kto dokonuje ewaluacji można podzielić ją na ewaluację wewnętrzną i zewnętrzną. Pierwszej dokonuje wyodrębniona jednostka, stanowiąca jednak

⁶ M. Mark, G. Henry i G. Julnes „Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs”, Jossey-Bass, San Francisco 2000, ss.54-58.

część samego programu lub organizacji, która odpowiada za jego działanie. Druga przeprowadzana jest przez ewaluatorów, pracujących dla innej organizacji, która nie ma żadnych związków z administratorami programu, ani z jego personelem. Czasem wyróżnia się też autoewaluację, gdy ewaluator należy do personelu programu, czyli w jego planie przewidziano takie stanowisko. Autoewaluacja bez zatrudnienia ewaluatora może polegać na tym, że każdy członek personelu ma obowiązek przeprowadzać ewaluację swojej własnej pracy lub pracy własnej jednostki organizacyjnej.

Udział personelu programu, a także jego klientów w zarządzaniu lub w konsultowaniu planu i działań ewaluacyjnych służy do ich podziału na współzarządzane, uczestniczące i nieuczestniczące.

Ze względu na kryterium można podzielić dociekania ewaluacyjne na ewaluacje cząstkowe, np. skuteczności, efektywności, trafności itd. W zależności od liczby kryteriów możemy mówić o ewaluacjach wielokryterialnych i jednokryterialnych. Kryteria mogą być ustalone z góry przez zamawiającego, ale mamy też ewaluacje, w których są one dopiero odkrywane i ustalane w trakcie ewaluacji. Częstym źródłem kryteriów są skonkretyzowane cele programu w postaci poziomu wskaźników, które mają być osiągnięte, ale uzasadniona jest też ewaluacja, która czerpie je z innych źródeł (*goal-free*)⁷, np. z określenia potrzeb klientów programu. Są też cele niezgodne z etosem ewaluacji, np. zdobywanie dowodów za lub przeciwko danemu programowi społecznemu⁸.

Źródłem podziału może być również to, czy ocenie podlega proces programu (ewaluacja procesu), czy też jego rezultaty (ewaluacja wyniku). Biorąc pod uwagę podstawowe elementy składowe każdego programu ewaluacji mogą podlegać zasoby, działania i produkt i stąd możliwość wyróżnienia kolejnych rodzajów dociekań ewaluacyjnych. Wyróżnione podziały przedstawiam w tabeli

Podstawa podziału	Podział ewaluacji na rodzaje
fazy procesu programu	ex ante / interim (w tym mid-term, on-going) / ex post
cel	podsumowująca / usprawniająca / kontrolna / poznawcza
pozycja ewaluatora wobec programu	wewnętrzna / zewnętrzna
udział personelu, klientów	współzarządzana / uczestnicząca / nieuczestnicząca
liczba kryteriów	jednokryterialna / wielokryterialna
źródło kryteriów	kryteria określone z góry (m.in. poprzez cele programu) / kryteria odkrywane
składniki programu	planu / zasobów (w tym personelu, kosztów itp.) / działań / produktu (dobra i/lub usługi)
rodzaj wyników	wyników w krótkim okresie / w średnim / w długim

Metody badawcze w kontekście ewaluacji

Ze względu na to, że ewaluacja opiera się na wynikach dociekań badawcze ich metody uznaje się często również za metody ewaluacji. Chodzi o stosowanie technik zbierania i analizowania różnego rodzaju danych, ale w przypadku ewaluacji o jakiej tu mowa mamy zawsze do czynienia z danymi na temat programu społecznego. W jednym z podejść

⁷ Krytyka ewaluacji w oparciu o cele określone na wskaźnikach, m.in. E.J. Davidson „Evaluation Methodology Basics”, SAGE Publications, Thousands Oaks, London, New Delhi 2004, ss.24-27.

⁸ B. Worthen „Program evaluation”, w: H. Walberg i G. Haertel red. „The International Encyclopedia of Educational Evaluation”, Pergamon Press, Toronto 1990, s.42.

podzielono metody wspólne dla badań i ewaluacji na cztery rodzaje według celu ich stosowania: metody opisu, klasyfikacji, dociekania przyczyn i dociekania wartości.⁹

Metody stosowane do opisu umożliwiają m.in. pomiar cech klientów programu, jego produktu, zasobów lub pozycji klienta na skalach potencjalnych wyników programu. Przykładowe pytania, na jakie odpowiada się za pomocą tych metod: jaki procent uczniów miało więcej niż dziesięć dni nieobecności w szkole w poprzednim roku szkolnym, czy uczniowie z większą liczbą nieobecności różnią się czymś więcej od pozostałych uczniów, jaki procent dzieci byłych klientów pomocy społecznej nie ma ubezpieczenia zdrowotnego, czy dany program dostarcza usług tym, którzy są do nich uprawnieni na podstawie przepisów prawa? Dane potrzebne do sporządzania opisu mogą być zbierane poprzez rozmowy (informacje przekazywane ustnie przez personel i uczestników), przeprowadzanie sondaży (ankiety do wypełnienia), obserwację i analizę dokumentów (wytworzonych w trakcie realizacji programu).

