

Jarosław Górniak, Karolina Keler

Ewaluacja i wskaźniki w cyklu planowania i realizacji polityk publicznych¹

Celem artykułu jest przedstawienie funkcji ewaluacji i wskaźników w cyklu polityk publicznych oraz omówienie problemów związanych z realizacją tych funkcji. W szczególności uwzględniono fazę oceny i planowania polityk oraz omówiono funkcje pomiaru efektów programów przy pomocy wskaźników jako kluczowej funkcji ewaluacji. W pierwszej części artykułu opisano szczegółowo funkcje diagnozy, wskaźników i ewaluacji w cyklu polityk publicznych. W drugiej części zaprezentowano ewaluację i wykorzystanie wskaźników na etapie oceny zakończonych projektów. W tej części rozważań omówiono również koncepcję ewaluacji opartą o teorię programu oraz znaczenie pomiaru efektu netto dla zidentyfikowania przyczynowego oddziaływania interwencji. Autorzy stoją na stanowisku, że należy traktować jako rezultaty wszystkie zamierzone konsekwencje interwencji (programu, polityki) dla bezpośrednich grup (układów) docelowych, niezależnie od tego, czy występują natychmiast, czy po pewnym czasie od interwencji; jako oddziaływania należy traktować zamierzone i niezamierzone skutki interwencji dla szerszego układu, w którym wystąpił problem wymagający interwencji (Komisja Europejska traktuje odłożone w czasie rezultaty jako oddziaływanie). Wskazano trudności wiążące się z pomiarem oddziaływania interwencji, a jednocześnie podkreślono, że uzyskanie wymiernego rezultatu jest warunkiem koniecznym oczekiwania, że interwencja miała swoje oddziaływanie. W części trzeciej poddano dyskusji rolę wskaźników i ewaluacji w planowaniu polityk publicznych. Szczególną uwagę zwrócono na właściwe definiowanie celów w procesie programowania polityk oraz wagę ustalenia dobrych wskaźników. Położono nacisk na różnice w sposobie ujęcia celów jako ogólnych zamiarów i wizji, precyzyjnie ujętych zamierzonych efektów działań i zaplanowanych wartości przyjętych wskaźników tych efektów. Przedstawiono problemy z ustalaniem wartości wskaźników oraz związków występujących pomiędzy pomiarem produktów, rezultatów i oddziaływania interwencji. Autorzy podkreślają w artykule, że ewaluacja i wskaźniki mogą być użytecznym narzędziem prowadzenia polityk publicznych, ale zastosowanie ich w praktyce wymaga poradzenia sobie z omówionymi w tekście problemami.

Słowa kluczowe: ewaluacja, wskaźniki, cykl polityk publicznych.

1. Wstęp

Wraz z przystąpieniem Polski do Unii Europejskiej pojawiły się – co można uznać za wartość samą w sobie – nowe wymogi dotyczące projektowania, implementacji i ewaluacji polityk, programów czy projektów publicznych².

Instytut Socjologii UJ.

¹ Referat prezentowany na Uniwersytecie Ekonomicznym w Krakowie w ramach programu naukowego „Dobre rządzenie”.

² Dla uproszczenia wyводу będziemy odnosić się do polityk publicznych, choć, o ile nie zostanie to specjalnie

Komisja Europejska sformułowała wiele zaleceń metodologicznych opisujących standardy programowania i ewaluacji³. Sprostanie tym wymogom wymusza wprowadzenie właściwych metod projektowania opartych na poprawnym diagnozowaniu sytuacji, jasno sformułowanych

podkreślone, wszystko, co piszemy, odnosi się do wszystkich rodzajów interwencji rozumianych jako celowe działania władz publicznych, aby wywołać pożądaną zmianę społeczną.

³ Zob. np. EC 2006a, EC 2006b, EC 2006c.

celach działania i systemie oceny prowadzonych działań.

Ewaluacja jest ważnym etapem „cyklu życia” polityk publicznych. Nie należy jej traktować jako narzuconego biurokratycznie, zbędnego i kosztownego elementu. Poprawnie prowadzona, pozwala przede wszystkim ograniczyć alokację środków na działania nietrafne, nieskuteczne, nieefektywne i nieprzynoszące korzyści społecznych. W przypadku decyzji regulacyjnych i programów finansowanych ze środków publicznych ich ewaluacja jest istotnym warunkiem możliwości rozliczenia władz publicznych przez wyborców. Z menedżerskiego punktu widzenia może przynieść wiele korzyści na etapie planowania nowych działań oraz rewizji i usprawniania procesu wdrożenia działań już realizowanych, jeżeli tylko wplecie się ją poprawnie w cykl realizowanych programów. Niezbywalnym narzędziem ewaluacji programów publicznych i w ogóle procesu programowania oraz zarządzania wdrożeniem są wskaźniki, za pomocą których dokonujemy diagnozy, ustanawiamy wymierne cele, monitorujemy proces wdrożenia i dokonujemy etapowej i całościowej oceny efektów programu.

W niniejszym artykule stawiamy sobie dwa powiązane ze sobą cele. Jednym z nich jest omówienie miejsca ewaluacji w cyklu polityk publicznych oraz korzyści, jakie przynosi na każdym z etapów wspomnianego cyklu. Drugi to przedstawienie funkcji wskaźników w planowaniu i ocenie prowadzonych działań oraz czynników, które wpływają na spełnianie tych funkcji⁴.

Poniższe rozważania można odnieść zarówno do ewaluacji polityk publicznych, jak i ewaluacji programów realizowanych w ramach tych polityk. Polityki i programy publiczne kojarzą się zwykle działaniami alokacyjnymi. Mówiąc o ewaluacji programów publicznych, koncentrujemy zwykle uwagę na ocenie skuteczności i efektywności wydatkowania środków publicz-

nych. Niestety, zazwyczaj nie poddaje się podobnej ewaluacji działań regulacyjnych – tworzenia ustaw, rozporządzeń i innych aktów prawnych – choć ich skutki są na ogół wymierne, zarówno w kategoriach finansowych, jak i niefinansowych. Wprowadzanie regulacji prawnych należy traktować jako przedsięwzięcie, które także powinno mieć jasno zdefiniowane cele i system pomiaru efektów. W istocie rzeczy interwencje publiczne złożone są często nie tylko z działań regulacyjnych, lecz także z działań alokacyjnych. Szersze omówienie tego zagadnienia wykracza poza ramy niniejszego artykułu, ale większość poruszonych w tym artykule kwestii można odnieść do sfery oceny działań regulacyjnych i alokacyjnych.

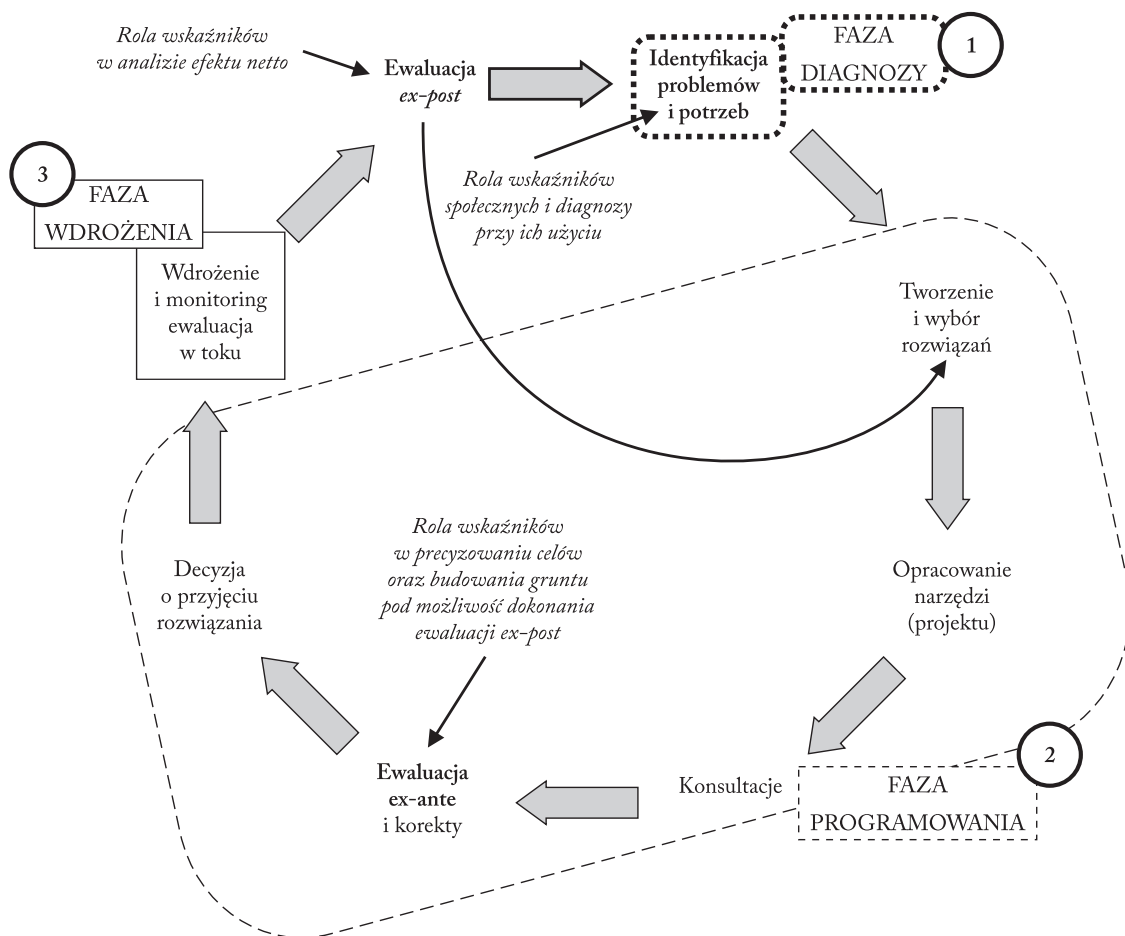
2. Miejsce ewaluacji i wskaźników w cyklu planowania i realizacji polityk publicznych

