OFICYNA NAUKOWA WARSZAWA 1997

EWALUACJA W EDUKACJI

REDAKTOR NAUKOWY
LESZEK KORPOROWICZ

E. R. House, *Evaluating with Validity*, s. 15–43, 139–157, tłumaczenie za zgodą Sage Publications, Inc/Corwin Press, Inc.
I. Simons, *Getting to Know Schools in Democracy*, s. 7–53, tłumaczenie za zgodą Falmer Press
E. R. House (ed.), *New Directions in Educational Evaluation*, s. 15–29, tłumaczenie za zgodą Falmer Press
N. K. Denzin, Y. S. Lincoln (ed.), *Handbook of Qualitative Research*, s. 236–248, tłumaczenie za zgodą Sage Publications, Inc/Corwin Press, Inc.
C. Robson, *Real World Research. A Resource for Social Scientists and Practitioner-Researchers*, s. 170–186, tłumaczenie za zgodą Blackwell Publishers Ltd.
M. Q. Patton, *How to use Qualitative Methods in Evaluation*, ch. 4, s. 70–106, tłumaczenie za zgodą Sage Publications, Inc/Corwin Press, Inc.
H. J. Rubin, I. S. Rubin, *Qualitative Interviewing. The Art of Hearing Data*, s. 42–64, tłumaczenie za zgodą Sage Publications, Inc/Corwin Press, Inc.
E. Chelimsky, *Where We Stand Today in the Practice of Evaluation: Some Reflections*, „Knowledge Policy" 1995, t. 8, nr 3, tłumaczenie za zgodą Transaction Publishers & „Knowledge and Policy"

© Copyright for the Polish edition by Program TERM FRSE, Ministerstwo Edukacji Narodowej

Wszelkie formy kopiowania są prawnie zastrzeżone

Redakcja merytoryczna
Wanda Lipnik

Projekt graficzny okładki i stron tytułowych
Piotr Mądrach

Wydawnictwo finansowane z funduszy Programu PHARE Unii Europejskiej — PROGRAM TERM

OFICYNA NAUKOWA
ul. Mokotowska 65/3, 00-533 Warszawa
tel. (0-22) 622-02-41; fax (0-22) 622-02-42

ISBN 83-85505-67-9

Spis treści

Przedmowa 7

Rozdział I. Czym jest i dlaczego potrzebna jest ewaluacja
Ernest R. House: Ewaluator w społeczeństwie 15
Helen Simons: Specyfika i rozwój studiów ewaluacyjnych 21

Rozdział II. Elementy teorii
David Nevo: Konceptualizacja ewaluacji edukacyjnej. Analityczny przegląd literatury 51
Helen Simons: Polityczne implikacje teorii ewaluacyjnych. Przybliżenie problemu 65

Rozdział III. Różnorodność podejść i ujęć
Ernest R. House: Ewaluacja i jej uprawomocnianie. Główne podejścia 95
Robert E. Stake: Studium przypadku 120

Rozdział IV. Projektowanie i metody badawcze
Colin Robson: Projektowanie ewaluacji 147
Michael Q. Patton: Obserwacja — metoda badań terenowych 163
Herbert J. Rubin, Irene S. Rubin: Jak zmierzać do celu nie wiążąc sobie rąk. Projektowanie wywiadów jakościowych 201

Rozdział V. Społeczny kontekst ewaluacji
Ernest R. House: Demokratyzacja ewaluacji 225
Eleanor Chelimsky: Praktyka ewaluacyjna dzisiaj. Kilka refleksji .. 244

Bibliografia 261
Słownik podstawowych pojęć 278
Indeks 281

Przedmowa

Przedstawiamy Państwu zbiór tekstów dotyczących dziedziny, która nie jest nowa. Nie pretenduje do odkrywania nie zbadanych czy choćby nie docenianych aspektów rzeczywistości. Korzenie systematycznie pojętej ewaluacji sięgają lat dwudziestych naszego wieku, a ostatnie dwadzieścia–trzydzieści lat zaowocowały tak szybkim i wszechstronnym jej rozwojem, iż nawet najlepsi znawcy problemu mają istotne trudności z przedstawieniem i analizą współczesnej ewaluacji jako całości. W znacznej mierze wynika to z faktu, iż jest ona tworem wysoce niejednorodnym, wyrastającym z pogranicza działalności naukowo-badawczej i administracyjnej, wkomponowanej w różnorodne strategie interwencji społecznej, działań planistycznych i kontrolnych wraz z ich kulturowym i politycznym kontekstem.

Ten ostatni sprawił właśnie, że stosunkowo wysoka popularność, a nawet swoiste zadomowienie ewaluacji w systemie instytucji publicznych stało się fenomenem jedynie krajów rozwiniętych, w których zaawansowanej gospodarce rynkowej towarzyszy ugruntowana demokratyzacja wszystkich podstawowych dziedzin życia. Nie bez powodu więc znajomość praktyk, jak i rozwój teorii ewaluacyjnych musiał być w polskich warunkach ostatnich dziesięcioleci niezwykle ograniczony, a ich obecny renesans jest ściśle związany z dynamiką przeobrażeń nie tyle świata nauki, ile całego systemu tych właśnie instytucji.

Analiza, a następnie zastosowanie profesjonalnych instrumentów oceny różnorodnych działań i programów społecznych poza bezpośrednim naciskiem scentralizowanych i wysoce upolitycznionych ośrodków władzy staje się wielkim wyzwaniem współczesnych strategii zarządzania, ale także elementem stymulującym rozwój, efektywność, a przede wszystkim jakość owych działań.

Uwaga ta w szczególny sposób dotyczy systemu edukacji. Od wielu już lat wymaga on gruntownej reformy nie tylko w sferze programowej i instytucjonalnej, ale w równej mierze pod względem wyposażenia w techniki analizy i oceny rzeczywistych jego efektów, widzianych przy tym znacznie szerzej niż przez pryzmat rutynowych wskaźników standardowej sprawozdawczości. Dlatego ewaluacja w edukacji stoi przed zadaniem szczególnym, opisuje bowiem „efekty", które nie zawsze są mierzalne, jako że dotyczą nie tylko przyrostu wiedzy, ale także zmian osobowości, postaw, kultury pedagogicznej, stylu pracy, a więc wartości, które ujawnić może analiza sposobu doświadczania, a nie „pomiar" ludzkiej rzeczywistości.

To właśnie przekonanie legło u podstaw wyboru zagadnień oraz tekstów niniejszego zbioru, który ma być wprowadzeniem w zagadnienia ewaluacji w ogóle, ze szczególnym jednak uwzględnieniem ewaluacji jakościowej, która w dziedzinie edukacji ma wyjątkowe znaczenie, pozwala bowiem nie tylko obserwować, ale i rozumieć opisywane działania, strategie i wybory. W ten sposób, być może, uda się też ustrzec przed zrobieniem z ewaluacji jednego z technokratycznych instrumentów restryktywnie pojętej kontroli.

Ograniczone rozmiary tej książki spowodowały jednak, że nawet tak sprofilowana wizja ewaluacji nie mogła być przedstawiona w pełni, a wiele problemów rozpatrywanych początkowo jako zasadnicze nie znalazło w niej należytego miejsca. Takim zagadnieniem jest niewątpliwie genealogia, a następnie historia studiów i praktyk ewaluacyjnych. Pozwoliłyby one lepiej zrozumieć ich obecny status oraz istotne często różnice wynikające nie tylko z aktualnych potrzeb, ale i odmiennych tradycji, np. amerykańskiej lub brytyjskiej, przejawiające się między innymi w terminologii i sposobach konceptualizacji zadań.

Konsekwencje redukcji tej problematyki niweluje w znacznej mierze rozdział I, „Czym jest i dlaczego potrzebna jest ewaluacja", zwłaszcza tekst Helen Simons *Specyfika i rozwój studiów ewaluacyjnych*. Przedstawia on punkty zwrotne rozwoju jakościowo zorientowanej ewaluacji, jej dylematy i wyzwania. Szeroki kontekst spojrzenia na ewaluację, jaki proponuje Helen Simons, pogłębia Ernest R. House w tekście zatytułowanym *Ewaluator w społeczeństwie*. Taki dobór tekstów do rozdziału wprowadzającego wynika z chęci pokazania możliwie wielostronnych relacji między ewaluacją a jej instytucjonalnym i społecznym kontekstem, między celem a metodą ewaluacji. Warto także zauważyć, że jakiekolwiek sposoby oceny zawsze odwołują się do mniej lub bardziej ukrytego systemu wartości. Pozostaje jednak problemem otwartym, w znacznej

mierze określającym społeczną adekwatność ewaluacji, w jakim stopniu są one przedmiotem wspólnej akceptacji ewaluujących i ewaluowanych, zamawiających, finansujących i realizujących ewaluowany program.

Podobnym ograniczeniom ulec musiał także rozdział II zatytułowany „Elementy teorii". Jego intencją jest wprowadzenie w obręb refleksji teoretycznej w dwojakiej perspektywie: wewnętrznej, odnoszącej się do ewaluacji jako takiej, i zewnętrznej, pozwalającej rozpatrywać ewaluację jako czynnik praktyki społecznej i naukowo-badawczej.