Metody klasyfikacji (dzielenie podobnych obiektów na klasy, typy, rodzaje, kategorie) stosowane są do analizy podobieństw i różnic oraz wyróżniania na tej podstawie np. rodzajów programów czy typów klientów: „Klasyfikacja pozwala zajrzeć głębiej pod powierzchnię zaobserwowanych różnic w celu odkrycia różnic strukturalnych, które odpowiedzialne są za powstawanie ważnych regularności”. W metodach tego rodzaju mamy do czynienia z dwiema strategiami: 1) dedukcyjną (a priori), gdy wyróżniamy pewne kategorie (klasy, typy) z góry w oparciu o m.in. teorię lub wcześniejsze badania, 2) indukcyjną, gdy kategorie (klasy, typy) wyprowadzamy na podstawie danych opisowych. W tym drugim przypadku używa się technik ilościowych takich, jak m.in. analiza skupień (ang. *cluster analysis*), oparta na pomiarze podobieństwa między przypadkami; analiza czynnikowa; taksonometria rozwijana przez psychologów.

Metody używane dla zbadania i testowania związków przyczynowo-skutkowych (np. pomiędzy usługami czy produktami programu a funkcjonowaniem jego klientów) lub badania mechanizmów, poprzez które dochodzi do takich a nie innych skutków programu to metody dociekania przyczyn. Są one bardzo istotne dla ewaluacji, gdyż na ich podstawie orzeka się o tym, co było a co nie było wynikiem programu. Przykładowe pytania, na które odpowiada się za pomocą takiej metody: jaki jest wpływ reformy pomocy społecznej na dobrobyt ubogich dzieci, czy programy wymiany strzykawek redukuje rozpowszechnianie się HIV/AIDS, który spośród wielu programów nauczania w przedszkolach najlepiej służy szansom edukacyjnym dzieci, który z alternatywnych programów przeznaczonych dla młodocianych przestępców pozwala na osiągnięcie przez nich lepszych wyników w społecznym funkcjonowaniu? Poza stwierdzeniem przyczyny i skutku ważne jest również określenie mechanizmu, który je ze sobą łączy (dlaczego dany program społeczny miał takie a nie inne wyniki).

Znaczenie analizy przyczynowo-skutkowej w kontekście ewaluacji programów społecznych zostanie dodatkowo wyjaśnione na prostym przykładzie. Załóżmy, że w pewnej społeczności jest 100 bezrobotnych absolwentów szkół średnich i przyjęto program Pierwsza Praca, żeby pomóc tej grupie znaleźć pracę. Po roku działania programu stwierdzono, że liczba bezrobotnych absolwentów zmniejszyła się o 30 osób. Stwierdzenie, że jest to wynik programu Pierwsza Praca może być błędem. Dlaczego? Z dwóch powodów. Założenie, że w programie uczestniczyło te 30 osób może być fałszywe. Założenie, że uczestnik programu zdobył pracę dzięki programowi również może być fałszywe. Prawda może być taka, że z tych 30 bezrobotnych, którzy znaleźli pracę uczestniczyło w programie 20, a z tych 20 tylko 10 zdobyło pracę w wyniku tego, że uczestniczyli w programie (np. w programie byli szkoleni na budowlańców, a znaleźli pracę w ochronie). Zawsze trzeba brać pod uwagę to, że

⁹ Charakterystyka metod głównie za: M. Mark, G. Henry i G. Julnes „Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs”, Jossey-Bass, San Francisco 2000, ss.173-323.

program nie jest jedynym czynnikiem oddziałującym na sytuację ludzi, dla dobra których go przyjęto. W skrajności wszystkie domniemane wyniki programu mogłyby być skutkami innych czynników, np. gdyby programu z naszego przykładu w ogóle nie zrealizowano pracę znalazłoby 20 bezrobotnych absolwentów. Jeżeli ktoś nie wiedział, że program nie został zrealizowany, mógłby wyciągnąć wniosek, że jest to wynik owego programu. W praktyce informacje na temat skutków programów społecznych uzyskuje się za pomocą badań o charakterze eksperymentalnym. W naszym przykładzie można dobrać losowo dwie podobne w liczebności i składzie grupy bezrobotnych absolwentów, ale tylko jedna z nich będzie uczestniczyła w programie, drugą grupę nazywa się grupą kontrolną. Po zrealizowaniu programu badamy sytuację obu grup, co pozwala sformułować wniosek o wynikach programu, np. jeżeli w grupie uczestniczącej w programie znalazło pracę 10 osób, a w grupie kontrolnej 5, to wynik programu jest taki: 5 osób znalazło pracę dzięki programowi.