W cyklu planowania i realizacji polityk publicznych daje się wyróżnić kilka sytuacji, w przypadku których ewaluacja oraz użycie wskaźników mogą pełnić użyteczną rolę. Pisząc o cyklu, mamy na myśli łańcuch działań zaprezentowany na ryc. 1, składający się z trzech zasadniczych elementów: (1) diagnozy, (2) programowania oraz (3) wdrożenia. Opieramy się na popularnym schemacie cyklu polityk publicznych (por. np. Bridgman, Davis 2004), który na ogół zamyka fazę wdrożenia ewaluacją *ex-post*. Uzupełniliśmy schemat o dodatkowe elementy. Poniżej omawiamy pokrótce najważniejsze pojęcia jako wstęp do prezentacji kluczowych tez artykułu.

2.1. Diagnoza

Punktem wyjścia wszelkiego rodzaju planów działań jest diagnoza. Cele polityk publicznych nie wynikają jednak automatycznie z diagnoz prowadzonych przez surowych, obiektywnych analityków. Cele są wynikiem wizji politycznych, ideologii rozwojowych, gry interesów i konkurencji politycznej. Naszym zdaniem, ekspercka diagnoza i analiza polityk publicznych powinna sprawdzać prawidłowość diagnoz i definicji sytuacji przyjmowanych przez polityków, usta-

⁴ Punktem wyjścia tego tekstu była nasza prezentacja wstępnej wersji artykułu na seminarium „Dobre rządzenie” prowadzonym przez Katedrę Gospodarki i Administracji krakowskiego Uniwersytetu Ekonomicznego pod kierownictwem prof. Jerzego Hausnera. Dziękujemy wszystkim uczestnikom tego seminarium, a w szczególności Jerzemu Hausnerowi, Łukaszowi Mamicy i Marcinowi Zawickiemu za istotne uwagi, które przedstawił podczas dyskusji nad koncepcją artykułu.



Ryc. 1. Miejsce ewaluacji i wskaźników w cyklu planowania i realizacji polityk publicznych

Źródło: opracowanie własne w nawiązaniu do Bridgman, Davis (2004).

łać zależności przyczynowe, bilansować zasoby niezbędne do działania, rysować scenariusze wynikające z proponowanych działań. Stanowi ona jednak zawsze tylko wsparcie dla inicjatywy polityków. Ważną funkcją diagnozy jest także analiza celów formułowanych w programie pod kątem ich adekwatności wobec potrzeb i problemów społecznych oraz ich osiągalności w danych warunkach. Dobra diagnoza dostarcza także wiedzy o kontekście programu, który może mieć wpływ na ostateczne efekty, zamierzone i niezamierzone. Gdy konfrontujemy taki punkt widzenia na zadania diagnozy z tym, jak są one zwykle zapisywane w formułowanych rozmaitych dokumentach programowych, zauważamy, że ich autorzy zasypują decydentów danymi sta-

tystycznym, nieodnoszącymi się do ostatecznie przyjmowanych, *explicite* bądź *implicite*, celów działania, a jedynie lepiej lub gorzej opisują obszar, w którym cele te są ustanawiane.

Jedną z dróg ulepszenia diagnoz sporządzanych na potrzeby polityk publicznych jest opracowanie i przyjęcie zestawu wskaźników dla poszczególnych dziedzin (częściowo będą w Polsce przyjęte rozwiązania wypracowane w Unii Europejskiej, ale nie zaspokoją one wszystkich potrzeb), po to, by dysponować możliwością spójnego, trafnego i rzetelnego opisu oraz analizy rzeczywistości społeczno-ekonomicznej w kategoriach empirycznych. Trzeba także zwiększyć zaangażowanie statystyki publicznej w gromadzenie danych potrzebnych

do regularnego wyliczania wartości kluczowych wskaźników, aktualnych w takim stopniu, by miały praktyczne znaczenie dla podejmowania decyzji oraz usprawniały system gromadzenia i przetwarzania danych w ramach poszczególnych instytucji. Alternatywą, ale nie zawsze dostępną i często kosztowną, jest zamówienie specjalnych badań empirycznych.

Po to, by móc ocenić program po jego wdrożeniu, musimy w diagnozie ustalić, jaki był punkt wyjścia, w stosunku do którego miała nastąpić zmiana. Diagnoza powinna posługiwać się wyraźnymi ocenami opisywanych zjawisk: oczekujemy w niej określenia, czego jest za dużo, czego jest za mało, a czego w sam raz. W diagnozie należy także ujmować potencjalne konsekwencje podjęcia i niepodjęcia określonych działań. Opis rzeczywistości dokonywany jest w diagnozie także za pomocą wskaźników, pozwalających scharakteryzować analizowane zjawiska. Oprócz kwestii bezpośrednio związanych z celami interwencji, diagnoza powinna obejmować także omówienie szerszego kontekstu problemowego, trendów (np. demograficznych czy ekonomicznych) mogących wpłynąć na realizację programu i na jego ostateczny efekt. Na takim szerszym tle można jasno wskazać, które problemy czy wyzwania społeczne zostają wybrane jako przedmiot interwencji

Diagnozę należy tworzyć w toku procesu iteracyjnego: opierając się na debacie wokół podstawowych problemów społecznych i wstępnej wizji strategicznych celów określonej polityki, przystępujemy do konstrukcji ram diagnozy (w tym wskaźników), przeprowadzamy diagnozę, korygujemy definicję sytuacji i koncepcje celów, wracamy do diagnozy w celu dokonania stosownych uzupełnień itd. Ostatecznie interwencje publiczne są pewnymi wizjami dotyczącymi wagi określonych problemów społecznych i koncepcjami ich skutecznego rozwiązania. Ich podmiotowy, a nierzadko kreatywny charakter powoduje, że, z jednej strony, ich kształt nie wynika nigdy samoistnie z diagnozy, z drugiej zaś – w związku z niepewnością – wymagają, przynajmniej w warunkach demokracji, przyjęcia odpowiedzialności stojącej wobec zaufania społecznego. Diagnoza powinna służyć temu, by odpowiedzialność i zaufanie nie były konfrontowane z nadmiernym poziomem niepewności.

2.2. Wskaźniki

Ze względu na wieloznaczne użycie terminu „wskaźnik” w literaturze naukowej, konieczne jest przyjęcie w miarę jednoznacznej definicji dla całego tekstu, a przynajmniej takiej, która nie budzi większych kontrowersji. Pragniemy uniknąć debaty nad pojęciem wskaźnika, wykraczającej poza ramy niniejszego opracowania⁵. Przyjmujemy uproszczone znaczenie tego pojęcia, wystarczające dla wyłożenia tez artykułu. W tym miejscu bardziej interesuje nas kontekst użycia wskaźników w ramach polityki publicznej oraz szanse i zagrożenia, jakie się z tym wiążą i tym zagadnieniom chcemy poświęcić maksymalnie dużo miejsca. Przez wskaźniki rozumimy tu zmienne odzwierciedlające badany obszar rzeczywistości, których wartości, ustalone w wyniku pomiaru, pozwalają określić występowanie lub rozmiar zjawisk będących obiektem zainteresowania.

Wskaźniki służą diagnozie, programowaniu oraz ewaluacji programów: są narzędziem umożliwiającym pomiar stanu przed wdrożeniem działań, pozwalają na określenie skali zjawisk problemowych i ich związku z innymi wymiarami kontekstu, umożliwiają jasną prezentację celów operacyjnych działania, a także pomiar efektów działań, co warunkuje rozliczenie podmiotów za nie odpowiedzialnych. Dobrze skonstruowane wskaźniki umożliwiają wreszcie eliminację działań nieskutecznych lub nieefektywnych – ułatwiają społeczne uczenie się. Należy jednak pamiętać, że dobór wskaźników jest raczej konsekwencją wstępnego zidentyfikowania i konceptualizacji problemów lub wyzwania niż sposobem na ich odkrycie.