Pierwszą z tych perspektyw reprezentuje tekst Davida Nevo *Konceptualizacja ewaluacji edukacyjnej. Analityczny przegląd literatury*. Tekst ten skonstruowany jest na planie prostych, ale fundamentalnych pytań, które znakomicie ułatwiają rozpoznanie struktury oraz potrzeb każdej systematycznie rozważanej ewaluacji.

Drugą perspektywę ukazuje tekst Helen Simons *Polityczne implikacje teorii ewaluacyjnych. Przybliżenie problemu*. Znaczenie tego tekstu polega na tym, iż ukazuje on rozliczne uwikłania, jak i konsekwencje teorii ewaluacyjnych, widzianych tym razem z pewnego dystansu. Umożliwia to krytyczną refleksję zarówno nad nie zawsze jawnymi społecznymi funkcjami tych teorii, ale także spojrzenie na ewaluację z punktu widzenia wielu innych perspektyw i nauk, np. epistemologii, socjologii, politologii, antropologii społecznej czy też nauk o komunikowaniu.

Następny rozdział, „Różnorodność podejść i ujęć", w pewnym przynajmniej zakresie pokazać ma wielorakie możliwości projektowania przedsięwzięć i strategii ewaluacyjnych, ich łączenia bądź też wzajemnej weryfikacji. Różnorodność tę przedstawia drugi tekst Ernesta House'a, *Główne podejścia*. Pod wpływem sugestii Helen Simons w tym właśnie rozdziale umieszczony został tekst Roberta E. Stake'a. Prezentuje on w sposób monograficzny historię oraz zasadnicze cechy coraz częściej praktykowanego zarówno w naukach społecznych, jak i w ewaluacji tak zwanego „studium przypadku". Stąd też właśnie jego tytuł *Studium przypadku*. Umieszczenie go w tym, a nie innym miejscu książki wynika z założenia, że szeroko pojęte studium przypadku jest w swej istocie nie tyle jedną z metod, ile raczej strategią lub też rodzajem ujęcia analizowanego problemu, które odwoływać się może do bardzo wielu standardowo rozumianych „metod" oraz technik badawczych, dając jednocześnie szansę na zbudowanie bardziej integralnego obrazu rzeczywistości.

Największym ograniczeniom uległa problematyka dwóch połączonych z konieczności w jednym rozdziale zagadnień, które ostatecznie tworzą blok zatytułowany „Projektowanie i metody badawcze". Naj-

ogólniejsze kwestie związane z pierwszym z tych problemów prezentuje Colin Robson w tekście *Projektowanie ewaluacji*. Samym metodom poświęcone jest studium Michaela Q. Pattona *Obserwacja — metoda badań terenowych* oraz tekst Herberta J. Rubina i Irene S. Rubin zatytuowany *Jak zmierzać do celu nie wiążąc sobie rąk. Projektowanie wywiadów jakościowych*. Ograniczenia te nie pozwalają szerzej omówić jednej z kluczowych faz procesu ewaluacyjnego, jaką jest właśnie procedura projektowania, ani też zasygnalizować wielości ewaluacyjnych metod. Na rzecz dokonanego wyboru przemawiają natomiast następujące argumenty:

1. Efektywna popularyzacja jakościowo zorientowanej ewaluacji wymaga wybiórczej choćby, ale konkretnej prezentacji jej metod.

2. Obserwacja analityczna oraz odpowiednio przygotowany i przeprowadzony wywiad to zasadnicze metody badań jakościowych; wkomponowane w metametodę, jaką jest studium przypadku, stanowią istotny element ewaluacji w edukacji i wbrew potocznym przekonaniom wymagają dużej wiedzy i systematyzacji.

3. Pomimo rosnącej wiedzy na temat metod jakościowych polski czytelnik w dalszym ciągu ma wiele trudności z dotarciem do fachowej literatury przedmiotu.

Ostatni rozdział zatytułowany „Społeczny kontekst ewaluacji" jest próbą pokazania, że ona sama, traktowana często jako instrument pozwalający budować sądy o konkretnych działaniach, jest także rodzajem zorganizowanego i celowego działania. Ewaluacja sama w sobie ulegać więc może tym samym wpływom co ewaluowana przez nią rzeczywistość, wspomagać autorytarne lub też demokratyczne procesy i struktury społeczne. Problemy te przedstawia tekst Ernesta R. House'a *Demokratyzacja ewaluacji* oraz artykuł Eleanor Chelimsky *Praktyka ewaluacyjna dzisiaj. Kilka refleksji*.

Na ostateczny kształt prezentowanej Państwu książki, jak i na sam fakt jej powstania miało wpływ wiele osób. Jest ona bowiem jednym z wielu rezultatów pionierskiego programu o nazwie „Training for Education Reform Management" (TERM) finansowanego z funduszy PHARE, a organizowanego i koordynowanego przez Biuro Programu TERM w Fundacji Rozwoju Systemu Edukacji, powołanej do życia przez Ministerstwo Edukacji Narodowej. Dyrektorem jednego z komponentów programu, odpowiedzialnym zarówno za jego bezpośrednią ewaluację, jak i wykreowanie polskiego środowiska ewaluatorów, została jedna z najbardziej znanych międzynarodowych ekspertek w dziedzinie ewaluacji, zwłaszcza edukacyjnej, autorka wielu cenionych prac, prof. Helen Simons z Uniwersytetu w Southampton. Do współpracy

zaproszono także polskich członków „personelu podstawowego": prof. Eugenię Potulicką i dra hab. Zbyszka Melosika z Uniwersytetu im. A. Mickiewicza w Poznaniu, dra Henryka Mizerka z Wyższej Szkoły Pedagogicznej w Olsztynie oraz piszącego te słowa z Uniwersytetu Warszawskiego. Niebawem w skład grupy polskich ewaluatorów weszły następne osoby: Beata Ciężka, Mirosława Cylkowska-Nowak, Aldona Hildebrandt, która wraz z Eugenią Potulicką i Zbyszkiem Melosikiem współpracowały z Helen Simons jeszcze przed uruchomieniem programu TERM, Jolanta Kłoczewska, Michał Kowalski, Beata Królewicz, Renata Sobolewska i Elżbieta Weyna. Zasadniczym celem tego zespołu była systematyczna ewaluacja programu TERM, ale równolegle rozpoczął on poszukiwanie, selekcję i dyskusję nad pracami, których przynajmniej fragmenty powinny się znaleźć w wydawnictwie przygotowanym dla czytelnika polskiego. Kolejne wizyty w instytucjach akademickich i ewaluacyjnych w Southampton, Norwich i Londynie pozwoliły zgromadzić przy pomocy brytyjskich współpracowniczek Helen Simons, Jany Hyatt, a zwłaszcza Barbary Zamorski, która zapoznała nas z pracami członków prawie legendarnego już ośrodka studiów ewaluacyjnych na Uniwersytecie Wschodniej Anglii w Norwich, a mianowicie Ośrodka Badań Stosowanych w Dziedzinie Edukacji (Centre for Applied Research in Education — CARE), pokaźny zbiór tekstów. Może być on podstawą dalszych publikacji, studiów i programów szkoleniowych.

Przygoda z ewaluacją, bo tak właśnie określić można doświadczenie wspomnianej wyżej grupy osób, prowadziła nie tylko do stopniowego odkrywania tajników sztuki ewaluacji, ale także ujawnienia kulturowych aspektów dynamiki nauczania i uczenia się tej profesji. Dlatego w początkowej wersji książki planowany był rozdział poświęcony międzykulturowym aspektom ewaluacji realizowanej w kontekście międzynarodowym, co z pewnością będzie coraz częściej spotykaną praktyką. Okazało się jednak, że jest to problem opisany jak do tej pory stosunkowo słabo, choć z pewnością godny zainteresowania zarówno ze strony badaczy, jak i praktyków. W efekcie, nie tylko na skutek braku miejsca, pozostawiliśmy to zagadnienie kolejnym, być może bardziej specjalistycznym publikacjom.

Na zakończenie chciałbym wyrazić szczególne słowa podziękowania osobom, bez których finalizacja wydania niniejszej antologii pozostałaby jedynie w sferze planów. Myślę więc o Helen Simons, której rozległa wiedza, doświadczenie i sugestie, jak również znajomość międzynarodowego środowiska ewaluatorów pozwoliły na ostateczny wybór tekstów, Renacie Sobolewskiej, której kompetencje i orientacja w realiach mię-

dzynarodowych programów i organizacji, a także zaangażowanie oraz zdolność motywowania w chwilach trudnych podnosiła efektywność pracy zespołu, wydawcy książki, Ewie Pajestce, która zdołała pokonać szereg problemów wydawniczych, kreując przy tym twórczą atmosferę współpracy, oraz pani Wandzie Lipnik za jej nieoceniony wkład w skomplikowane prace redakcyjne.

Leszek Korporowicz

Rozdział I

Czym jest i dlaczego potrzebna jest ewaluacja

Ernest R. House

Ewaluator w społeczeństwie

Ernest R. House, „The Evaluator in Society", rozdział z książki tegoż autora *Evaluating with Validity*, Sage, Beverly Hills — London 1980.

Oddzielanie dobrego od złego już od tak dawna jest przedmiotem ludzkiego zainteresowania, że wydaje się stanowić jeden z ważnych składników naszego życia zbiorowego. Wprawdzie rozmaite instytucje często obdarzane są autorytetem pozwalającym im na formułowanie ocen, jednakże dopiero od niedawna ewaluacja programów interwencji społecznych stała się na tyle sformalizowana, że można ją uznać za odrębny obszar ludzkiej działalności — a może nawet jej nową dziedzinę.