Metody stosowane do określenia i analizy wartości oraz sądów wartościujących to metody dociekania wartości. O ile analiza przyczynowo-skutkowa może dać nam informację o skutkach (wynikach) programu, to nie daje odpowiedzi na pytanie o to, za pomocą jakich kryteriów mamy ocenić te skutki, czyli ocenić czy program osiągnął sukces, czy też porażkę, albo które ze skutków są ważniejsze od innych. Wartości i sądy wartościujące można odkrywać m.in. poprzez sondaże opinii dotyczące oceny danego programu przeprowadzane wśród jego interesariuszy (administratorów, sponsorów, personelu), klientów, opinii publicznej; wywiady grupowe; symulowanie demokratycznej debaty, czyli przedstawianie grupie problemu do rozwiązania (grupy dyskusyjne); krytyczne przeglądy teorii filozoficznych w docieraniu do często ukrytych wartości leżących u podstaw pewnych programów i polityk, np. zastosowanie teorii J. Rawlsa, każde nam szukać wpływu programu na sytuację grup, które znajdują się w najgorszej sytuacji, a teorii neomarksistowskich – wpływu na strukturę władzy w społeczeństwie.

Metoda ewaluacji

Poza metodami, które są wspólne dla dociekań badawczych i ewaluacyjnych, są również metody dedykowane wyłącznie dla ewaluacji. W badaniach nie jest konieczne określanie jakości i wartości przedmiotu badania, a w przypadku ewaluacji to właśnie decyduje jej sensie i specyfice. Można zebrać mnóstwo różnych informacji o programie społecznym i jego wynikach, które jednak same w sobie nic jeszcze nie mówią o jego jakości i wartości, np. informacja uzyskana za pomocą metod opisowych i przyczynowych o tym, że program pomocy dla 1000 bezrobotnych kosztujący 1 mln zł spowodował zatrudnienie 60% z nich przez 3 lata nie jest wystarczająca do określenia, jakiej jest on jakości i jaką ma wartość dla społeczeństwa. Wybór kryteriów ewaluacyjnych, określenie ich miar (wymiarów) oraz progów na tych miarach wyznaczających poziom jakości, określenie ważności (wag) dla poszczególnych wymiarów, ocena programu na poszczególnych wymiarach (oceny cząstkowe) i synteza ocen cząstkowych to specyficzne dla ewaluacji działania, które powinny być przeprowadzane w przemyślany sposób, a więc zgodnie z uzasadnionymi wskazówkami i procedurami. To są właśnie metody ewaluacji.

Prześledźmy tę problematykę najpierw na przykładzie ewaluacji osiągnięć szkolnych ucznia. Jakie powinny być kryteria ewaluacyjne w tym przypadku? Przyjmijmy, że ostateczna ocena powinna zależeć od uczestnictwa w zajęciach oraz wiedzy i umiejętności, które należą do programu przedmiotu. Dla każdego z tych kryteriów powinniśmy teraz dobrać wymiar, np. obecność i aktywność dla uczestnictwa, wynik testu wiadomości dla wiedzy, a wynik samodzielnej pracy i zadania praktycznego dla umiejętności. Dla każdego z wymiarów trzeba określić skalę pomiaru i progi na tych skalach, np. skalą obecności może być odsetek liczby obecności do wszystkich zajęć, a progami na tej skali 50% - minimum, 50%-64% -

dostatecznie, 65%-75% - dobrze, powyżej 75% - bardzo dobrze; skalą aktywności może być liczba zgłoszeń do odpowiedzi, a progami na tej skali do 5 zgłoszeń mało aktywny, między 5 a 10 – średnio aktywny, powyżej 10 - bardzo aktywny; skalą testu może być odsetek dobrych odpowiedzi do liczby wszystkich pytań, a progami na tej skali 50% - minimum, powyżej 50% - dostatecznie, powyżej 60% - dostatecznie plus itd.; skala dla pracy i zdania praktycznego może być skalą ocen szkolnych od celującej do niedostatecznej (pomijam problem oceny samej pracy i wykonania zadania praktycznego, tzn. ustalenia miary i progów w tych przypadkach). Należy zwrócić uwagę na to, że próg minimum, np. obecności poniżej 50% lub wynik testu poniżej 50%, jest dość szczególny i zwykle ma on istotne dla ucznia konsekwencje.

Pozostało jeszcze określenie ważności poszczególnych kryteriów, a także ważności wymiarów, jeżeli jest ich więcej niż jedno dla danego kryterium. Skala ważności może być następująca: bardzo ważne, średnio ważne i mało ważne. Załóżmy teraz, że uczestnictwo jest mało ważne, wiedza - średnio ważna a umiejętności - bardzo ważne. W ramach uczestnictwa obecność jest mało ważna a aktywność bardzo ważna; w ramach umiejętności – praca jest średnio ważna a zadanie praktyczne bardzo ważne. Załóżmy, że było tylko dwóch uczniów i wiemy jakie są ich wyniki na poszczególnych wymiarach, a więc możemy przypisać im częściowe oceny, które wyglądają tak, jak w poniższej tabeli.