W przypadku diagnozy, wskaźniki wiążą ją z celami programów i ich analizą. Źródłem strategicznych celów programowych są wizje polityczne. Analizy dostarczają inspiracji i możliwości krytycznej (samo)oceny wizji. W celu zapewnienia krytycznej diagnozy i weryfikacji koncepcji warto posługiwać się wskaźnikami umożliwiającymi np. porównanie państw lub regionów, tzn. stosować wypracowane na polu międzynarodowym (międzyregionalnym) wspólne, standardowe portfele wskaźników.

⁵ Więcej na ten temat piszemy w Górniak, Keler 2007.



Ryc. 2. Logiki wykorzystania wskaźników społecznych

Źródło: opracowanie własne w nawiązaniu do Land (2000).

Standardowo używane zestawy wskaźników, stosowane głównie dla celów porównawczych, należy uzupełniać o wskaźniki odzwierciedlające zmianę zawartą w wizji strategicznej, szczególnie zaś gdy standardowe nie opisują specyfiki lokalnej sytuacji.

Użycie wskaźników w działalności związanej z prowadzeniem polityki publicznej ma długą tradycję. Ważny wkład w rozwój wykorzystania wskaźników miał zapoczątkowany w latach 60. XX w. tzw. „ruch wskaźników społecznych”⁶. Sposoby wykorzystania wskaźników we współczesnych politykach publicznych dobrze klasyfikuje koncepcja Kennetha Landa. Wskazał on na trzy sposoby użycia wskaźników społecznych, które określił mianem „logik wykorzystania” (*rationales*) wskaźników, w kontekście prowadzonych polityk: logika zmiany społecznej, logika raportowania społecznego i logika polityk publicznych. Logiki te mają charakter hierarchiczny – przyjęcie drugiej oznacza także przyjęcie pierwszej, a przyjęcie trzeciej – dwóch pozostałych (Ekos 1998, s. 3)⁷.

⁶ Podstawowym bodźcem rozwoju ruchu było pytanie o koszty społeczne wzrostu gospodarczego (por. Noll 2002) i generalnie – działalności państwa. Wskaźniki społeczne miały pozwolić na uzupełnienie obrazu, jaki wyłaniał się z na bazie danych ekonomicznych, o dodatkowe wymiary. Wskaźniki społeczne rozumiane są zwykle jako „dane liczbowe, serie statystyk, i inne formy dowodów, które pozwalają nam ocenę stanu obecnego i tendencji zmian w odniesieniu do wyznawanych przez nas wartości i stawianych przez nas celów strategicznych” (Bauer (red.) 1966, s. 1, „statistics, statistical series, and all other forms of evidence – that enable us to assess where we stand and are going with respect to our value and goals”).

⁷ Cytowane w publikacji, na którą się powołujemy, za: K.C. Land. (1975). „Social Indicator Models: An Overview”, rozdział 2 publikacji *Social Indicator Models*. New York: Russell Sage Foundation.

- (1) Pierwsza z nich – logika zmiany społecznej – wiąże się silnie z funkcją poznawczą wskaźników. Wskaźniki miały pozwolić na badania porównawcze, ukazywanie różnic czy zmienności zjawisk i procesów w różnych społecznościach. Funkcja wskaźników w ramach tej logiki ma charakter czysto opisowy. Wiąże się ona ściśle z możliwościami diagnostycznymi. Wszystkie wskaźniki winny wpisywać się w tę logikę.
- (2) Druga – logika raportowania społecznego – oznacza budowanie powiązania między wskaźnikami służącymi diagnozie z ich ważną rolą, jaką jest „oświecania społeczeństwa” (Land 2000, s. 2686), tę logikę można nazwać podejściem „ostrzegawczym” (Anderson Moore, Brown, Scarupa 2003, s. 3). W ramach tego stanowiska pomiar i analiza zjawisk wiążą się z wartościami, które ważne są w danym społeczeństwie. Dzięki wskaźnikom można sprawdzić, czy ogólnie efekty różnorodnych podejmowanych działań są pozytywne czy negatywne. Wskaźniki pozwalają także patrzeć w przyszłość, przy czym niekoniecznie, jak to podkreślał Raymond Bauer, na zasadzie ścisłej predykcji naukowej, a raczej na „antycypacji”. Predykcja polega na identyfikacji najbardziej prawdopodobnego stanu. Wskaźniki (w ramach raportowania społecznego) nie są bynajmniej narzędziami rejestracji wydarzeń historycznych, akcent położony jest na przyszłość, na dostarczanie bazy dla planowania przyszłych polityk. Planowanie wymaga uwzględnienia nie tylko najbardziej prawdopodobnych stanów rzeczy, ale także tych w ogóle prawdopodobnych a zarazem dość istotnych (zob. Bauer 1966, s. 17–19). W praktyce oznacza to wskazywanie obszarów ryzyka i ukierunkowanie działań polity-

ki na wyróżnione obszary, co może zapobiec eskalacji problemów.

- (3) W przypadku trzeciej – logiki polityk publicznych – punktem odniesienia i drogowskazem analiz staje się wpływ konkretnej interwencji. Logika polityk publicznych czerpie z logiki raportowania, jednakże zdecydowanie mocniej wiąże wskaźniki z polityką, z programami realizowanymi w jej ramach. Kristin Anderson Moore i Brutt V. Brown wyróżniają trzy obszary wykorzystania wskaźników w ramach logiki polityk publicznych – choć się do niej literalnie nie odnoszą: (a) ustanawianie celów (*setting goals*), (b) zwiększanie znaczenia sensownego planowania (i skutecznego działania) przez sukcesywne rozliczanie z osiągnięć (*increasing accountability*) i wreszcie – (c) włączanie refleksji w praktykę (*reflective practice*) przez informowanie społeczności o zachodzących zmianach, a także poprzez ciągłe weryfikowanie wdrażanych programów, uczenie się z praktyki (zob. Anderson Moore, Brown 2003, s. 3). Ten podział zdaje się precyzyjnie uwypuklać najważniejsze elementy podejścia do wykorzystania wskaźników właściwego logice polityk publicznych.

Wskaźniki użyte zgodnie z logiką raportowania pozwalają na identyfikowanie, ocenę i antycypowanie zmian w skali całych obszarów zainteresowania polityk publicznych. Wobec wielości interakcji zjawisk społecznych i ekonomicznych z trudem, jeśli w ogóle, zmiany te można precyzyjnie wiązać z określonymi interwencjami. Możemy jednak przynajmniej oceniać, czy sprawy idą w dobrym, czy w złym kierunku. W przypadku ściślejszego powiązania wskaźników z celami określonych programów – mamy do czynienia z wykorzystaniem wskaźników zgodnie z logiką polityk publicznych. Oczywiście i w przypadku takiego podejścia powstaje problem z możliwością jednoznacznego przypisania „chwytanych” przez wskaźniki zmian konkretnym interwencjom, zwłaszcza na poziomie wpływu wykraczającego poza bezpośrednich beneficjentów. Niemniej, intencją wykorzystania wskaźników w tym kontekście jest uchwycenie przyczynowego wpływu interwencji na wystąpienie zmian ujętych w postaci celów programu. Oczywiście jest tu związek pomiędzy wykorzystaniem wskaźników a ewaluacją.

2.3. Ewaluacja

Jak definiuje się ewaluację? Rozumienie pojęcia ewaluacji nie jest jednolite, co wynika między innymi z rozmaicie określanymi funkcjami tego przedsięwzięcia. Tutaj przyjmiemy dość szeroko i pragmatycznie stosowane rozumienie ewaluacji jako przedsięwzięcia:

- a) zmierzającego do określenia na podstawie właściwie zgromadzonych i przetworzonych informacji, w jakim stopniu dane rozwiązanie (np. interwencja publiczna: polityka, program lub projekt) spełnia ustalone kryteria;
- b) w tym, w szczególności:
 - o w jakim stopniu osiągnęło cele, dla których zostało podjęte
 - o jakie są relacje pomiędzy nakładami, działaniami i efektami tego rozwiązania.

Tak rozumiana ewaluacja wpisuje się w ciąg działań związanych z kreowaniem programów publicznych, określonego przez nas w tym artykule mianem cyklu programów publicznych: domyka ten cykl, przynosząc wiedzę na temat zrealizowanego programu i wnioski dla kolejnych przedsięwzięć.

W cyklu polityk publicznych (ryc. 1) ewaluacja jest nie tyle zakończeniem łańcucha działań, ile raczej spoiwem pozwalającym na stałe doskonalenie programów zorientowanych na rozwiązanie problemów społecznych. W takim ujęciu ewaluacja występowałaby tylko raz w zaprezentowanym cyklu, w miejscu, w którym została umieszczona ewaluacja *ex-post*. Tak też jest lokowana na ogół w prezentacjach cyklu polityk publicznych (por. np. Bridgman, Davis 2004). W przedstawionym tutaj zmodyfikowanym ujęciu cyklu polityk (programów) publicznych ewaluacja pojawia się trzykrotnie:

- a) w fazie wstępnej, gdy dokonywana jest analiza przyjętych rozwiązań pod kątem potencjalnej zdolności osiągnięcia efektów i poprawności konstrukcji wszystkich elementów projektu – ewaluacja *ex-ante*;
- b) w trakcie realizacji, zwłaszcza w przypadku przedsięwzięć złożonych, o długim okresie wdrożenia, gdy możliwe i celowe są korekty – ewaluacja śródk okresowa lub w toku programu (*on-going*);

c) po zakończeniu wdrożenia – ewaluacja *ex-post*; jest to ewaluacja *sensu stricto* w klasycznym cyklu polityk publicznych.