Podjęcie w ciągu ostatnich dwudziestu lat wielkich programów zabezpieczenia socjalnego zwiększyło znaczenie ewaluacji. W samych Stanach Zjednoczonych przeprowadza się dziesiątki tysięcy ewaluacji programów publicznych rocznie. Tysiące ludzi znajduje w nich zatrudnienie, a setki uniwersytetów i korporacji konkuruje ze sobą o ewaluacyjne kontrakty. Większość ewaluacji finansuje rząd federalny, resztę — władze stanowe lub lokalne. Ewaluacja programów publicznych stała się usankcjonowaną działalnością, a każdego roku wydaje się na nią setki milionów dolarów.

Zarówno wszechobecność, jak rola ewaluacji decydują o jej społecznym znaczeniu. Ewaluacja bowiem przekształca, legitymizuje albo dyskredytuje programy publiczne. Ogromna liczba przeprowadzanych ewaluacji sprawia, że nadzór nad nimi staje się trudny. Ewaluacje programów publicznych nie zawsze są wysokiej jakości. Zbyt często ewaluatorzy robią po prostu to, czego oczekują sponsorzy. Ewaluatorzy zbyt często błędnie pojmują istotę swojego zadania, wyrządzając przez to szkodę ocenianym przez siebie programom społecznym. Nie

jest oczywiście tak, że ewaluatorzy są mniej kompetentni albo bardziej ograniczeni niż np. przedstawiciele innych zawodów, ale w ich przypadku szkody, jakie mogą wyrządzić, są często rozleglejsze, mniej oczywiste i trwalsze.

Celem tej książki jest rozwój autorefleksji ewaluacyjnej. Wydaje się konieczne, aby ewaluacja była oparta na jakimś poczuciu odpowiedzialności moralnej, a praktyka ewaluacyjna — ukierunkowana przez refleksję nad jej sprawiedliwością, prawdziwością, a nawet pięknem. Kiedy w perspektywie mamy setki organizacji i tysiące ewaluatorów ubiegających się o poparcie kilku sponsorów, nieodzowna jest refleksja moralna. Ponadto jej przesłanki nie mogą być ani arbitralne, ani narzucane. Jeśli ma być ona podstawą uczciwej praktyki, to musi brać początek z racjonalnych analiz tego, czym jest ewaluacja i co ewaluatorzy uznają za rozsądne.

Ewaluacja prywatna i publiczna

Współczesna ewaluacja jest dzieckiem nowoczesności. Modernizacja była uwolnieniem od tradycji, przejściem od niekwestionowanej rzeczywistości danej właśnie przez tradycję do sytuacji społecznej, w której wszystko poddane być może weryfikacji i zmianie. Zamiast przyjmować to, co „jest zrządzeniem losu" zaczęto „wybierać" (Berger, 1974). We współczesnym świecie można wybierać więcej rzeczy. Przypuszczalnie daje to w rezultacie poprawę życia materialnego i głębsze spełnienie się jednostkowe. W praktyce jednak modernizacja doprowadziła do tego, że samo dokonanie wyboru stało się problemem. Ewaluacja przychodzi z pomocą w dokonywaniu wyboru.

Ostatnimi czasy dokonywanie formalnej ewaluacji na skalę masową, finansowanie ewaluacji przez rząd i rosnąca liczba agencji dokonujących takich ocen stworzyły problemy dotąd nie spotykane. Przyjrzyjmy się, jak przebiega proces indywidualnego wyboru. Pewna kobieta chce kupić nowe auto. Przede wszystkim odwiedzi sprzedawców samochodów oferujących auta, na które ją stać, i obejrzy dostępne. Po krótkim targu o cenę wybierze mały, najtańszy samochód. Kobieta ta lubi bowiem małe samochody, no i nie chce czekać dwa miesiące na fabryczną dostawę. Dla niej ważne są cena, wielkość i dostępność auta. Nie obchodzi ją zbytnio kolor czy wyposażenie. Jednostka dokonuje własnej ewaluacji (oceny) kierując się własną satysfakcją.

Teraz rozważmy sytuację, kiedy jest tyle możliwości wyboru, że jednostka nie może sobie dać z tym rady. Dostępnych jest tak wiele

samochodów, ze nie wiadomo od czego zacząć, kupująca zaś nie jest ekspertem w dziedzinie mechaniki samochodowej. Wtedy właśnie swoje usługi oferują takie grupy, jak Unia Konsumentów. Dokonają ewaluacji samochodów dla klientki i wskażą, który jest najlepszy. Klientka może więc uzyskać raport ewaluacyjny o samochodach. Tutaj ewaluator i podejmująca decyzję to dwie różne osoby. Są to więc sytuacje pod pewnymi ważnymi względami różne.

Jak powinien postępować ewaluator? Jeśli ewaluacja ma być przydatna, musi wybrać kryteria oceny (*standards*), które będą istotne dla klientki. Bez poznania klientki ewaluator nie będzie mógł odgadnąć jej osobistych preferencji. Ewaluator nie bierze pod uwagę mody, ponieważ w tym zakresie klientka dokonać może wyboru bez niczyjej pomocy. Dostępność jest kwestią lokalną. Ewaluator pomija ją również, chyba że traktuje ją jako czynnik ograniczający. Z pewnością każdy będzie zainteresowany ceną zakupu. Ponadto ewaluator jako istotne kryterium ewaluacyjne przyjmuje zużycie paliwa itp. Pomija jako kryterium maksymalną prędkość.

Ewaluator ma inne ważne decyzje do podjęcia. Które samochody ma oceniać? Kiedy konsumentka oceniała na własny użytek, określała rodzaj samochodów, wśród których mogłaby wybrać. Dla ewaluatora nie jest to problem bez znaczenia, gdyż rezultaty będą zasadniczo różne w zależności od tego, które samochody uwzględni porównując je między sobą. Musi on wybrać pewien zestaw do oceny. Do tego zestawu samochodów odniesie przyjęte przez siebie kryteria. Odpowiednio do tego, jak spełniają te kryteria, sklasyfikuje samochody jako dobre lub złe albo uszereguje je od najlepszego do najgorszego. Na koniec sporządzi raport.

Teraz konsumentka ma jeszcze jeden wybór: musi zadecydować, czy ta ocena jest dla niej do przyjęcia. Zależy to od kilku spraw. Czy te właśnie samochody są jedynymi, spośród których chciałaby wybierać? Czy te kryteria są ważne jako przesłanka jej decyzji? Czy ewaluator wydając ostateczną opinię wyważył i zbilansował kryteria tak, jak uczyniłaby to ona sama? A zasadniczo, czy ocena jest prawdziwa? Czy można mieć do niej zaufanie? Czy sam ewaluator był wiarygodny? Czy nie opłacił go któryś z koncernów samochodowych? Czy ocena była rzetelna? Konsumentka jest wciąż wolna i może zaakceptować lub odrzucić wyniki i rady wynikłe z ewaluacji. Chociaż ewaluacja stała się społecznym procesem włączającym więź między ewaluatorem i konsumentką, to ostateczny wybór należy do niej.

Rozważmy trzecią sytuację, kiedy rząd opłaca ewaluatora mającego oceniać program działań publicznych. Dla kogo w tym przypadku jest przeprowadzana ocena? Dla rządu? Dla narodu? Społeczeństwa? Dla odbiorców tego programu? Dla jego administratora? Odpowiedź nie jest prosta, bo ewaluacja ta przez swój wynik oddziała na nich wszystkich. Nie ma tu już jednostki, jak nabywczyni samochodu, ku której ewaluacja mogłaby być skierowana. Zainteresowanych jest wielu, ponieważ decyzja dotycząca programu, a oparta na ewaluacji, będzie miała wpływ na każdego. Ewaluacja jest zasadniczą i nierozerwalną częścią sytuacji społecznej, jest decyzją zbiorową.

Inaczej niż w kontekście prywatnym, niezadowolony konsument nie może po prostu wycofać się, nawet jeśli ewaluacja okaże się dla niego nie do przyjęcia, ponieważ i tak wywrze ona określony wpływ. Publiczny charakter tej sytuacji narzuca pewne ograniczenia ewaluacji. Wszystkie problemy związane z wyborem właściwych kryteriów i selekcją odpowiednich porównań są niejednokrotnie wyolbrzymiane. Ewaluatorzy stworzyli różnorakie sposoby radzenia sobie z tym problemem, ewaluator bowiem musi wybrać. [...]

Proces ewaluacji

Najkrócej mówiąc: ewaluacja prowadzi do powzięcia opinii, że coś przedstawia się tak a tak. Nie musi prowadzić do decyzji w kwestii sposobu działania, chociaż dziś często o to właśnie chodzi. Można ocenić Cezara jako wielkiego wodza, nie podejmując w wyniku tej oceny żadnych decyzji. Ewaluacja prowadzi do wyrobienia sobie sądu o wartości jakiegoś zjawiska, działania lub rzeczy.