Kryteria	Ważność kryteriów	Wymiar	Ważność wymiarów	uczeń A	uczeń B
Uczestnictwo	mało ważne	obecność	mało ważna	bardzo dobrze	poniżej minimum
		aktywność	bardzo ważna	bardzo aktywny	średnio aktywny
Wiedza	średnio ważna	test wiadomości		celujący	dobry
Umiejętności	bardzo ważne	praca pisemna	średnio ważna	dobra	bardzo dobra
		zadanie praktyczne	bardzo ważna	dostateczny plus	dobry

Pozostaje nam zadanie dokonania syntezy ocen częściowych z uwzględnieniem ważności kryteriów i wymiarów. Jeżeli przyjęto by, że brak osiągnięcia minimum na którymkolwiek wymiarze oznacza, że osiągnięcia ucznia mają być ocenione najniżej, to uczeń B dostałby taką łączną ocenę mimo, że radził sobie w innych wymiarach dobrze. Widzimy jednak, że kryterium uczestnictwa jest mało ważne, a wśród jego wymiarów mało ważna jest obecność. W takich przypadkach nie jest uzasadnione automatyczne przypisanie uczniowi najgorszej oceny łącznej. Można przyjąć, że tylko niespełnienie minimów w wymiarach uznanych za bardzo ważne powinno mieć takie konsekwencje (tzn. pozostałe oceny częściowe nie mają wówczas znaczenia dla oceny łącznej). Jeżeli wyniki na uznanych za ważne wymiarach przekroczyły minimum, wtedy ocena łączna musi uwzględnić wszystkie oceny częściowe.

Jeżeli wszystkie oceny częściowe są najwyższe, to łatwo stwierdzić, że ocena łączna jest najwyższa, problemy zaczynają się wówczas, gdy oceny częściowe są mniej lub bardziej rozbieżne. Uczeń A na jednym z bardzo ważnych wymiarów miał niską ocenę, a więc powinno mieć to istotny wpływ na jego ocenę łączną. Problemy tego rodzaju rozwiązuje się poprzez agregowanie ważonych ocen częściowych sprowadzonych do wymiaru liczbowego, np. według skali ocen szkolnych wyniki ucznia A za wiedzę i umiejętności to 6, 4 i 3,5, skale obecności i aktywności są trzystopniowe, przypiszmy 6 do oceny wyższej, 4 do średniej i 2 do niższej, stąd na pozostałych wymiarach uczeń A ma 6 i 6. Ze względu na to, że wagi występują też w obrębie danego kryterium najpierw trzeba mieć częściowe wyniki łączne dla poszczególnych kryteriów. Skala wag też jest trzystopniowa, złożmy, że waga 1 odpowiada ocenie bardzo ważne, waga 0,8 – średnio ważne, a waga 0,6 – mało ważne. Najpierw agregujemy wyniki dla kryterium uczestnictwa $0,6 \cdot 6 + 1 \cdot 6 = 9,6$ i umiejętności $0,8 \cdot 4 +$

$1 \cdot 3,5 = 6,7$, gdyż składa się na nie po kilka wymiarów z ocenami¹⁰. Pozostaje wyliczenie wyniku łącznego: $0,6 \cdot 9,6 + 0,8 \cdot 6 + 1 \cdot 6,7 = 5,76 + 4,8 + 6,7 = 17,26$ (dla ucznia B wynik łączny wyniesie $0,6 \cdot 4 + 0,8 \cdot 4 + 1 \cdot 8 = 13,6$, a maksymalny do uzyskania przy tych założeniach $5,76 + 6 + 10,8 = 22,56$) oraz wyznaczenie progów na skali wyniku łącznego, aby znowu dojść do wniosków ewaluacyjnych, tzn. odpowiedzieć na pytanie, jak dobry jest wynik 17,26 lub wynik 13,6. Porównując wyniki obu uczniów widzimy, że nieco lepszy poziom osiągnąć ma uczeń A.

W omówionym przykładzie wystąpiło wiele istotnych kwestii metodologicznych dla dociekań ewaluacyjnych niezależnie od tego, czy ich przedmiotem są osiągnięcia szkolne uczniów, czy programy społeczne i ich osiągnięcia. Oto tylko niektóre z pytań, na jakie musi odpowiedzieć metodologia ewaluacji: jak wybrać właściwe dla danego przedmiotu ewaluacji kryteria, jak dobrać wymiary (miary) właściwe dla każdego kryterium oraz jakie powinny mieć one skale, jak przypisać progi ocen na skalach poszczególnych wymiarów, jak ustalić ważność dla poszczególnych kryteriów i dla wymiarów w ramach kryteriów, jak agregować oceny cząstkowe z uwzględnieniem ich ważności?¹¹