Na każdym etapie ewaluacja oznacza badanie i ocenę programu pod kątem stosownego zestawu kryteriów i wyprowadzanie wniosków dla decydentów. W przypadku dokonywania ewaluacji na różnych etapach cyklu inaczej rozkłada się akcenty, ale ewaluacja w swej istocie jest zawsze ocenianiem pewnego projektu rozwiązań.

Przydatne do rozważań nad rolą ewaluacji w cyklu (programów) polityk publicznych jest wskazanie na trzy zasadnicze jej funkcje (zob. Chelimsky 1997):

- a) rozliczenia i egzekwowania odpowiedzialności (*accountability*);
- b) poznawczą (*knowledge*);
- c) stymulowania usprawnień i rozwoju organizacyjnego (*development*).

Funkcji rozliczenia i egzekwowania odpowiedzialności nie należy mylić z kontrolą – ma ona sprzyjać użyteczności prowadzonych działań i legitymizacji sposobu wydawania publicznych pieniędzy. Funkcja poznawcza wiąże się z gromadzeniem wiedzy dotyczącej związków przyczynowo-skutkowych pomiędzy podejmowanymi działaniami a problemami, które mają rozwiązać⁸. Funkcja stymulowania usprawnień i rozwoju organizacyjnego – zwana rozwojową, formatywną, optymalizującą – wiąże się z dążeniem do usprawnienia własnego działania w toku podejmowanych działań, uczeniem się z doświadczeń. Musi to zakładać dopuszczalność błędów i niepowodzeń. Jednak każde niepowodzenie powinno być przedmiotem starannej analizy jego przyczyn i prowadzić do konkluzji dla przyszłych interwencji. W ten sposób niepowodzenia nie muszą oznaczać marnotrawstwa środków publicznych, lecz koszt uczenia się, który prowadzi do skuteczniejszego rozwiązywania problemów społecznych w przyszłości⁹.

Warto dokładniej wskazać powiązanie między funkcjami ewaluacji a wskaźnikami w przy-

padku dwóch podstawowych kwestii: oceny efektów programów i planowania programów. Tym dwóm obszarom poświęcone są dwie kolejne części artykułu.

3. Rola ewaluacji i wskaźników w ocenie zrealizowanych programów

Dokonanie ewaluacji *ex-post*, czyli po wdrożeniu działań w życie i zrealizowaniu wszystkich etapów programu, niesie kilka zasadniczych korzyści. Ewaluacja pozwala sprawdzić sensowność i trafność założeń ujętych w planie, wskazać na najskuteczniejsze i najefektywniejsze sposoby ich realizacji oraz zakomunikować społeczeństwu, jakie efekty przynoszą działania finansowane z funduszy publicznych. Jak już to ujęli Osborne i Gaebler (1992), jeśli potrafisz zaprezentować efekty, możesz pozyskać poparcie społeczne. Ewaluacja wiąże się zatem z funkcją komunikacji społecznej, przedstawienia wartości dodanej naszych działań, wskazania, że są one korzystne dla społeczeństwa.

3.1. Od nakładów do wpływu

Pierwszym warunkiem możliwości realizacji ewaluacji *ex-post* jest to, by cele interwencji sformułowane zostały w sposób precyzyjny i pozwalający na dobór wskaźników (szerzej omawiamy to zagadnienie poniżej, w części pt. „Rozróżnienie celów na *goals*, *objectives*, oraz *targets*”). Musimy wiedzieć, co mierzymy, oraz być w stanie dokonać pomiaru efektów. Jakie aspekty programu mierzymy w tym zakresie?¹⁰:

- a) nakłady – poniesione koszty, bezpośrednie i pośrednie, finansowe i rzeczowe;
- b) produkty – wykonane działania;
- c) rezultaty – zmiany uzyskane w układzie docelowym (na poziomie grupy beneficjentów lub obiektów objętych działaniami), które można przypisać podjętym działaniom;
- d) wpływ/oddziaływanie – stopień faktycznego i względnie trwałego rozwiązania problemu, który miał być rozwiązany przez oddziaływanie na układ docelowy.

⁸ Pisali o tym Sabatier i Mazmanian w swoim klasycznym artykule (zob. Sabatier, Mazmanian 1982). Odkrywanie związków przyczynowo-skutkowych uznali za najważniejszy warunek – po jasności sformułowanych celów – skutecznej implementacji polityk publicznych.

⁹ Szerzej na temat funkcji ewaluacji zob. Górniak (2007).

¹⁰ Jest to standardowy sposób ujęcia aspektów programu w dokumentach publikowanych przez Komisję Europejską, zob. np. European Commission 2006b).

W toku planowania programów postępuje się od celów najogólniejszych do tych najbardziej szczegółowych. W przypadku ewaluacji postępujemy na odwrót – najpierw dokonujemy ewidencji zmiany na poziomach najbardziej szczegółowych – najbardziej konkretnych, a dopiero później możemy zająć się badaniem wpływu działań na osiągnięcie celów z poziomów wyższych.

Przejsie pomiędzy rezultatami a wpływem jest najtrudniejsze. Przypisanie rezultatom – czyli efektom programu dla grupy jego beneficjentów ostatecznych, będących ułamkiem populacji docelowej – wpływu na szersze układy społeczne jest zadaniem praktycznie w tej chwili niewykonalnym. Brakuje nam narzędzi pozwalających wyodrębnić wpływ danego programu czy polityki z wpływu wielu innych interwencji i procesów nieobjętych oddziaływaniem polityk publicznych. Badanie tego rodzaju efektów, jak efekt substytucji, wyparcia czy podwójnego liczenia w takich szerokich układach jest niezwykle ambitnym i często niewykonalnym zadaniem.

Warto w omawianym kontekście rozróżnić ewaluację na poziomie strategicznym polityk publicznych od tej na poziomie operacyjnym programów i projektów. Pomiar rezultatów i pomiar wpływów jest możliwy, ale dokonujemy go na poszczególnych poziomach – strategicznym i operacyjnym – w innym celu i za pomocą innych narzędzi.

Pomiar rezultatów jest kluczowym przedmiotem ewaluacji operacyjnej. W miarę możliwości rezultat należy mierzyć jako zmianę wywołaną interwencją, a więc zmianę netto, ustaloną jako różnica między uzyskanym wynikiem a jego wielkością, której można by oczekiwać w sytuacji braku działań skierowanych do danej grupy ostatecznych beneficjentów. Brak rezultatu netto oznacza niemożność uznania zmiany wartości wskaźników wpływu/oddziaływania jako skutku działań podjętych w ramach programu. Jeżeli nie odnotowano zmiany na poziomie wskaźników rezultatu (zmiany na poziomie ostatecznych beneficjentów działań), to należy uznać, że zmiana odnotowana za pomocą wskaźników wpływu/oddziaływania jest efektem innych czynników niż program.

Pomiar wpływu jest podstawą ewaluacji strategicznej. Ostateczny cel każdej polityki publicz-

nej to wywołanie zmiany stanowiącej całkowite lub częściowe rozwiązanie pewnego problemu społecznego, czego wyrazem jest osiągnięcie stosownych wartości docelowych wskaźników wpływu/oddziaływania. Zmiana na poziomie tych wskaźników powinna być trafnym, rzetelnym i obiektywnym odzwierciedleniem zmiany społecznej, która została określona jako strategiczny cel programu. W tym celu używamy zwykle zestawów („portfeli”) wskaźników. Wynika to z tego, że cele strategiczne mogą mieć charakter wielowymiarowy i być definiowane w kategoriach trudnych do pomiaru za pomocą tylko jednego wskaźnika (np. konkurencyjność, innowacyjność, kapitał ludzki, ubóstwo, wykluczenie społeczne). Pomimo problemów z przypisaniem polityce publicznej zmian wykraczających poza bezpośrednie oddziaływanie interwencji, należy te zmiany analizować i oceniać, gdyż brak pozytywnych efektów społecznych (oddziaływania) stawia w wątpliwość sensowność interwencji. Mamy tu do czynienia z asymetrią: z większą pewnością możemy wykazać brak oddziaływania polityki publicznej, niż ją potwierdzić (jeśli to jest w ogóle możliwe).