Ten sąd jest często wynikiem klasyfikacji albo uszeregowania czegoś według stopnia, w jakim obiekt ten spełnia wymogi zestawu standardów lub kryteriów. Ewaluacja jest z natury porównawcza. Oznacza to, że musi istnieć jakiś zestaw kryteriów i jakaś klasa, z którą obiekt jest porównywany. Jeśli obiekt klasyfikuje się jako „dobry" albo „zły", to porównuje się go albo z ogółem obiektów danej klasy, albo z przeciętnym obiektem tejże. Jeśli zostaje zakwalifikowany jako „najlepszy" albo „najgorszy", to porównuje się go bezpośrednio z określonym podzbiorem obiektów podobnych, np. samochód z czterema innymi samochodami w tej samej cenie. Samochody zostają uszeregowane według danych kryteriów w porządku wskazującym miejsce każdego z nich w porównaniu z pozostałymi czterema, z czego nie musi wynikać, jaką wartość przedstawiają te samochody jako zbiór. Dobór klasy porów-

nawczej może spowodować drastyczne różnice w ocenie, nawet przy stosowaniu tych samych kryteriów (Taylor, 1961).

Tak więc ewaluacja z istoty wymaga obrania zestawu kryteriów, ich zdefiniowania, skonkretyzowania klasy porównawczej, a następnie ustalenia stopnia, w jakim dany obiekt odpowiada kryteriom. Wszystkie te kroki powinny umożliwić ewaluatorowi wyrobienie sobie sądu o wartości ocenianego obiektu (Taylor, 1961). Jednak nie wszystkie ewaluacje są tak zracjonalizowane. Wiele z nich [por. niżej *Studium przypadku*] wypracowuje kryteria, porównania i sądy bardziej pośrednio i intuicyjnie. Inne dążą raczej do rozumienia niż sądu.

Sąd wartościujący wyrażany jest często za pomocą takich określeń, jak „dobry" lub „zły" albo w innym wartościująco obciążonym języku. Do wyrażenia sądu wartościującego w danym kontekście można użyć codziennych słów opisowych. Wartościujący komunikat mogą nieść przykłady i anegdoty. Sąd wartościujący zasadza się raczej na sposobie użycia słów niż na ich doborze.

Sądy wartościujące będące rezultatem ewaluacji nie są równoważne z dyrektywami. Ewaluator nie jest w mocy kierować czyimś postępowaniem. Nie każdy też sąd wartościujący poparty jest znajomością rzeczy. Tym niemniej we współczesnej praktyce ewaluacyjnej ewaluatora często prosi się o pomoc w podjęciu decyzji osobie, która próbuje dokonać wyboru co do zasadniczego kierunku działania. Możliwe kierunki działania stają się obiektami oceny. Konkluzjami sądów wartościujących mogą być rekomendacje, ale i wtedy człowiek podejmujący decyzje jest na tyle wolny, że może odrzucić radę ewaluatora.

W najprostszych przypadkach ewaluacji, kiedy ewaluator i podejmujący decyzje są jedną i tą samą osobą, dobór klasy odniesień i przyjęcie zestawu kryteriów przebiega łatwiej. Wciąż jednak rzeczywisty proces ewaluacyjny jest bardziej skomplikowany niżby się mogło wydawać. Ludzie oceniają odwołując się do ustalonych kryteriów, ale również konfrontując swoje sądy z sytuacjami, w których nie mają wątpliwości co do słuszności swoich ocen. W rzeczywistości rozumowanie ewaluacyjne jest raczej procesem, którym rządzi dialektyczna relacja między abstrakcyjnymi zasadami i konkretnymi przykładami, niż prostym wnioskowaniem w oparciu o kryteria (Barry, 1965).

Co więcej, w rzeczywistej ewaluacji kryteria mogą kolidować ze sobą. Ktoś pragnie wygodnego, dużego samochodu, ale jednocześnie ma to być samochód zużywający mało paliwa. Nie można tu, niestety, maksymalizować obu pożądanych cech równocześnie. W takich przypadkach osoba zainteresowana jest skłonna relatywizować oba kryteria wzglę-

dem siebie: ilość miejsca i zużycie paliwa. Gotowa jest raczej przyznać względne pierwszeństwo któremuś z kolidujących ze sobą kryteriów, intuicyjnie wyważając ich walory rynkowe, niż uznać bezwzględną wyższość jednego kryterium nad drugim. Z jakiej ilości miejsca potencjalny klient chciałby zrezygnować na korzyść mniejszego zużycia paliwa? Różne osoby różnie odpowiedzą na to pytanie. Ewaluacja jest procesem złożonym także wtedy, gdy wszystkie jego komponenty pozostają w gestii jednej osoby.

Kiedy ewaluator i podejmujący decyzję są dwiema różnymi osobami, dochodzi jeszcze jeden wymiar: wiarygodność ewaluatora. Co więcej, ewaluacja musi być zakomunikowana osobie podejmującej decyzję. Występuje tu większa niepewność, co powinno być klasą porównawczą, jakie powinny być kryteria i jak należy je porównywać między sobą. Podejmujący decyzję musi trapić się nie tylko tym, czy fakty są prawidłowe, ale i tym, czy ewaluator przedstawia je prawdziwie. W nowoczesnych liberalnych społeczeństwach wciąż traktuje się tę relację jako działanie o charakterze prywatnym, ale interpersonalnym.

Na koniec trzecia sytuacja, kiedy przedmiotem oceny jest program publiczny. Tu ewaluator musi troszczyć się o to, czy ewaluacja odpowiada normom przewidzianym w procedurach podejmowania decyzji w interesie publicznym. Ewaluacja staje się interesem nie tylko interpersonalnym, ale także kolektywnym. W tym przypadku występuje szersza wspólnota istotnych interesów, które trzeba uszanować. Jest to kłopot, z którym najtrudniej uporać się współczesnemu ewaluatorowi. Ocena musi być nie tylko prawdziwa i wiarygodna ; musi być sprawiedliwa.

Kłopot ten ewaluatorzy rozwiązują na wiele sposobów. Rozwiązania te wynikają z ich filozoficznych, zawodowych i osobistych przekonań. Znaczący wpływ ma również wyszkolenie i nabyta praktyka. Główne podejścia do ewaluacji są przedstawione w rozdziale następnym, a łączą się one już z systemem przekonań [prezentujemy go niżej w rozdziale zatytułowanym „Różnorodność podejść i ujęć"].

tłumaczyła *Monika Koniecko*

Helen Simons

Specyfika i rozwój studiów ewaluacyjnych

Helen Simons, „The Nature and Development of Evaluative Inquiry", rozdział z książki tejże autorki *Getting to Know Schools in a Democracy*, The Falmer Press, London – New York – Philadelphia 1987.

„Oddzielanie dobrego od złego już od tak dawna jest przedmiotem ludzkiego zainteresowania, że wydaje się stanowić jeden z ważnych składników naszego życia zbiorowego. Wprawdzie rozmaite instytucje często obdarzane są autorytetem pozwalającym im na formułowanie ocen, jednakże dopiero od niedawna ewaluacja programów interwencji społecznych stała się na tyle sformalizowana, że można ją uznać za odrębny obszar ludzkiej działalności — a może nawet jej nową dziedzinę" (House, 1980, s. 15).

Bez względu na to, czy studia ewaluacyjne powołują do życia nową dyscyplinę — jak sugeruje House — czy też nie, z całą pewnością w ciągu minionego ćwierćwiecza stały się one wyodrębnionym obszarem działalności, bogatym w swoiste podejścia analityczne, modele, we własne kryteria, zasady i wzorce. Dziedzina ta rozwijała się bardzo szybko, początkowo z uwagi na potrzebę obserwowania rezultatów dużych programów socjalnych, ostatnio zaś z uwagi na potrzebę uzasadniania decyzji podejmowanych na różnych szczeblach systemów edukacyjnych.

Wprawdzie studia ewaluacyjne ciągle jeszcze są w stadium rozwoju, ale zgromadzony z biegiem czasu zasób uzgodnień pozwala na wskazanie przynajmniej kilku teoretycznych własności ewaluacji edukacyjnych. Studia ewaluacyjne to działalność praktyczna (Stake, 1980; Glass, 1972), nastawiona na uchwycenie odrębności tego, co ewaluowane (Stake, 1967b; Guba, Lincoln, 1981), czasem o charakterze politycznym (House, 1973a; MacDonald, 1974; Weiss, 1975). Już w stosunkowo wczesnej fazie

rozwoju studiów ewaluacyjnych podkreślano, że ich podstawową funkcją jest reagowanie na potrzeby klientów (Cronbach, 1963), za których początkowo uznawano osoby czy zespoły pracujące nad doskonaleniem programów edukacyjnych. Funkcję tę określano jako formatywna (Scriven, 1967), ukazującą kierunki doskonalenia programów edukacyjnych już w toku ich opracowywania. Scriven sugerował oddzielenie tej „formatywnej" funkcji studiów oceniających od funkcji „konkluzywnej", polegającej na wspieraniu wyboru pomiędzy programami alternatywnymi. W miarę jak w latach sześćdziesiątych studia ewaluacyjne zyskiwały coraz większą popularność, niewspółmierną w stosunku do wolno narastającej świadomości ich znaczenia politycznego, teoretycy zaczęli stawać przed kłopotliwym pytaniem, komu właściwie mają one służyć. Stufflebeam (i współaut., 1971) utrzymywał, że za podstawową kategorię klientów należy uznać urzędników administrujących oświatą, natomiast Stake usiłował poszerzyć ją o rozmaite kategorie wyborców (Stake, 1967a). W naszym kraju MacDonald najpierw podkreślał rolę studiów ewaluacyjnych we wspomaganiu decyzji podejmowanych przez lokalnych „nabywców" programów edukacyjnych (MacDonald, 1970), później zaś wskazywał, że pozwalają one na ich szerszą wymianę między zainteresowanymi stronami (MacDonald, 1974). Narastająca świadomość tego, że wyniki studiów ewaluacyjnych mogą mieć wpływ na podział środków finansowych, a także organizacyjne i personalne konsekwencje dla osób zaangażowanych w programy poddawane ocenie, powodowała także, że zaczęto dostrzegać nowe problemy etyczne (House, 1980; Adelman, 1984).