Rekomendowanie programów społecznych¹²

Przykład z poprzedniego punktu może nam posłużyć również do przedstawienia problematyki analizy programów społecznych, czyli dociekań, których celem jest sformułowanie rekomendacji. Wyobraźmy sobie, że zamiast uczniów mamy dwa programy społeczne A i B, pierwszy z nich jest obecnie funkcjonującym (status quo), a drugi został zaproponowany jako alternatywa dla niego, a jego autorzy uważają, że będzie on lepszy. Wyobraźmy sobie, że stosujemy kryteria ewaluacyjne odpowiednie dla programów społecznych (skuteczność, efektywność, legalność, akceptowalność itd.) a ich wymiary są dostosowane do charakterystyk tych dwóch programów. Po przeprowadzeniu analizy, otrzymujemy łączne wyniki punktowe dla programu A – 17,26 i dla programu B – 13,6.

Nie rozstrzygając już problemu jak dobre są oba te wyniki (czyli bez przechodzenia do wniosków ewaluacyjnych) wyraźnie widzimy, że program A ma wyższy wynik łączny od programu B, co powinno też mieć odzwierciedlenie w różnicach w ocenie łącznej, przynajmniej takie, że ocena łączna A powinna być lepsza od oceny łącznej B (mogą się jednak mieścić w tym samym przedziale ocen). Jeżeli tak, to należy sformułować rekomendację o następującej treści: z wyników badań, ewaluacji i analizy programów społecznych A i B można wyciągnąć wniosek, że nie należy podejmować reformy, która będzie miała na celu zastąpienie obecnie działającego programu A zaproponowanym programem B. Dodatkowo można wskazać, które z cech programu B odpowiadają za jego gorszą ocenę, co może być wykorzystane do poszukiwania rozwiązań, które nie miałyby tych cech i byłyby jednak lepsze od A.

Właśnie w ten sposób przebiega wielokryterialna analiza programów społecznych. Schematycznie przyjmijmy, że mamy 3 kryteria ewaluacyjne (K1...K3), każdemu odpowiada dla uproszczenia tylko jeden wymiar z własną skalą i progami ocen (D1...D3, a więc odpada osobna kwestia ważności wymiarów), mamy określoną ważność dla kryteriów (W1...W3), a przedmiotem analizy są trzy alternatywne programy społeczne A, B i C, które mają realizować podobne cele. Odpowiednia tabela ewaluacyjna z ocenami cząstkowymi (OC) i łącznymi (OŁ) alternatyw oraz rekomendacją („*”, przy założeniu, że ocena łączna

¹⁰ Problem, który od razu widać przy takim sposobie agregowania ocen cząstkowych jest następujący, liczba wymiarów dla poszczególnych kryteriów ma wpływ na wynik, dlatego w tym przypadku lepszym sposobem agregacji byłaby średnia.

¹¹ Na wiele z tych pytań daje odpowiedź E.J. Davidson „Evaluation Methodology Basics”, SAGE Publications, Thousands Oaks, London, New Delhi 2004.

¹² W świecie anglojęzycznym jest to problematyka nazywana *policy analysis*, spośród wielu podręczników można polecić D.L. Weimer A.R. Vining „Policy Analysis. Concepts and Practice”, Prentice Hall, Upper Saddle River 2004 (kilka wydań).

alternatywy B jest lepsza od pozostałych) będzie wyglądała następująco (podobnie jak wyżej pominąłem informacje, które są podstawą do wydania ocen cząstkowych, tzn. osiągnięcia alternatyw na poszczególnych skalach poszczególnych wymiarów).

Kryteria	Wymiary	Wagi	Oceny alternatywnych programów społecznych		
			A	B	C
K1	D1	W1	OC1A	OC1B	OC1C
K2	D2	W2	OC2A	OC2B	OC2C
K3	D3	W3	OC3A	OC3B	OC3C
Ocena łączna			OŁA	OŁB	OŁC
Rekomendacja				*	

Względna prostota tej tabeli (nazywanej w kontekście analizy polityki macierzą cele/alternatywy) nie powinna nas zmylić. Jak widzieliśmy wyżej jej skonstruowanie dla rzeczywistych programów społecznych wymaga badań, ewaluacji i analizy, czyli również zasobów i działań, aby dać produkt w postaci informacji, ocen i rekomendacji.

Badania dostarczyć muszą nam informacji o poszczególnych alternatywach, tak abyśmy mogli je przedstawić w świetle kryteriów na poszczególnych ich wymiarach. Jeżeli wśród alternatyw są też takie, których jeszcze nie zrealizowano, informacje o nich są dodatkowo niepewne, gdyż dotyczą tego, co się zdarzy w przyszłości, gdy zastąpią one status quo. Stanowi to dodatkowy problem i potencjalne źródło błędów, które powinny być kontrolowane. Wszystkie problemy, jakie mamy z klasycznymi metodami dociekań badawczych, w tym szczególnie z metodami dociekań przyczynowych, niezależnie od tego, do czego wykorzystujemy ich wyniki, mają też znaczenie dla ewaluacji i analizy programów społecznych.