Stosujemy tutaj inną definicję rezultatu i oddziaływania (wpływu) niż przyjęta w dokumentach metodologicznych Komisji Europejskiej (por. np. European Commission 2000). Proponujemy traktowanie jako **rezultatów** wszystkich zmian wywołanych interwencją na poziomie bezpośrednich beneficjentów zarówno zmian natychmiastowych (które także Komisja Europejska traktuje jako rezultaty), jak i zmian trwałych lub pojawiających się po pewnym czasie po zakończeniu działań. Takie rezultaty możemy umownie nazwać trwałymi. Komisja Europejska każe traktować trwałe rezultaty jako oddziaływanie (*impact*), dodając przy tym kwalifikację: oddziaływanie lokalne. Natomiast wpływem globalnym określa oddziaływanie wykraczające poza bezpośrednich beneficjentów czy układ, na który skierowane są działania. Takie ujęcie jest zdecydowanie bardziej arbitralne i prowadzi do wielu kłopotów, a najczęściej do zaniechania oceny szerszych konsekwencji programów, pozytywnych i negatywnych, oczekiwanych i niezamierzonych, na rzecz koncentracji uwagi na „oddziaływaniach lokalnych”, czyli na pomiarze rezultatów z opóźnieniem w sto-

sunku do zakończenia programu. To ostatnie nie jest czymś złym czy złym, wręcz przeciwnie: na poziomie operacyjnym o ocenie programu powinna decydować trwałość rezultatów i ich poziom wywołany rzeczywiście działaniami w ramach programu. Nie należy jednak tych wysiłków ewaluacyjnych traktować jako pomiaru stopnia globalnego rozwiązania problemu społecznego przez program, a raczej jako uchwycenie efektów lokalnych. Lepiej unikać zamieszania terminologicznego.

3.2. Ewaluacja oparta na teorii programu

Ustalenie związków przyczynowo-skutkowych w projektach publicznych jest trudne ze względu na złożoność procesów społecznych i problemy z kontrolowaniem wpływu czynników kontekstowych na uzyskiwane wyniki. Sugerujemy, by skoncentrować się w odkrywaniu takich zależności na poziomie powiązania działań/produktów z rezultatami (także trwałymi); kolejnym krokiem w ewaluacji jest powiązanie rezultatów z wpływem, ale – jak już wyżej wspomnieliśmy – jest to trudne lub niewykonalne, a już na pewno obciążone dużym błędem. Nie chcemy przez to powiedzieć, iż wysiłki w tym zakresie są chybione – wręcz przeciwnie, uważamy, że należy je podejmować i w niektórych, wybranych dziedzinach być może przyniosą efekty (np. dobre wyniki uzyskano w dziedzinie ustalania wpływu polityki monetarnej banków centralnych na poziom inflacji, choć i tu sukces nie jest gwarantowany). Dobre podejście to budowanie teorii dotyczących mechanizmu wpływu określonego rodzaju interwencji i formułowanie na ich podstawie prognoz dotyczących oczekiwanych zmian ujmowanych za pomocą wskaźników. Konfrontacja takich prognoz z danymi i faktycznymi efektami pozwala eliminować błędne koncepcje polityk publicznych.

Klasyyczny, pozytywistyczny paradygmat ewaluacyjny silnie promował stosowanie quasi-eksperymentu jako podstawowego planu badawczego pozwalającego zmniejszyć ryzyko nieprawidłowego ustalenia związków przyczynowych pomiędzy działaniami i efektami. Kłopoty z jego stosowaniem w realiach implementacji programów publicznych oraz łatwość popadania w formalizm badawczy (mamy tu na

myśli sprowadzanie wyjaśnienia do raportu statystycznego dotyczącego istotności związków pomiędzy bodźcami i reakcjami, bez uzyskiwania wglądu w mechanizm transmisyjny) doprowadziły do krytyki tego podejścia. Wskazuje się na konieczność oparcia ewaluacji na teorii dotyczącej mechanizmów wiążących działania z wynikami i na potrzebę jakościowego wglądu w procesy związane z realizacją programów czy projektów, wpływające zarówno na ich powodzenie, jak i na porażki.

Każda interwencja publiczna, niezależnie od jej genezy, wynika ze swoistej „teorii”: opiera się na hipotezie, że w danych warunkach podjęcie działań wynikających z projektu interwencji doprowadzi do osiągnięcia założonych skutków. Polityki publiczne opierają się na takich przyczynowych konstrukcjach. Nie zawsze są one wyrażane *explicite*, wręcz przeciwnie, często ukryte są pod powierzchnią formułowanych celów i przyjmowanych sposobów działania. Po to, by przeprowadzić dobrze ewaluację, należy właściwie uchwycić istotę analizowanej interwencji, czyli jej „teorię”. „Teorii programu”, to jest założeń co do związków przyczynowych pomiędzy podejmowanymi działaniami a pożądanymi efektami, towarzyszy na ogół zbiór przekonań dotyczących najbardziej właściwych sposobów wdrożenia interwencji – „teoria wdrożenia”. Efektem takiej „teorii wdrożenia” są przyjmowane rozwiązania w zakresie struktur wdrożenia (np. centralizowane *vs* zdecentralizowane), sposobu zarządzania, organizacji monitoringu, kontroli finansowej itp. Rozpoznanie „teorii wdrożenia” dominującej w gronie osób projektujących interwencję i decydentów pozwala na lepsze zrozumienie polityki czy programu i jej pogłębioną ewaluację. Ewaluacja, w szczególności końcowa, sprawdza, czy w świetle uzyskanych efektów „teoria programu” daje się utrzymać, czy też należy ją odrzucić. Koncepcje działania, które pomyślnie przejdą ewaluację, tworzą korpus wiedzy podnoszącej skuteczność polityk publicznych. Faktycznie, spontaniczny proces weryfikacji występuje zawsze – tak czy owak – w postaci społecznej oceny skutków prowadzonych polityk. Ewaluacja jest metodologicznie kontrolowanym i usystematyzowanym sposobem takiej weryfikacji prowadzonym przez ekspertów. Z natury rzeczy może się zatem przyczynić do legitymizacji

lub delegitymizacji sposobu rządzenia w oczach wyborców, którzy oceniają na podstawie kryterium skuteczności i efektywności.

Polityki i programy publiczne powstają w wyniku splotu inicjatywy i twórczych idei polityków ze zracjonalizowanym, opartym na wiedzy eksperckiej działaniem administracji i ograniczeniami narzucanymi przez warunki działania. Ich cele nie wynikają automatycznie z diagnoz prowadzonych przez surowych, obiektywnych analityków. Cele polityk publicznych są pochodną wizji politycznych i ideologii rozwojowych, rezultatem konkurencji politycznej i ścierania się różnych interesów. Ekspertka diagnoza i analiza polityk publicznych, jak już wspomnieliśmy, powinna sprawdzać prawidłowość diagnoz i definicji sytuacji przyjmowanych przez polityków i osoby formułujące programy, ustalać zależności przyczynowe, badać potrzeby społeczne, bilansować zasoby niezbędne do działania, rysować scenariusze wynikające z proponowanych działań. Jednakże zarówno ekspertka diagnoza, jak i analiza polityk publicznych stanowią zawsze tylko wsparcie dla inicjatywy polityków. Ewaluacja *ex-ante* zamyka taki proces intelektualnej „przymiarki” proponowanych rozwiązań i wypowiada się na temat szans zrealizowania postawionych celów za pomocą takiego, a nie innego zestawu środków. Zakładając idealną sytuację, ewaluator *ex-ante* powinien korzystać z bazy wiedzy zgromadzonej przez ewaluację *ex-post*. Oparcie się na sprawdzonych „teoriach programów” podnosi prawdopodobieństwo sukcesu, choć ze względu na zmienność warunków działania nie gwarantuje idealnego powtórzenia efektów.

Ewaluacja ma dużo wyższy potencjał falsyfikacyjny – odrzucania interwencji nieskutecznych lub przynoszących negatywne skutki – niż confirmacyjny – potwierdzania, że osiągnięte efekty są wynikiem interwencji, a nie czynników od niej niezależnych. To powoduje pewną rezerwę decydentów wobec ewaluacji; wolą oni analizę polityk publicznych, która jest prospektywna (podczas gdy ewaluacja w ścisłym tego słowa znaczeniu jest retrospektywna), oparta na koncepcjach teoretycznych i wiedzy eksperckiej oraz nie związana aż tak bardzo z procedurami metodologicznymi. Także ewaluatorzy nie chcą pozostawać w roli krytycznych „falsyfikatorów” i rozwijają koncepcje ewaluacji, która jest bliższa

uczestnikom procesu wdrożenia rozwiązań, staje się faktycznie moderowaną przez „ewaluatorów” refleksją uczestników interwencji i jej beneficjentów nad swoją rolą w realizowanym programie i prowadzi do ciągłego ulepszania prowadzonych działań. Takie koncepcje pojawiają się w ramach tzw. nurtu konstruktywistycznego czy też ewaluacji partycypacyjnej, zwanej też ewaluacją czwartej generacji. Nie negując samej możliwości uwzględnienia partycypacji wszystkich zainteresowanych stron w procesie ewaluacji i pragmatycznej wartości takiego podejścia, trzeba stwierdzić, że ewaluacja nie może odchodzić od samej swej istoty – oceny efektów (potencjalnych i ostatecznych) interwencji.