House zwrócił uwagę, że studia oceniające „stanowią istotny element publicznych relacji między ludźmi, a także służą doskonaleniu zbiorowo podejmowanych decyzji" (1980, s. 18). Później teza ta uzyskała wsparcie wielu innych ewaluatorów, którzy kładli nacisk na wpływ procesów ewaluacji na ogólną politykę zmian społecznych. Opracowana przez House'a koncepcja studiów ewaluacyjnych traktowanych jako procesy perswazji pozwoliła również na bardziej przejrzyste określenie charakteru wiedzy powstającej w toku tych studiów (tamże, s. 72). Wreszcie ostatnio dostrzeżona potrzeba efektywnego komunikowania wyników studiów ewaluacyjnych doprowadziła do uznania, że mają one same przez się wartości edukacyjne (Cronbach, 1982).

Tak więc, odwołując się do opinii czołowych teoretyków studiów ewaluacyjnych, możemy charakteryzować je jako praktyczne, konkretne, polityczne, perswazyjne i edukacyjne działania zmierzające do zaspokojenia oczekiwań określonej klienteli. Ta wyrwana z kontekstu

charakterystyka niewiele nam powie o odrębności studiów ewaluacyjnych jako szczególnego obszaru dociekań badawczych. Potrzeba tu czegoś więcej — trzeba zrozumieć powody, dla których poszczególne cechy ewaluacji zostały uznane za warte podkreślenia: jakie argumenty chciało się przy tej okazji podważyć, jakie skutki dla praktyki studiów ewaluacyjnych brali pod uwagę ci, którzy kładli nacisk na ich właściwości. Spróbuję więc osadzić w stosownym kontekście większość problemów, jakie pojawiały się w toku rozwoju teorii ewaluacji, biorąc zarazem pod uwagę te zagadnienia, z którymi zetknęłam się w toku własnej praktyki, i to, co interesuje mnie najbardziej, to jest sposób, w jaki studia ewaluacyjne mogą być wsparciem dla profesjonalistów usiłujących doskonalić system szkolny. Skoncentrują się zatem na trzech cechach studiów ewaluacyjnych, tj. na konkretności i perswazyjności ewaluacji oraz na ich politycznym kontekście. Sądzę, że łącznie tworzą one ramy pozwalające na uogólnienie mojego praktycznego doświadczenia.

KONKRET

Ewaluacja nastawiona na konkret

Większość opisów studiów ewaluacyjnych kładzie obecnie nacisk na to, że — niezależnie od tego, czy są to nowe programy szkolne czy interwencje społeczne o szerszym zasięgu — ewaluacja koncentruje się na *konkretnych* projektach, które swoim wpływem obejmują takie struktury społeczne, jak klasy szkolne, szkoły czy regiony kraju. Ekstrapolację lub uogólnienie rezultatów tych studiów wykraczające poza ich konkretny przedmiot można traktować jako zagadnienie o wtórnym znaczeniu i kwestię odpowiedzialności, która obciąża bardziej ich odbiorców, mniej zaś samych ewaluatorów. Ten nacisk na *konkretny* charakter studiów ewaluacyjnych oznacza istotną zmianę w myśleniu o ewaluacjach programów interwencji społecznych. Postaram się zilustrować tę tezę, odwołując się do ich niedługiej historii.

Należy pamiętać o tym, że studia oceniające traktowane jako działalność specjalistyczna, wyodrębniona najwyraźniej w Stanach Zjednoczonych w okresie szeroko zakrojonych interwencji społecznych podejmowanych przez władze federalne (por. Atkin, House, 1981), stanowiły pierwotnie domenę psychologów eksperymentalnych posługujących się koncepcją hipotezy zerowej. Nie trzeba było zbytnio modyfikować tej tradycji, aby zmiany programów szkolnych zinterpretować jako zróżnicowanie bodźców mające na celu wywołanie nowych zachowań uczniów. Nowe programy szkolne były przygotowywane tak, aby

stosunkowo łatwo można było mierzyć ich efekty (Hamilton i współ-aut., 1977), a psychometrycy-ewaluatorzy skwapliwie się tym zajmowali wspierani przez dominujące w tamtym okresie, technologiczne podejście do zagadnień interwencji społecznych.

Sceptycyzm wobec tego podejścia do rozwoju programów szkolnych po raz pierwszy wyraził Myron Atkin (1963), ale Stake pierwszy zasugerował konieczność ponownego przemyślenia charakteru całego przedsięwzięcia. Gdy w swoim wstępie do pierwszego numeru „AERA Monograph Series on Curriculum Evaluation" (Stake, 1967b) zaakcentował on potrzebę przedstawiania przez ewaluatorów swoich „opowieści" i podkreślał, że ich klienci powinni być świadomi tego, jaki „towar" nabywają, stało się to bodźcem do ponownego przemyślenia celów stawianych ewaluacji i osadzenia jej w bardziej idiograficznej tradycji badawczej oraz do odejścia od takiej koncepcji rozwoju programów szkolnych, która sprowadzała się do mało przemyślanych eksperymentów nad procesami uczenia się. Nacisk na narratywny aspekt ewaluacji oznaczał dostrzeżenie tego, że klienci powinni coś więcej wiedzieć o cieszących się znacznym prestiżem innowacjach edukacyjnych aniżeli to, że sprostały testom hipotezy zerowej. Apel Stake'a nie wpłynął na amerykańskich psychologów behavioralnych, ale zyskał życzliwy odzew po drugiej stronie Atlantyku. W tym samym czasie w Anglii MacDonald, Hamilton i Parlett usiłowali zgłębić sens i opisać zmieniającą się rolę studiów ewaluacyjnych. Powstanie anglo-amerykańskiego sojuszu dysydentów stanowiło reakcję na kłopoty, jakie wiązały się z nową konceptualizacją ewaluacji.

W końcu lat sześćdziesiątych, gdy Stake apelował o zwrócenie większej uwagi na specyfikę i złożoność zmian w programach nauczania, po obu stronach Atlantyku narastała świadomość tego, że realizacja tych zmian wymaga czegoś więcej aniżeli prostej inżynierii społecznej wspartej wielkimi pieniędzmi i zaangażowaniem poważnego potencjału akademickiej nauki. Towarzyszyło temu zainteresowanie i pewne poparcie ze strony inicjatorów i fundatorów zmian programowych w szkolnictwie: ci nieliczni ewaluatorzy, którzy odstępowali od tradycyjnego, behavioralnego podejścia, skoncentrowanego na pomiarze reakcji na zestaw określonych bodźców edukacyjnych, mogli przedstawić bardziej przekonujące argumenty wyjaśniające powody, dla których nie udaje się osiągnąć założonych celów projektów zmian w szkolnictwie. W końcu lat sześćdziesiątych Stake skłaniał się ku takiemu rozumieniu ewaluacji, które kładło nacisk na to, że służy ona „obrazowaniu" procesu zmian (Stake, 1977). MacDonald (1970) posługiwał się studiami przypadków

służących jako przewodnik dla zainteresowanej publiczności, Smith zwracał uwagę na konieczność zwracania uwagi na kontekst wdrażania innowacji edukacyjnych (Smith, Keith, 1971). Wspólną podstawą zmian w sposobach konceptualizacji studiów ewaluacyjnych (choć powinniśmy pamiętać o tym, że zmiany te dokonywały się raczej równolegle, nie zaś w określonej kolejności) stało się stosunkowo wczesne określenie ich celu przez Stake'a:

„Cel edukacyjnych studiów ewaluacyjnych ma charakter prezentacji: idzie o zapoznanie adresatów z pracą nauczycieli i ich uczniów. Różni się od celu typowych badań nad funkcjonowaniem systemów edukacyjnych tym, że studia oceniające są skoncentrowane na *konkretnych* przedsięwzięciach, w mniejszej zaś mierze zwracają uwagę na zmienne wspólne wielu przedsięwzięciom edukacyjnym. Pełne studium oceniające przynosi wynik w postaci «opowieści», czasem wspartej danymi o charakterze statystycznym. Opowiada ono o tym, co się wydarzyło. Ujawnia punkty widzenia i przekonania rozmaitych grup i osób w sposób, mam nadzieję, obiektywny. Opowiada o zaletach i brakach. Dodatkową korzyścią, jaką przynosi pełne studium ewaluacyjne, jest to, że jego wyniki poddają się uogólnieniu («Z opowieści tej wynika...»), mogą służyć usprawnieniu następnych programów zmian edukacyjnych" (Stake, 1967b, s. 5).