Dodatkowe źródło trudności kryje się za ustalaniem ważności kryteriów ewaluacyjnych (i ewentualnie wymiarów, gdy przypada ich więcej na jedno kryterium). Administratorzy programu, jego personel, klienci, sponsorzy, politycy i opinia publiczna mogą odmiennie hierarchizować ważność. W skrajności może dojść do sprzeczności w tym względzie, tj. odwrotności hierarchii ważności – co dla jednej grupy jest najważniejsze, dla innej jest najmniej ważne itd. Ponieważ wagi wpływają na oceny cząstkowe, a te z kolei na oceny łączne i rekomendacje, to odwrotne ich uporządkowanie może dać sprzeczne rekomendacje. Według hierarchii ważności grupy 1 rekomendować powinniśmy program B, ale według hierarchii ważności grupy 2 zalecić będzie trzeba program A. Szczególnie jest to prawdopodobne, gdy wyniki programów na poszczególnych wymiarach nie różnią się znacząco. Obiektywizacja tej kwestii wymagałaby osobnej procedury agregowania różnych uszeregowania ważności kryteriów i wymiarów. W praktyce przypadki skrajne jednak rzadko występują, stąd problem sprzeczności hierarchii ważności może mieć mniejsze znaczenie praktyczne.

Problemy nie kończą się wraz z uzyskaniem ocen cząstkowych, gdyż potrzebna nam jeszcze ich synteza z uwzględnieniem hierarchii ważności i ewentualnie innych czynników, które związane są z niepewnością informacji o analizowanych programach. Wyżej widzieliśmy, że wymaga to dodatkowych decyzji i rozwiązań, które same w sobie mogą mieć wpływ na ocenę łączną. Głównym problemem jest sprowadzenie ocen częściowych do wspólnego mianownika oraz wybór kształtu funkcji ważenia i funkcji agregowania (mnożenie, dodawanie, średnia itd.).

Wpływ produktu badań, ewaluacji i analizy w postaci informacji, ocen i rekomendacji na proces polityczny, w trakcie którego zapadają decyzje o losach różnych programów społecznych może być niewielki, gdyż w demokratycznej polityce uzasadnione argumenty merytoryczne mogą być słabsze niż argumenty o charakterze politycznym związane z cyklem wyborczym i ciągłą grą o władzę między zmiennymi koalicjami parlamentarnymi oraz grą

grup różnych interesów o wpływ na politykę w celu zwiększenia swoich korzyści i/lub zmniejszenia swoich strat. Jeżeli potraktujemy systematyczne dociekania nad programami społecznymi jako jeszcze jeden przedmiot badań, ewaluacji i analiz może się okazać, że ich wpływ jest niewielki w stosunku do przeznaczanych na nie zasobów. Gdyby tak rzeczywiście było, a oceny w świetle innych kryteriów ewaluacyjnych byłyby również niedobre, może to być sytuacja niepokojąca. Problem ten może być częściowo rozwiązany poprzez wprowadzenie obowiązku przeprowadzania ewaluacji i analizy programów społecznych oraz uwzględniania ich wyników przy decyzjach o kształcie programów społecznych.

Rozwój i perspektywy ewaluacji i analizy programów społecznych

Ewaluacja i analiza w sensie pokazanym wyżej rozwija się już od co najmniej kilku dziesięcioleci w USA i innych krajach. Stała się już nie tylko działalnością naukową, ale profesjonalną. W USA jednym z głównych bodźców do tego rozwoju był wzrost zaangażowania rządu federalnego w rozwiązywanie problemów i kwestii społecznych, któremu towarzyszyło szybkie zwiększanie wydatków na programy społeczne. Zdecentralizowane państwo i kultura pragmatyzmu oraz wzrost wydatków na cele społeczne, który zaczął być postrzegany jako problem sam w sobie, a także rosnące wątpliwości dotyczące skuteczności programów społecznych (wyrażane przez neokonserwatystów amerykańskich takich jak m.in. Ch. Murray i G. Gilder) sprzyjały rozwojowi dociekań nastawionych na dostarczanie informacji o nich, ocen ich działania oraz rekomendacji, które miały je usprawnić lub pokazać alternatywne sposoby rozwiązywania problemów społecznych. Dodatkowym bodźcem było wprowadzenie prawnego obowiązku przeprowadzania „analiz wpływu regulacji” (*regulatory impact analysis*, RIA) dla inicjatyw legislacyjnych o kosztach powyżej 100 mln \$¹³.