3.3. Badanie efektu netto

Można słusznie zadać sobie pytanie, jak w praktyce wykorzystać podejście oparte na teorii. Jednym z podstawowych zadań ewaluacji jest ustalenie związku przyczynowego między podjętą interwencją a osiągniętymi wynikami. Ustalenie przyczynowego wpływu interwencji wymaga:

- a) zmierzenia efektu całkowitego interwencji jako ogólnej zmiany w zakresie zdefiniowanym przez cel (zmiany wartości przyjętych wskaźników);
- b) oddzielenia zmiany niezależnej od podjętej interwencji od zmiany, którą można przypisać wpływowi interwencji (wyodrębnienia efektu netto).

Dla ustalenia efektu netto niezbędne jest oszacowanie, co by się stało, gdyby nie podjęto interwencji. Tak formułuje to podręcznik wydany przez Komisję Europejską: „Gdy mówimy, że dane efekty zostały wyprodukowane lub spowodowane przez program, oznacza to, że gdyby program nie miał miejsca lub przeprowadzono by go w innej formie lub stopniu, te efekty nie wystąpiłyby lub nie wystąpiłyby w tym samym stopniu. Oznacza to, że ważne jest, by mieć precyzyjny obraz tego, co stałoby się bez programu. Nazywa się to sytuacją kontrfaktyczną” (Komisja Europejska 1997, s. 39).

Skutki, które wystąpiłyby nawet w sytuacji, gdyby program wydatków publicznych nie miał miejsca (sytuacja kontrfaktyczna) nazywa-

my efektem *deadweight*¹¹. Ponadto, chcąc wyodrębnić wpływ podjętego działania, podczas pomiaru zmiany społecznej musimy zbadać, czy nie zaszedł tzw. efekt substytucji lub wyparcia (przesunięcia) – sytuacja, w której pozytywne wyniki programu w odniesieniu do pewnych osób, grup lub obszarów zostały uzyskane kosztem innych osób, grup lub obszarów. Ten rodzaj efektów bardzo trudno uchwycić. Dodajmy, że jego wystąpienie jest bardziej prawdopodobne w sytuacji, gdy interwencja wdziera się na obszar normalnego funkcjonowania konkurencji rynkowej, a nie pozostaje w obszarze, na którym rynek zawodzi. Przykładem interwencji naruszających konkurencję rynkową i dlatego na ogół nieskutecznych w sensie efektu netto są bezpośrednie dotacje dla przedsiębiorstw, które odwracają uwagę firm od konkurowania o względy klientów, zastępując to zabieganiem o względy jednostek przyznających środki publiczne. Efektem takiej orientacji interwencji może być w szczególności obniżenie jakości nawet całej oferty branży (tak się może stać wkrótce z rynkiem usług szkoleniowych – skutecznie będą na nim oferowane głównie szkolenia, na które zostanie zdobyta dotacja, a nie te, które w warunkach konkurencji wybrałyby klienci).

Nacisk na ustalanie efektu netto wyraźnie wskazuje na oparcie ewaluacji na fundamentach przyczynowości. Pociąga to za sobą wymóg ilościowego pomiaru oraz dążenia do obiektywizmu. Aby zmierzyć efekt netto, trzeba dobrze zrozumieć mechanizmy zachodzące podczas realizacji projektu, wyodrębnić zarówno czynniki sprzyjające, jak i hamujące. Dobrze jest włączyć do tego wyjaśnienia wiedzę jak najszerszej grupy interesariuszy, dostępną przede wszystkim za pomocą metod jakościowych.

Wspomnieliśmy już, że badanie efektu netto na poziomie najogólniejszym – całego programu – jest zadaniem niezwykle trudnym i należy rozpocząć go na poziomie rezultatów, natychmiastowych i po pewnym czasie. W ustaleniu efektu programu na poziomie wpływu mogą pomóc tzw. wskaźniki kontekstowe, które pokazu-

ją szerszy zakres zmian, jakie zachodzą w społeczeństwie (np. demograficzne czy ekonomiczne). Takie ujęcie pozwala precyzyjniej zinterpretować zmianę ujętą przez pomiar wartości wskaźników przed dokonaniem interwencji i po jej ukończeniu. Skupienie się na poziomie rezultatów daje realną szansę na przełożenie wniosków z ewaluacji na użyteczne rekomendacje dla przyszłych interwencji tego samego rodzaju prowadzonych w zbliżonym kontekście instytucjonalnym. Przede wszystkim umożliwia to wyeliminowanie z repertuaru programów publicznych działań ewidentnie nieskutecznych, co pozwala poprawić proces programowania. Warto pamiętać, że skuteczność na poziomie rezultatu nie gwarantuje osiągnięcia celów strategicznych odzwierciedlanych wskaźnikami wpływu/oddziaływania. Natomiast brak skuteczności na poziomie rezultatów gwarantuje, że ewentualna zmiana wartości wskaźników oddziaływania jest efektem innych czynników niż program.

3.4. Dobre wskaźniki

Wszystkie omawiane problemy prowadzą nas do zagadnienia odpowiedniego ustalenia wskaźników. Bez nich ewaluacja nie może do końca spełniać wskazanych wyżej funkcji, a w konsekwencji nie stanie się istotnym ogniwem w cyklu programów publicznych. Sam ten cykl będzie przez to niepełny i bardziej podatny na błędy.

Wiele się mówi i pisze o cechach, jakie powinny mieć dobre wskaźniki¹². W istocie większość z tych „zasad” dotyka oczywistych kwestii metodologicznych lub prakseologicznych. O tym, że wskaźnik czy cały zestaw wskaźników musi trafnie oddawać to, co mierzy, lub o tym, że dane muszą być rzetelnie zebrane, wiemy z niemal każdego podręcznika metodologii nauk społecznych. O tym, że użyteczność wskaźników dla procesu podejmowania decyzji zależy od ich aktualności – wiemy na podstawie zdrowego rozsądku. Tutaj warto jeszcze zwrócić uwagę na aspekt normatywny wskaźników, zwłaszcza wskaźników dotyczących diagnozy i oceny polityk publicznych na poziomie strategicznym. Jeżeli wskaźniki mają służyć artykulacji celów,

¹¹ W polskich tłumaczeniach używa się terminu angielskiego (por. European Commission 2006a, s. 13 i Komisja Europejska 2006, s. 16). Można go tłumaczyć jako efekt zdarzenia niezależnego.

¹² Więcej na temat kryteriów oceny wskaźników piszemy w Górniak, Keler (2007).

a cele są wypadkową diagnozy opisującej, jak jest, i normatywnych założeń dotyczących tego, jak być powinno, to jasne stanie się powiązanie wskaźników wykorzystywanych w programowaniu i ewaluacji z wartościami przyjmowanymi przez decydentów. Wybór konkretnych wskaźników wiąże się z systemem przekonań i wartości oraz ma wpływ na wybór środków prowadzących do celu. Dobrym tego przykładem może być dylemat, czy poziom ubóstwa należy mierzyć w odniesieniu do relatywnej czy też do absolutnej linii ubóstwa; zakładamy, że podzielaną powszechnie wartością jest zmniejszenie zasięgu ubóstwa. Wybór relatywnych wskaźników ubóstwa dla potrzeb programu walki z ubóstwem odzwierciedla założenie, że jest ono (i jego korelat, jakim jest wykluczenie społeczne) funkcją nierówności dochodowych, a drogą do jego przezwyciężenia staje się polityka transferowa prowadząca do zmniejszenia różnicowania dochodów. Wybór bezwzględnej linii ubóstwa i dążenie do zmniejszenia odsetka osób o dochodzie poniżej poziomu przez nią wskazywanego może wygenerować inny repertuar polityk publicznych, w tym zwrócenie większej uwagi na wzrost gospodarczy. Ustalenie wskaźników jest więc jednym z podstawowych elementów konstruowania programów publicznych, który powinien angażować wszystkich interesariuszy. Dlatego mówimy, że wskaźniki powinny być „uzgodnione”. Oczywiście, wszystkie te argumenty mają sens tylko wtedy, gdy poważnie podchodzi się do procesu planowania programów i ich implementacji zgodnie z przyjętymi założeniami. Tego niestety nie można zagwarantować.

4. Rola wyników ewaluacji i wskaźników w planowaniu polityk publicznych

W procesie planowania polityk (także programów i projektów) dużą rolę odgrywa (a może: powinna odgrywać) diagnoza, wyniki ewaluacji zrealizowanych już przedsięwzięć i ewaluacja *ex-ante* przygotowanego programu czy polityki. W każdym z tych przedsięwzięć istotną rolę odgrywa właściwy dobór wskaźników. Ewaluacja *ex-ante* pozwala dopracować sam program; powinna także sprawdzać możliwość oceny jego efektów na podstawie przyjętego zestawu podstawowych wskaźników i zaplanowanego sys-

temu monitoringu. Sprawdzenie „oceniałości” (*evaluability*) programu jest bardzo dobrym egzaminem jego konkretności, wymierności efektów i rozliczalności (*accountability*). Ewaluacja na etapie przygotowania projektów pozwala określić cele programu, jego model, na który składają się powiązane ze sobą cele i działania, zbadać, czy taki model jest odpowiedni i wreszcie wskazać, gdzie plan wymaga modyfikacji, co powinno być ewaluowane w trakcie i *ex-post* i jaki z tego będzie pożytek (Levitom i in. 1998).