Wprawdzie ewaluatorów-dysydentów wspierali ci, którzy byli coraz bardziej niezadowoleni z korzyści niewspółmiernych do znacznych nakładów na zmiany w programach szkolnych, ale „naukową" legitymizację ruch ten uzyskał dopiero w 1975 r. w sposób zgoła nieoczekiwany. Właśnie wtedy Lee Cronbach, jedna ze znaczących postaci psychologii wychowawczej, opublikował artykuł, w którym utrzymywał, że psychologia eksperymentalna posługująca się wyrafinowanymi metodami statystycznymi jest niezdolna wnieść wkład w tworzenie wiedzy potrzebnej do realizacji zmian edukacyjnych. Sugerując potrzebę studiów opisowych Cronbach pisał:

„Powinniśmy zastanowić się nad zmianą naszych priorytetów badawczych, które dotychczas polegały na poszukiwaniu uogólnień. Sugerowałbym odwrócenie naszych praktyk. Obserwator zbierający dane w konkretnej sytuacji, dostrzegając określone reakcje w specyficznym kontekście, jest w stanie sformułować ocenę lub hipotezę dotyczącą tejże sytuacji. [...] Wraz z przechodzeniem od jednej sytuacji do kolejnej jego podstawowym celem staje się opisanie reakcji w zmieniających się okolicznościach — nie jest wykluczone, że dostrzeże on przy tym to, co stanowi o unikalności kotekstów. [...] W miarę kumulowania się danych

CASE STUDIES

tego rodzaju możliwe stanie się zrozumienie tego, jak niekontrolowane okoliczności mogą powodować brak oczekiwanych rezultatów. Oznacza to m.in., że uogólnienia stanowią końcowy efekt procesu badawczego, a wyjątki należy traktować równie poważnie jak zawarte w tych uogólnieniach prawidłowości" (Cronbach, 1975, s. 124–125).

Cronbach opisywał wprawdzie schemat studium ewaluacyjnego MacDonalda opartego na analizie przypadku zmiany społecznej realizowanej w odmiennych kontekstach, a dotyczącej nauczania wspomaganego komputerem, tj. projektu realizowanego wtedy od blisko dwóch lat w Anglii (MacDonald i współaut., 1975), ale miał na uwadze całą wyspiarską tradycję psychologii stosowanej, jej koncentrację na specyficznych zmiennych oraz na formułowaniu teorii zakładających ich uniwersalistyczne zastosowanie.

Sceptyczne uwagi Cronbacha dotyczące uogólnień wyników badań edukacyjnych odzwierciedlały nacisk, jaki kładli ewaluatorzy na konkretny charakter studiów ewaluacyjnych, i ich opór wobec tezy mówiącej, że przynajmniej od pewnego poziomu w hierarchicznych strukturach zarządzania systemami edukacyjnymi podejmowanie decyzji musi odwoływać się właśnie do uogólnień wyników dotyczących poszczególnych, konkretnych przypadków. Można było przy tym utrzymywać — jak robił to np. Stake formułując swoją koncepcję „uogólnień naturalistycznych" (Stake, 1980) — że wielu ważnych adresatów tych wyników może, nieco na własną odpowiedzialność, odnosić wyniki określonych studiów ewaluacyjnych do własnych problemów, zwracając uwagę na to, co je łączy z konkretnym, wyraźnie określonym kontekstem. Jednakże bez rozbicia całej struktury wpływów, w ramach której podejmowane były decyzje dotyczące programów szkolnych, oczekiwania administratorów decydujących o wielu tego typu przypadkach musiały — w ramach tej koncepcji — pozostawać nie spełnione. Stake wzbogacił swój oryginalny pomysł o koncepcję „raportu wykonawczego" (Stake, Easley, 1978), MacDonald dodał doń zalecenie formułowania uogólnień dla administratorów oświaty (MacDonald, Jenkins, 1980), ale były to jedynie próby rozszerzenia zastosowań metody opartej na koncepcji studiów przypadku.

Zastanawiając się nad użytecznością studium przypadku jako szczególnej techniki badawczej, Stake potwierdził swoje przekonanie, że właściwą drogą rozwoju analiz ewaluacyjnych jest zachowanie ich konkretności, nadając im zarazem nieco inny sens, to jest zarysowując granice między klasycznymi badaniami o charakterze akademickim, nastawionymi na wiedzę formalną, a studiami ewaluacyjnymi nasta-

wionymi na wiedzę praktyczną. Ta druga miała odwoływać się do „osobistej wiedzy" (*tacit knowledge*) (Polanyi, 1962) zainteresowanej publiczności i jej zdolności do rozumienia dzięki nabytemu uprzednio doświadczeniu (Stake, 1980). W gruncie rzeczy zupełnie świadomie i — jak sam mówił — nieco „na wyrost" podkreślał on odmienność badań akademickich i praktycznych, „dostrzegając w tym nadzieję na doskonalenie i legitymizację takich form prac badawczych, które przynoszą rozumienie i wspierają rozwiązywanie praktycznych problemów, ale niewiele wnoszą do tworzenia dających się obronić uogólnień" (tamże, s. 7). Stake usilnie dążył do poszerzenia tradycyjnej koncepcji ewaluacji tak, aby wykroczyła ona poza ramy wyznaczane przez społeczność naukową. Doprowadziło to do konfrontacji, najpierw z psychologami ze szkoły behawioralnej, potem zaś z etnografami, którzy również zainteresowali się problematyką ewaluacji w systemach oświatowych i coraz częściej wkraczali w stworzoną przezeń przestrzeń badawczą. MacDonald (1982b) odnotował spięcie, do jakiego doszło w trakcie kongresu AERA w 1981 r. między Harrym Wolcottem, który mówił o potrzebie ratowania etnografów przed agresją pseudobadaczy, a Stake'em, który zarzucił temu pierwszemu operowanie frazesami godnymi działacza związkowego[1].

Wprawdzie badacze zajmujący się ewaluacją różnie zapatrują się na ewentualną powinność wnoszenia wkładu w ogólne teorie zmian edukacyjnych czy — bardziej ogólnie — w teorie społeczne, to obecnie w większości są zgodni co do tego, że ich podstawowym zadaniem jest wspieranie konkretnych, osadzonych w określonym miejscu i czasie projektów zmian, programów edukacyjnych bądź kierunków polityki oświatowej. Mówiąc nieco inaczej, studia ewaluacyjne uznają oni za studia poświęcone jednoznacznie określonym obiektom, nie zaś ich mniej czy bardziej ogólnym klasom.

Innym sposobem ujęcia problemów, przed jakimi stały studia ewaluacyjne usiłując określić swoją tożsamość, jest osadzenie dyskusji dotyczącej tych zagadnień w kontekście sporów o „scjentyzm" — to jest, jak określił to Robert Nisbet, „nauki wyzbytej ducha twórczości i odkryć" (Nisbet, 1976, s. 4). Wprawdzie Nisbet atakował pewną tradycję badań socjologicznych, ale to, co miał w tej sprawie do powiedzenia, ma znaczenie dla badań edukacyjnych i ewaluacji, zwłaszcza że znaczną część pierwszego pokolenia ewaluatorów interwencji społecznych

[1] Harry Wolcott przedstawił na kongresie American Educational Research Association w 1981 r. referat o prowokacyjnym tytule *How to look like an anthropologist without really trying* (Jak bez większych wysiłków udawać antropologa).

w systemach edukacyjnych stanowili badacze mocno osadzeni w tradycji zdominowanej przez dążenie do naśladowania nauk przyrodniczych.

Oczywiście nie wszystkie badania edukacyjne mają „scjentystyczny" charakter, ale te formy tych badań, które niewolniczo naśladują to, co zwykło się nazywać „metodą naukową", i kładą nacisk na analizy statystyczne, testowanie hipotez i konstruowanie teorii narażają się łatwo na takie określenie. Analizując związki między sztuką i nauką Nisbet wskazywał na to, iż problemy nauk społecznych biorą się z ich niezdolności do rozróżnienia *logiki odkrycia* i *logiki prezentacji*:

„Ta druga podlega regułom i zaleceniom; pierwsza natomiast jest od nich wolna. Jednakże najcięższym grzechem przeciwko Muzom jest założenie, a przynajmniej mocna teza zawarta w podręcznikach metodologii i budowy teorii, że (ciągle żywotna) logika odkrycia może w jakiś sposób zostać zmobilizowana przez respektowanie reguł logiki prezentacji. To nieporozumienie może prowadzić jedynie do intelektualnej pustki i wyjałowienia" (tamże, s. 5).

„Teoria ewaluacji w ciągu ostatniego ćwierćwiecza kładła coraz wiekszy nacisk na logikę odkrycia, można twierdzić, że dokonywało się to kosztem respektu dla logiki prezentacji. Na obszarze studiów oceniających rozróżnienie proponowane przez Nisbeta znalazło wyraz w podziale na informacje umożliwiające rozumienie (*vindicative*) i dane o charakterze wskaźnikowym (*indicative*), przy czym te pierwsze dotyczą zmian już zrealizowanych, drugie zaś mają służyć jako materiał wspierający podejmowanie decyzji" (MacDonald, 1982b, s. 14).

Polityczne uwikłanie ewaluacji

„Komponentem ewaluacji są jakieś zapatrywania społeczne. Ludzie różnią się w swoich ocenach, ponieważ wyznają odmienne poglądy w takich kwestiach jak to, czym społeczeństwo jest i czym być powinno. Wiele sporów o wyniki ewaluacji to spory ideologiczne, którym nadaje się pozór sporów o techniki analityczne" (Hamilton i współaut., 1977, s. 25).