Wśród wymogów narzuconych w ten sposób agencjom rządowym były m.in. takie: „działania regulacyjne nie powinny być podejmowane o ile potencjalne korzyści płynące dla społeczeństwa z regulacji nie przeważają nad ich potencjalnymi kosztami”, „cele regulacyjne powinny być wybierane tak, aby maksymalizować korzyści netto dla społeczeństwa”, „spośród alternatywnych podejść do realizacji każdego z celów regulacyjnych wybrane powinny być te, które nakładają najmniejsze koszty netto dla społeczeństwa”.¹⁴ Są to podstawowe zasady jednej z metod dociekań nastawionych na rekomendację, a mianowicie analizy koszty-korzyści¹⁵, w której głównym kryterium ewaluacyjnym jest efektywność, a jej wymiary przedstawia się w jednostkach pieniężnych. Pokazana wyżej metoda analizy wielokryterialnej jest jej rozwinięciem i uzupełnieniem o inne kryteria.

W propagowanie szeroko rozumianego RIA zaangażowała się również Organizacja Współpracy Gospodarczej i Rozwoju (OECD). W 1995 Rada OECD przyjęła zalecenie w sprawie poprawy jakości regulacji rządowych, w którym m.in. zaprasza rządy państw członkowskich do składania co trzy lata raportów w tym zakresie.¹⁶

Unia Europejska w połowie lat 1990. zamówiła badania na temat praktyk ewaluacyjnych swoich własnych organów, a głównie Komisji Europejskiej i jej Dyrekcji Generalnych wyspecjalizowanych w prowadzeniu poszczególnych działów polityki

¹³ Rozporządzenie wykonawcze 12291 podpisane przez R. Reagana w 1981 r. W. Clinton podpisał w 1993 rozporządzenie o planowaniu i przeglądzie regulacji nr 12866 rozwijające zasady RIA. Treść obu aktów dostępna jest na stronach The National Archives www.archives.gov.

¹⁴ Rozporządzenie 12291.

¹⁵ Więcej o tego rodzaju analizie można się dowiedzieć z dokumentu opracowanego w kontekście polityki regionalnej UE „Analiza kosztów i korzyści projektów inwestycyjnych: Przewodnik”, Dyrekcja Generalna – Rozwój Regionalny, Komisja Europejska. Dokument dostępny na stronach Portalu Funduszy Strukturalnych www.fundusze-strukturalne.gov.pl.

¹⁶ Szczegóły tej inicjatywy i towarzyszące jej dokumenty oraz sprawozdania państw członkowskich można znaleźć na stronach OECD www.oecd.org.

wspólnotowej. Efektem tego był komunikat skierowany do Komisji, który miał poprawić sytuację w tej dziedzinie. Rozpoczyna go następujące zdanie: „rosnące rozmiary budżetu i troska podatników UE o to, aby za swój wkład do niego otrzymali odpowiednie usługi (*value*), a także rozwój w Państwach Członkowskich i gdzie indziej w kierunku zarządzania sektorem publicznym nastawionego na wyniki, stworzyły potrzebę przeglądu i unowocześnienia podejścia Komisji do zarządzania budżetem”.¹⁷

Jedno z zaleceń sformułowane w tym kontekście zobowiązywało wszystkie dyrekcje generalne do wyodrębnienia własnych funkcji ewaluacyjnych. Jak widać, głównym impulsem do rozwijania ewaluacji na poziomie UE były względy finansowe. Wspomniany komunikat powstał w kontekście inicjatywy „zdrowe i efektywne zarządzanie finansowe” zwanej w skrócie SEM 2000. Jej wyniki zostały uwzględnione w głównym akcie regulującym kwestie ogólnego budżetu UE z 2002 i tam też mowa o ewaluacji w art. 27 pkt. 4: „W celu poprawy procesu podejmowania decyzji instytucje powinny przeprowadzać ewaluacje *ex ante* i *ex post* zgodnie ze wskazówkami Komisji. Powinny być one stosowane do wszystkich programów i działań, które wiążą się ze znacznym wydatkowaniem środków, a wyniki ewaluacji powinny być rozpowszechniane...”.¹⁸

Ewaluacja programów i projektów społecznych w Polsce zaczęła się dynamicznie rozwijać w związku z włączeniem Polski do UE. Społeczny wymiar integracji był obecny w największym stopniu w Sektorowym Programie Operacyjnym Rozwój Zasobów Ludzkich 2004-2006 oraz w Programie Inicjatywy Wspólnotowej EQUAL, które finansowane są z Europejskiego Funduszu Społecznego (EFS). W dokumencie będącym podstawą działania pierwszego programu zapisano m.in.: „wzrost potencjału ewaluacyjnego polskiej administracji jest jednym z głównych celów w okresie programowania 2004-2006”.¹⁹ W instrukcji dla projektodawców, którzy chcieli korzystać ze środków EFS znalazł się natomiast taki punkt: „należy opisać, czy i jak będą badane rezultaty projektu, w szczególności efekty długoterminowe; w jaki sposób i kiedy zostanie przeprowadzona ewaluacja projektu po zakończeniu wsparcia, jeśli tak, to należy opisać sposób oraz termin przeprowadzenia oceny”.²⁰