4.1. Rozróżnienie celów na *goals*, *objectives* oraz *targets*

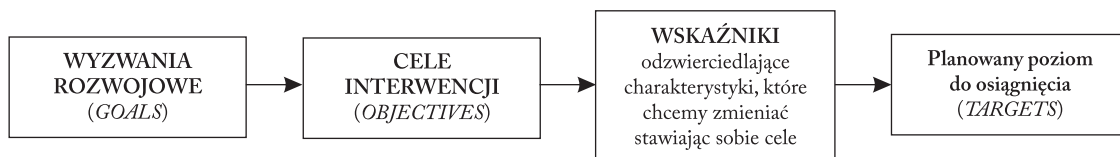
Polityki publiczne, programy czy projekty podejmuje się dla osiągnięcia pewnych celów. Powinny być one zapisane w dokumentach programowych. Jednakże samo pojęcie celu jest wieloznaczne i sprawia problemy. W języku angielskim, w kontekście programowania interwencji, funkcjonują trzy różniące się znaczeniem terminy, którym w języku polskim odpowiada jeden termin – „cel”:

- a) *goals* – czyli strategiczne cele rozwojowe, cele-wyzwania: na ogół pojedyncze lub nieliczne i często wieloaspektowo ujęte jakościowe wizje;
- b) *objectives* – czyli cele interwencji, cele-zadania: ujęte w sposób konkretny, mierzalny i określony w czasie;
- c) *targets* – czyli docelowe wartości wskaźników, które w danym czasie zamierzamy osiągnąć.

Konkretyzacja celów przebiega według linii wskazanej na ryc. 3. Cele rozwojowe związane z przyjmowanymi założeniami politycznymi przekładają się na cele poszczególnych programów, które w istocie dookreślają, co konkretnie ma znaczyć „rozwój”. Te drugie z kolei są dookreślane przez wskaźniki. Sprawą dyskusyjną jest to, czy ustalać cele w kategoriach *targets*. Do tej ostatniej kwestii wrócimy za moment.

Wymaga się, by cele-zadania (*objectives*), spełniały kryteria określane jako *SMART*, a więc by były:

- a) *Specific* – konkretne (co chcemy osiągnąć);
- b) *Measurable* – mierzalne (trzeba umieć zmierzyć to, co się chce osiągnąć);
- c) *Achievable* – osiągalne;



Ryc. 3. Wyzwania, cele interwencji i planowany poziom realizacji celów a wskaźniki

Źródło: opracowanie własne.

d) *Realistic* – realistyczne przy istniejących zasobach;

e) *Time-bounded* – określone w czasie.

Już na pierwszy rzut oka widać, że aby cel-zadanie (*objective*) mógł być *SMART*, powinien zostać opisany za pomocą wskaźników. Jednak nie ma mowy o dobrych wskaźnikach, jeśli cele nie są ustalone w sposób nie tylko merytorycznie sensowny, lecz także formalnie poprawny. Maniera formułowania celów w postaci ogólnych, jakościowych, niemierzalnych deklaracji prowadzi do niemożliwości dokonania ewaluacji (*evaluability*). Zasadność wyrażania celów interwencji w postaci, która pozwoli na pomiar efektów działań ukierunkowanych na realizację tych celów, trafnie podkreślili Osborne i Gaebler (1992):

- co da się zmierzyć, będzie zrobione;
- jeśli nie mierzysz efektów, nie odróżnisz sukcesu od porażki;
- jeśli nie możesz zaobserwować sukcesu, nie możesz go nagrodzić;
- jeśli nie możesz nagrodzić sukcesu, prawdopodobnie nagrodzisz porażkę;
- jeśli nie umiesz zaobserwować sukcesu, nie potrafisz wyciągnąć z niego nauki;
- jeśli nie możesz rozpoznać porażki, nie dasz rady jej skorygować;
- jeśli potrafisz zaprezentować efekty, możesz pozyskać poparcie społeczne.

Jeżeli do jednego celu dobierzemy jeden wskaźnik, wówczas, siłą rzeczy, to wybór tego, a nie innego wskaźnika będzie oznaczał ostateczne określenie celu, skoro to wskaźniki będą służyły jako narzędzia oceny programu. Aby uniknąć sytuacji, w której wybór wskaźników prowadzi do nadmiernego uproszczenia celów i sprowadza je do zawężonego rozumienia problemu, warto korzystać z większej liczby wskaźników – takich, które w pełni ujmą ustalony cel. W istocie rzeczy chodzi nawet nie tyle o samą wielość wskaźników, ile o dobre przemyśle-

nie wszystkich istotnych aspektów celu, które wskaźniki winny ująć.

4.2. Problem: do jakich poziomów celów interwencji (*objectives*) określać konkretną wartość do osiągnięcia (*targets*)?

Wróćmy do pytania, które zasygnalizowaliśmy już wcześniej – czy w ogóle należy ustalać cele w kategoriach *targets*, czyli docelowych wartości wskaźników, a jeżeli tak, to na których poziomach logiki programu¹³ należy (i do kogo należałoby to zadanie)? Ustalając cel, zadajemy sobie pytanie, gdzie jesteśmy, jeżeli chodzi o daną kwestię (diagnoza sytuacji) i co chcemy osiągnąć (pożądaný stan w przyszłości). Stawiając sobie cele, chcemy zmienić zastaną sytuację. Ewaluacja ma nam odpowiedzieć, które rezultaty osiągnęliśmy, które produkty przyczyniły się do osiągnięcia rezultatów i – choć, jak wspomnieliśmy, jest to bardzo trudne – stwierdzić, czy i jakie wystąpiło oddziaływanie. Zmiana społeczna, jaką chcemy uzyskać, może być wyrażona kierunkowo: sukcesem określonym w kategoriach wskaźników, może być zmianą wartości wskaźnika w pożądanym kierunku, np. zwiększenia stopy zatrudnienia. Jednakże wchodzi w grę również ustalenie wartości docelowej wskaźników (*target*), czyli zamiast podawać jako cel wzrost zatrudnienia wyrażonego wskaźnikiem – stopa zatrudnienia, określamy, że np. za trzy lata wartość wskaźnika ma wynieść 80%.

Ustalenie *targets* na poziomie produktów (*outputs*) naszych działań nie stanowi problemu, a nawet powinno być obowiązkowe. Poprawne zarządzanie oznacza, że jesteśmy w stanie z góry zaplanować, jakie działania zrealizujemy.

¹³ Termin „logika programu” czy „logika interwencji” (*intervention logic*) użyty jest tu w sensie nadanym mu w dokumentach Komisji Europejskiej, zob. np. European Commission (2006b).

Uchybienia w tym zakresie będą świadczyć na niekorzyść autorów planów działania.

Trudniejszym zadaniem jest określenie *targets* na poziomie rezultatów, a zwłaszcza wpływu, w szczególności gdy mówimy o wpływie całego programu czy jego większych komponentów, a nie rezultatach pojedynczego projektu. Z jednej strony, ustalenie *targets* zwiększy możliwość rozliczenia programów – dotyczy to zwłaszcza rezultatów. Z drugiej strony, sporządzenie prognozy przyszłego stanu większych układów społecznych lub gospodarczych i ustalenie, w jakim stopniu odstępstwo od prognozy *ceteris paribus* można przypisać interwencji, jest trudne, a często wręcz niemożliwe przy założeniu jakiegoś przyzwoitego poziomu precyzji, a to ze względu na bogactwo i złożoność wzajemnych oddziaływań rozmaitych czynników. Dodatkowo, nawet na poziomie rezultatów istnieje pokusa, by podawać wartości „bezpieczne”, najlepiej możliwe do osiągnięcia nawet bez udziału interwencji. Takie zachowawcze podejście autorów planów programów zmiany społecznej pozostaje praktycznie bez konsekwencji, gdy nie podejmujemy podczas ewaluacji badań nad efektem netto interwencji. Nie ma bowiem w takiej sytuacji możliwości wykazania, że fundusze publiczne wydawane są bezsensownie i nieefektywnie. Pojawia się w związku z tym pytanie: kto i dlaczego miałby ustalać *targets*, organizatorzy działania czy zewnątrzni eksperci? Stoimy na stanowisku, że ustalanie *targets* jest i celowe, i możliwe, a powinno to być wyrazem aspiracji decydentów politycznych skonfrontowanym z ekspercką diagnozą warunków realizacji. Tylko w przypadku produktów zaplanowane wartości docelowe należy poddawać rozliczeniu o charakterze audytorskim, natomiast założone w planach docelowe wartości wskaźników rezultatów (a zwłaszcza wskaźników wpływu/oddziaływania) trzeba konfrontować z efektami programu w procesie ewaluacji. Ewaluacja powinna ustalić nie tylko stopień osiągnięcia tych celów, lecz także określić zakres przyczynowego oddziaływania programu i wskazać czynniki sprzyjające oddziaływaniu i hamujące go. Na poziomie wpływu/oddziaływania programu i ten ostatni cel wydaje się często zbyt ambitny. To, na co nas stać, to na ogół stwierdzenie, czy sprawy poszły w planowanym lub w innym

kierunku. Jeśli w założonym, to pozwala nam to podtrzymywać przekonanie, że nasz program był skuteczny (o ile wystąpiły w nim pozytywne rezultaty netto); jeśli w przeciwnym – to powinniśmy pewnie uznać fiasko programu, ale prawdę na ogół poszukujemy usprawiedliwień: czynników poza programem, które zdecydowały o braku widocznego sukcesu (zła pogoda, zła koniunktura itp.). Rzadko jednak potrafimy rzetelnie dowieść, że program naprawdę osłabił oddziaływanie tych negatywnych czynników zewnętrznych.