Dość powszechnie uznaje się obecnie, że ewaluacja ma swoje aspekty polityczne (House, 1973a; 1980; Apple, 1974; MacDonald, 1974; Weiss, 1975). Nie mówi się jednak o tym głośno, a doniosłość tego faktu jest ciągle kwestionowana. Mimo to tylko niewielu zainteresowanych studiami ewaluacyjnymi badaczy odrzuciłoby następującą opinię Cronbacha: „[...] teoria studiów ewaluacyjnych musi być teorią interakcji

politycznych w tej samej mierze, w jakiej jest teorią konstruowania wiedzy" (Cronbach i współaut., 1980, s. 52).

Wprowadzając pojęcie „kontekst wartościowania", House był jednym z pierwszych amerykańskich ewaluatorów, którzy dostrzegli znaczenie politycznego uwikłania studiów oceniających: „Wbrew powszechnemu przekonaniu wyniki ewaluacji nie są ostatecznym werdyktem, charakteryzującym się absolutną obiektywnością, akceptowanym jako orzeczenie nieodwołalne. Wyniki ewaluacji zawsze opierają się na jakichś przeświadczeniach, zawsze są w jakimś stopniu stronnicze. [...] Czy wyniki te zostaną zaakceptowane, zależy w dużej mierze od takich okoliczności jak to, czy wspierają one bądź stanowią zagrożenie dla tych, którzy je otrzymują. Ewaluacje są integralnym składnikiem procesu politycznego w naszym społeczeństwie" (House, 1973a, s. 3).

Weiss (1975), autorka dobrze znana ze swego wkładu w badania o orientacji praktycznej, prezentuje podobny pogląd, opisując ewaluację jako „racjonalne przedsięwzięcie realizowane w kontekście politycznym" (s. 13). Jej zdaniem, względy polityczne interweniują w procedury ewaluacyjne na trzy sposoby. Po pierwsze, przedmioty studiów ewaluacyjnych są politycznie wykreowanymi elementami rzeczywistości, są one „proponowane, definiowane, dyskutowane, akceptowane i finansowane w toku procesu politycznego". Po drugie, ponieważ studia ewaluacyjne są prowadzone *po to*, aby ich wyniki zostały włączone w mechanizmy podejmowania decyzji, raporty ewaluacyjne wkraczają więc na scenę polityczną, „współzawodniczą z innymi czynnikami, które w sumie składają się na proces polityczny". Wreszcie, po trzecie, wyniki ewaluacji mają same przez się polityczną doniosłość, choć zazwyczaj się tego nie zauważa: „Ze względu na swój charakter, wyniki ewaluacji zawierają opinie o charakterze politycznym w takich kwestiach, jak problematyczność jednych programów interwencji społecznych i pożytek płynący z innych, prawomocność celów tych programów i/lub strategii ich realizacji, ich udziału w realizacji cząstkowych reform, a nawet opinie o roli badaczy społecznych w formuowaniu tych programów" (Weiss, 1975, s. 13–14).

Jednakże to MacDonald był tym teoretykiem ewaluacji, który nadał absolutne pierwszeństwo kwestiom ich politycznego charakteru (MacDonald, 1974), odwołując się do orientacji politycznych w budowie typologii podejść do zagadnień ewaluacji. Proponując swoją klasyfikację pisał on: „Ewaluatorzy nie tylko żyją w realnym świecie politycznych sporów dotyczących systemów szkolnych, zmieniają oni także zastane stosunki władzy. Ich praca polega na gromadzeniu informacji, które

funkcjonują jako zasoby w promowaniu określonych interesów i wartości" (s. 16).

Po części zdając sobie z tego sprawę, po części odwołując się do swoich doświadczeń związanych z realizacją wielu studiów ewaluacyjnych, MacDonald usiłował starannie oddzielać je od typowo akademickich prac badawczych. Te pierwsze miały się różnić od drugich funkcjami, a także nieuchronnym uwikłaniem ewaluatorów w procesy polityczne. Aby podkreślić to rozróżnienie, MacDonald posłużył się metaforą, w której studia oceniające przyrównywane są do ulicznego spektaklu, sam zaś ewaluator jest jego aktorem, który „zachowuje się zgodnie ze scenariuszem naszkicowanych przez problemy edukacyjne, jego gra zaś ma wpływ nie tylko na przebieg spektaklu, lecz także na samą sztukę. Stwierdza on, że może poczynić bardzo niewiele założeń dotyczących tego, co się wydarzyło, co się aktualnie dzieje i tego, co się wydarzy w przyszłości. Staje w obliczu spierających się ze sobą grup interesów posługujących się odmiennymi definicjami sytuacji i odmiennymi, nie dającymi się jednocześnie zaspokoić potrzebami informacyjnymi. Jeśli zaakceptuje on waską definicję zakresu swojej działalności, może się okazać, że warunki kontraktu narzucają mu trudne do przyjęcia ograniczenia. Jeśli jednak okaże się, że ma pełną swobodę działania, to jego problemy wcale nie maleją. Musi zdecydować, któremu ze swoich potencjalnych klientów chce służyć, jakie informacje okażą się najbardziej użyteczne, kiedy będą one potrzebne i jak można je uzyskać. Usiłuję wykazać, że rozwiązując te problemy ewaluator nie uniknie zajęcia określonego stanowiska politycznego, wyrażenia swojej postawy wobec władz systemu edukacyjnego. [...] Tego rodzaju wymagań nie stawia się badaczowi akademickiemu. Dba on o to, by nie mieszać się w proces polityczny. Tworzenie nowej wiedzy i jej praktyczne stosowanie stanowią dlań dwa odrębne zagadnienia. Ewaluator jest nieuchronnie uwikłany w działanie, tj. w decyzje dotyczące alokacji środków oraz w określanie celów, ról i zadań" (tamże, s. 14–15).

Cronbach i współautorzy (1980) podtrzymują tezę o odmienności studiów ewaluacyjnych i badań akademickich, porównując te pierwsze z badaniami o charakterze eksperymentalnym: „Naukowiec spogląda na kolonię obiektów eksperymentalnych przez okular mikroskopu. W realiach studiów oceniających — ewaluator pływa w tej samej wodzie co obiekty jego badania i jest popychany przez te same prądy, które wyznaczają ich ruch" (s. 52).

Wprawdzie te rozróżnienia między badaniami akademickimi i ewaluacją trafnie opisują złożoność sytuacji, w jakiej znajdują się ewa-

luatorzy, sugerują jednak, że nauka akademicka nie jest powiązana z bieżącymi problemami społecznymi, co więcej — sprzeniewierzają się w ten sposób historycznym czy politycznym kontekstom nauki (Karier i współaut., 1973). Trzeba stwierdzić wyraźnie, że podział ten, w tej mierze, w jakiej trafnie zdaje sprawę z realiów, dotyczy przede wszystkim procesu ewaluacji i akcentuje ciągłe uwikłanie ewaluatorów w procesy polityczne oraz obciążenie wyników ich pracy tymi czy innymi wartościami. W społeczno-politycznym „spektaklu ulicznym" ewaluatorzy nie są niewrażliwi na uznanie, jakie pociąga za sobą stosowanie metody naukowej, uznanie, jakim obdarzani są badacze posługujący się „logiką prezentacji" zakorzenioną w określonej praktyce. Przedstawiciele nauk społecznych zachowują się tu nie inaczej niż przyrodnicy, choć ich bardziej elastyczne i w mniejszym stopniu bazujące na możliwości repliki wyniku metody pozwalają na posługiwanie się mniej wyraźnymi formami politycznej kooptacji aniżeli te, jakie kojarzymy zwykle z powojennym aliansem między naukami przyrodniczymi i militarno-przemysłowymi grupami interesów. W artykule z 1974 r. MacDonald przewidywał załamanie się wzorca bezstronnego badacza społecznego i wiązał to z narastaniem technokratycznego centryzmu. Współcześnie jego przewidywania wydają się trafne: okazuje się, że nieomal wszystkie badania społeczne finansowane ze środków publicznych lub przez organizacje pararządowe poprzestają na dość ściśle ograniczonych definicjach problemów, których dotyczą.

Niektórzy z badaczy społecznych charakteryzowali polityczny kontekst swojej pracy wskazując na to, że wymusza on na nich swoiste samookreślenie, skłaniając do opowiedzenia się po którejś ze stron (Becker, 1967). Własny wybór Beckera, socjologia zaangażowana (*partisan sociology*), został zakwestionowany w długiej replice Gouldnera, który koncepcję Beckera określił jako „socjologię drugorzędną" (Gouldner, 1973). Podobnie na obszarze studiów ewaluacyjnych przeświadczenia polityczne skłaniały ewaluatorów do akcentowania nierówności społecznych owocujących nierównościami w dostępie do systemu edukacyjnego. Wypada tu jednak odróżniać dwa zasadniczo odmienne zagadnienia: dostrzeganie tego, że studia ewaluacyjne są uwikłane w procesy polityczne, jest czymś innym aniżeli gotowość do podjęcia politycznej działalności. House utrzymuje, że każdy z trzech podstawowych sposobów uprawiania studiów ewaluacyjnych odwołuje się do różnych wersji liberalizmu (House, 1980, s. 45). Wydaje się, że zawęża to nieco obszar wyboru, niemniej w praktyce rozmaite formuły ewaluacji różnią sie od siebie dość znacznie, choć różnice te są dość trudno uchwytne w opisie

teoretycznym. Sam House (1976) odwołuje się do teorii sprawiedliwości Rawlsa (1994), stwierdzając, że powinna ona stanowić moralny układ odniesienia dla ewaluatorów. Postulował on także zastosowanie kryteriów moralnych właściwych koncepcji sprawiedliwości Rawlsa do technicznych decyzji dotyczących prawomocności wyników ewaluacji (House 1980, s. 257).