W PIW EQUAL zadeklarowano, że Polska zapewni, aby „wszystkie Partnerstwa na rzecz Rozwoju [organizacje realizujące projekty] opracowały metodologię i zastosowały odpowiednie mechanizmy dla oceny bieżącej swoich działań i osiągnięć, łącznie z opisem możliwych do weryfikacji wskaźników, które ilustrują, w jaki sposób cele i rezultaty będą mierzone i oceniane, a także odpowiednią metodologię i mechanizmy monitoringu i oceny wspólnych działań w ramach współpracy ponadnarodowej”.²¹ W innym dokumencie interpretuje się ten zapis jako obowiązek prowadzenia przez partnerstwa autoewaluacji. W PIW EQUAL znajdujemy jeszcze jedno istotne zdanie o wykorzystaniu okresu 2004-2006 na „stworzenie i rozwinięcie systemów ewaluacji programów oraz możliwość zastosowania ich w kolejnym okresie programowania”.²² W związku z tym zarówno na poziomie administracji centralnej, jak i w organizacjach z różnych sektorów powstał bardzo duży popyt na usługi ewaluacyjne. Szybko zaczęły powstawać podyplomowe studia ewaluacyjne, kursy i szkolenia w zakresie ewaluacji, coraz więcej instytucji i firm zaczyna wchodzić na stworzony w ten sposób rynek, co dobrze wróży rozwojowi całej branży oraz profesjonalizacji tej działalności.

¹⁷ „Concrete steps towards best practice across the Commission”, maj 1996, Komisja Europejska, SEC 96/659 final.

¹⁸ Regulacja Rady (EC, Euratom) nr 1605/2002 z 25 czerwca 2002 na temat Regulacji Finansowej mającej zastosowanie do ogólnego budżetu Wspólnot Europejskich.

¹⁹ Załącznik do rozporządzenia Ministra Gospodarki i Pracy z dnia 1 lipca 2004 r. (Dz. U. 166 poz. 1743), s.223. Dostępny również na stronach Portalu Funduszy Strukturalnych www.fundusze-strukturalne.gov.pl.

²⁰ „Europejski Fundusz Społeczny w Polsce - Poradnik dla projektodawców”, Ministerstwo Gospodarki, Pracy i Polityki Społecznej, luty 2004, s.76.

²¹ „Program Inicjatywy Wspólnotowej EQUAL dla Polski”, Ministerstwo Gospodarki, Pracy i Polityki Społecznej, kwiecień 2004, s.181. Dokument dostępny, m.in. na stronach Portalu Funduszy Strukturalnych www.fundusze-strukturalne.gov.pl.

²² Tamże.

Przejdźmy teraz do początków RIA w Polsce. Obowiązują już przepisy, które wprowadzają obowiązek „oceny skutków regulacji” (OSR). Pojawiły się one w uchwale Rady Ministrów o regulaminie jej pracy z 1997 r. w wyniku poprawki z 2001 r. W obecnie obowiązującej wersji tej uchwały z 2002 r. mowa o tym, że każdy projekt aktu normatywnego powinien zawierać w uzasadnieniu odrębną część z OSR, czyli oceną „przewidywanych skutków (kosztów i korzyści) społeczno-gospodarczych regulacji”. Ma ona zawierać „wskazanie podmiotów, na które oddziałuje akt normatywny”, „wyniki przeprowadzonych konsultacji” jeżeli są one wymagane, a także „przedstawienie wyników analizy wpływu aktu normatywnego” na: sektor finansów publicznych, w tym budżet państwa i budżety jednostek samorządu terytorialnego; rynek pracy; konkurencyjność gospodarki i przedsiębiorczość, w tym na funkcjonowanie przedsiębiorstw; sytuację i rozwój regionalny.²³

Ministerstwo Gospodarki opublikowało w lipcu 2006 r. dokument pt. „Program reformy regulacji”, gdzie zapowiedziano opracowanie „Wytycznych do oceny skutków regulacji”. Ma to być odpowiedź na problemy z dotychczasowym systemem, który „Nie funkcjonował efektywnie ze względu na mało użyteczne procedury, które powodowały, że system OSR nie miał istotnego wpływu na proces podejmowania decyzji”²⁴. Inicjatywa ta nawiązuje do komunikatu Komisji Europejskiej z 2005 „Lepsza regulacja dla wzrostu i miejsc pracy w UE”. Oznacza to, że OSR czyli RIA w polskim wydaniu ma również dobre perspektywy na rozwój i profesjonalizację. Dotyczyć to będzie również regulacji w zakresie polityki społecznej.

²³ Uchwała nr 49 z dnia 19 marca 2002 r. Regulamin pracy Rady Ministrów (M. P. nr 13, poz. 221), par. 9 i 10.

²⁴ Informacja prasowa pt. „Przedsiębiorca ma prawo - Program Reformy Regulacji” z 28.08.2006, Ministerstwo Gospodarki. Dokument dostępny jest na stronach www.mg.gov.pl.