5. Zakończenie

Ewaluacja powinna dostarczać nie tylko informacji, czy i w jakim stopniu zostały osiągnięte cele interwencji, lecz także – w jaki sposób udało się to zrobić, a w sytuacji porażek – dlaczego nie powiodło się ich uzyskanie (Sanderson 2000, s. 440). Wymaga to dobrego pomiaru wyników i nakładów osadzonego w jasno zdefiniowanym systemie celów działania i prawidłowo określonych i zaplanowanych zasobach potrzebnych do ich osiągnięcia. Trudno się nie zgodzić z popularnym wśród teoretyków ewaluacji stanowiskiem, że wymaga to sięgnięcia po bardziej złożoną metodologię, wykraczającą poza paradygmat pozytywistyczny, skupiony na pomiarze wyników i eksperymencie lub quasi-eksperymencie jako technice ustalania związków przyczynowo-skutkowych między działaniami a wynikami. Zapewne trzeba uwzględnić metody jakościowe i techniki ewaluacji partycypacyjnej. Trzeba silniej uwzględnić w ewaluacji kontekst społeczny i instytucjonalny działań oraz wartości i perspektywę beneficjentów interwencji. Ewaluacja musi być nastawiona na eksplorację i wyjaśnienie, a nie tylko na pomiar wyników.

Takie stanowisko nie zwalnia nas jednak z obowiązku poprawnego użycia wskaźników, gdyż aby ewaluacja stała się rzeczywiście użytecznym elementem cyklu polityk publicznych, musi być oparta na stabilnych fundamentach pomiaru efektów. Źle dobrane wskaźniki przynoszą więcej szkód niż całkowita z nich rezygnacja, mogą bowiem prowadzić do nieuzasadnionych wniosków, a te przekładać się na konsekwencje dla wdrażanych i przyszłych programów. Na poziomie strategicznym konieczne jest opraco-

wywanie portfeli wskaźników dla poszczególnych domen polityk publicznych i zasilanie ich przez służby statystyczne aktualnymi i rzetelnymi danymi. Na poziomie operacyjnym proces ustalania wskaźników wymaga metodycznego, często partycypacyjnego podejścia i ścisłego powiązania z procesem określania celów programów. Dobre ujęcie celów i ich konkretyzacja za pomocą wskaźników, umożliwiających pomiar efektów, czyni ewaluację logicznym i jasnym elementem cyklu polityk publicznych: dostarcza ona decydującym informacyjnego sprzężenia zwrotnego na temat efektów ich działań i poprawności „teorii programu”.

Bibliografia

- Anderson Moore K., Brown B.V., Scarupa H.J. (2003). *The Uses (and Misuses) of Social Indicators: Implications for Public Policy*. Washington: Child Trends, Research Brief.
- Bauer R.A. (red.) (1966). *Social Indicators*. Cambridge–London: MIT Press.
- Bridgman P., Davis G. (2004). *The Australian Policy Handbook* (wyd. 3). Sydney: Allen & Unwin.
- Chelimsky E. (1997). „The Coming Transformations in Evaluation”, w: E. Chelimsky, W.R. Shadish (red.), *Evaluation for the 21st Century*. London: Sage Publications.
- European Commission (2000). *Indicators for Monitoring and Evaluation: An indicative methodology. The New Programming period 2000–2006: methodological working papers. Working Paper*, nr 3.
- European Commission (2006a). *The New Programming Period 2007–2013. Indicative Guidelines On Evaluation Methods: Ex Ante Evaluation. Working Document*, nr 1, sierpień.
- European Commission (2006b). *The New Programming Period 2007–2013. Indicative Guidelines On Evaluation Methods: Monitoring And Evaluation Indicators. Working Document*, nr 2, sierpień.
- European Commission (2006c). *The New Programming Period 2007–2013. Indicative Guidelines On Evaluation Methods: Evaluation During The Programming Period. Working Document*, nr 5, październik.
- Ekos Research Associates Inc. (1998). *The Use of Social Indicators as Evaluation Instruments. Final Report*. Przygotowany dla Strategic Evaluation and Monitoring Human, Resources Development Canada, sierpień, SP-AH064E-08-98 (dostępny także po francusku).
- Górnjak J. (2007). „Ewaluacja w cyklu polityk publicznych”, w: S. Mazur (red.) *Ewaluacja funduszy strukturalnych – perspektywa regionalna*. Kraków: MSAP UE.
- Górnjak J., Keler K. (2007). „Wskaźniki w ewaluacji *ex-post* programów publicznych”, w: A. Haber (red.), „Ewaluacja *ex-post*: teoria i praktyka badawcza”. Warszawa: PARP.
- Komisja Europejska (1997). *Ewaluacja programów wydatków Unii Europejskiej: Przewodnik. Ewaluacja w połowie okresu i ex-post. Dyrekcja Generalna XLIX – Budżet*. (nieoficjalne tłumaczenie tekstu wykonane przez Urząd Komitetu Integracji Europejskiej oraz francuskie Krajowe Centrum Terytorialnej Służby Cywilnej).
- Komisja Europejska (2006). *Nowy okres programowania 2007–2013. Przewodnik po metodach ewaluacji: Wskaźniki monitoringu i ewaluacji. Dokument roboczy nr 2, sierpień* (polskie tłumaczenie European Commission 2006b).
- Land K.C. (2000). „Social indicators”, w: E.F. Borgatta, R.J.V. Montgomery (red.), *Encyclopedia of Sociology. Second Edition*, t. 4. New York–Detroit–San Francisco–London–Boston–Woodbridge: Macmillan Reference, s. 2682–2690.
- Leviton L.C. i in. (1998). „Teaching Evaluation Using Evaluability Assessment”, *Evaluation*, t. 4(4), s. 389–409.
- Noll H.-H. (2002). „Social Indicators and Quality of Life Research: Background, Achievements and Current Trends”, w: N. Genov (red.), *Advances in Sociological Knowledge over Half a Century*,. Paris: International Social Science Council, s. 168–206.
- Osborne D., Gaebler T. (1992). *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*. Reading: Addison-Wesley.
- Sabatier P.A., Mazmanian D.A. (1981). „The Implementation of Public Policy: A Framework of Analysis”, w: D.A. Mazmanian, P.A. Sabatier (red.), *Effective Policy Implementation*. Toronto: Heath, Lexington.
- Sanderson I. (2000). „Evaluation in Complex Policy Systems”, *Evaluation*, t. 6(4).

Evaluation and indicators in the cycle of planning and realization of public policies

The purpose of the article is to present the functions of evaluation and indicators in a public policy cycle, as well as discuss problems related to the execution of those functions. In particular, policy formulation and evaluation phases as well as measurement of programs' effects via indicators are taken into consideration. The first part includes a detailed description of the functions of diagnosis, indicators and evaluation in a public policy cycle. The second part of the paper outlines evaluation and the utilization of indicators at the evaluation stage of the cycle. The part concludes with a discussion of theory-driven evaluation and the significance of measuring net effects for the identification of causal impacts of a given intervention. The authors believe that we should treat as results all the intended effects of a given intervention (program, policy) for ultimate beneficiaries, regardless of when they occur, whether they are immediate or delayed, whereas impacts should be construed as intended and unintended effects for the broader system in which problems have arisen (the European Commission considers delayed results as impacts). Difficulties related to the measurement of impacts of an intervention impact are also discussed. The authors note that obtaining measurable results is a necessary condition for any impact of the intervention to take place. In the third part of the article, the authors discuss the role of indicators and evaluation in public policy formulation. Special attention is drawn to the need for a proper definition of objectives and the choice of relevant indicators in the process of public policy formulation. The authors underscore the fact that there exist considerable differences in defining purposes or intentions as goals, objectives and targets. The paper also discussed problems related to the setting of target values for indicators as well as the relationships among the measurement of outputs, results, and impacts of an intervention. In conclusion, the authors emphasize that evaluation and indicators can constitute useful tools in conducting public policies, but putting them into practice requires coping with problems discussed in the article.