Cronbach odrzuca stanowisko House'a jako stronnicze i polityczne niewłaściwe, proponując jednocześnie w to miejsce swoją koncepcję ewaluatora jako badacza odpowiedzialnego przed całym społeczeństwem (*multipartisan „public scientist"*) (Cronbach i współaut., 1980, s. 210). MacDonald usiłował wprowadzić swój model „demokratycznej ewaluacji" jako sposób na współkształtowanie stosunków władzy i wsparcie nieuprzywilejowanych grup społecznych. Od chwili przedstawienia swojej propozycji odnosił coraz większe sukcesy w zawieraniu kontraktów, które dawały mu dostęp do ważnych grup interesu, i utrzymywał, że ewaluator musi wykraczać poza granice analizowanego przypadku, aby dotrzeć do źródeł określonych decyzji w sprawach publicznych (MacDonald, Norris, 1981). Obstając przy tezie mówiącej, że zleceniodawcom studiów ewaluacyjnych nie przysługują jakieś szczególne przywileje, a wszystkie zainteresowane strony, niezależnie od ich położenia w strukturze władzy, mają prawo do tego, aby uzyskać dostęp do ich wyników, proponował także szereg procedur realizacyjnych zorientowanych na likwidowanie przewagi jednej ze stron. Podejście zbliżone do tego, jakie sugerował MacDonald, zastosowano stosunkowo niedawno w Stanach Zjednoczonych jako „ewaluację udziałową" (*stake-holder evaluation*), ale miało ono charakter eksperymentu, a jego rezultaty, opisywane przez Weiss i House'a, sugerują, że — stanowiąc swoisty kompromis — nie zadowoliły żadnej ze stron (Weiss, 1986; House, 1986).

Większość prezentowanych wyżej poglądów dotyczących politycznego wymiaru studiów ewaluacyjnych odnosiła się do ewaluacji programów edukacyjnych realizowanych na dużą skalę. Jednakże podobne problemy pojawiają się też wtedy, gdy bierze się pod uwagę ewaluację rozmaitych inicjatyw lokalnych.

Ewaluacja jako perswazja

„Skłonny jestem utrzymywać, że studia oceniające same w sobie mają charakter perswazyjny. [...] Racje prezentowane przez ewaluatorów są mniej pewne, bardziej cząstkowe, bardziej osobiste, a jednocześnie czę-

ściej są zaczynem praktycznych działań aniżeli informacje dostarczane przez badania naukowe" (House, 1980, s. 72–73).

Stake wprowadzając dystynkcję między badaniami naukowymi (formalnymi) i badaniami zorientowanymi praktycznie, a także MacDonald, usiłując wskazać na cechy różniące badania i studia ewaluacyjne, zakładali, że kryteria, wedle których oceniamy badania naukowe, oraz kryteria oceny ewaluacji powinny się od siebie różnić. MacDonald kładzie nacisk na tę kwestię ostrzegając zarazem, że myślenie o ewaluacji jako szczególnym rodzaju badań naukowych grozi ewaluatorom popadnięciem w pułapkę „restryktywnych oczekiwań naukowej szacowności", w której „poznawcza czystość ma zastępować przydatność, a trywialne uzasadnienia są przedkładane nad nieudolne próby ogarnięcia złożoności" (MacDonald, 1974, s. 15). Formułując to ostrzeżenie, proponuje on jednocześnie dwa kryteria oceny wartości wyników ewaluacji, tj. ich przydatność i kompletność.

Kilku innych ewaluatorów zwracało uwagę na podobne zagadnienia. Od badań naukowych nie oczekuje się, że ich rezultaty będą bezpośrednio przydatne w praktyce, natomiast jest to niemalże definicyjna cecha ewaluacji (Weiss, 1972; Stake, 1980; House, 1980; Patton, 1978). Ignorowanie kryterium przydatności praktycznej grozi tym, że wyniki studiów ewaluacyjnych uznane zostaną za nieistotne i bezużyteczne. Podobnie ignorowanie „języka praktyki", *vocabulatory of action*[2], grozi zakłóceniami w komunikowaniu wyników ewaluacji. Wprowadzenie tych kryteriów bierze się z postrzegania studiów ewaluacyjnych jako swoistej usługi. Jeśli jednak formułowanie „rzetelnych uogólnień" jest niemożliwe, a z „trywialnych uzasadnień" należy zrezygnować, bo w niczym nie przyczyniają się do ogarnięcia złożoności przedmiotu analizy, to w jaki sposób i dzięki jakiej wiedzy ewaluacja może wpływać na świat praktyki społecznej?

W jednej z najświetniejszych analiz specyfiki studiów ewaluacyjnych, jakie możemy znaleźć w literaturze przedmiotu, House (1980) proponuje swoją odpowiedź na to pytanie, wskazując na różnicę między badaniami akademickimi i pracami ewaluacyjnymi oraz na kryteria oceny tych drugich. W pluralistycznym społeczeństwie — powiada House — oczekiwania formułowane pod adresem ewaluacji muszą różnić się od oczekiwań kierowanych pod adresem badań akademickich. Biorąc pod uwagę to, że wyniki ewaluacji prezentowane są różnym odbiorcom,

[2] Jest to określenie wprowadzone przez House'a (1973b) po to, aby wskazać, że tylko wtedy, gdy wyniki ewaluacji są choćby w jakimś stopniu zbieżne ze sposobem myślenia ich adresata o obszarach jego działania, mają one szanse na to, że zostaną uwzględnione.

podzielającym odmienne wartości i integrowanym przez odmienne interesy, trzeba zdawać sobie sprawę, że nie mogą one stać się niekontrowersyjnymi dyrektywami działań praktycznych. Prezentowane w takim kontekście, wyniki ewaluacji „bardziej skłaniają do czegoś, mniej zaś przekonują, bardziej argumentują, niż ujawniają, są mniej lub bardziej wiarygodne, natomiast nigdy pewne, bardziej zalecają, aniżeli nakazują" (tamże, s. 73). Intencją Hause'a było najprawdopodobniej położenia nacisku na to, że wyniki studiów ewaluacyjnych powinny wspierać racje służące interesom ludzi nieuprzywilejowanych, stosownie do podzielanej przezeń teorii sprawiedliwości społecznej Rawlsa. Jednakże wypada zauważyć, że analizując logikę argumentacji, jaką zazwyczaj posługują się ewaluatorzy, dochodzi on do wniosków, które w konsekwencji wykraczają poza tę ideę. Z jego punktu widzenia zniesienie rygorów rozumowania indukcyjnego bądź dedukcyjnego pozwala ewaluatorom na posługiwanie się w znacznie większym stopniu perswazją. „Gdy zniesiemy ograniczenia, jakie narzuca nam dążenie do pewności — pisze House — możliwości rozsądnego działania zwiększają się raczej niż zmniejszają" (House, 1977, s. 7).

Gdy House prezentował swoją koncepcje perswazyjnych funkcji ewaluacji, praktykujący ewaluatorzy dyskutowali już zróżnicowanie form prezentacji swojej pracy, poczynając od opisów studiów przypadku poprzez szczegółowe, „fotograficzne" eseje czy rozmaite formy dialogu, na radiowej prezentacji kończąc (Pick, Walker, 1976). Wielu z nich odwoływało się do tradycji literackich, sięgając po metaforę i reportaż (Brauner, 1974; MacDonald, Walker, 1975; Stake 1975). Elliot Eisner rozwijał swą koncepcję studiów ewaluacyjnych jako form krytyki polityki edukacyjnej (Eisner, 1984).

Eksponowanie perswazyjnych aspektów studiów oceniających odbiega nieco, przynajmniej z uwagi na inne rozłożenie akcentów, od propozycji MacDonalda kładącego nacisk na to, że ich wyniki powinny służyć podniesieniu poziomu debaty publicznej w sprawach o znacznej doniosłości dla tych czy innych zbiorowości. „Tak rozumiany raport podsumowujący wyniki ewaluacji można traktować przede wszystkim jako początek, nie zaś podsumowanie dyskusji nad tym, co powinno się robić dalej. Na tym polega funkcja raportów ewaluacyjnych, które prezentują możliwe do zakwestionowania charakterystyki określonych programów interwencji społecznych" (MacDonald, Sanger, 1982, s. 182).

Studia ewaluacyjne a procesy podejmowania decyzji

Krótki przegląd refleksji teoretycznej opisującej studia ewaluacyjne jako prace nastawione na społeczny konkret, uwikłane w mechanizmy polityczne i perswazyjne, ujawnia m.in. to, że ewaluacja rozumiana jako praca na rzecz tych czy innych decydentów, a więc taka koncepcja ewaluacji, która tkwiła u podstawy jej projektowania przez ostatnie ćwierćwiecze, nie wystarcza ewaluatorom usiłującym odpowiedzieć na pytanie, w jaki sposób wnoszą oni swój wkład w funkcjonowanie liberalnego, demokratycznego społeczeństwa. Od połowy lat sześćdziesiątych studia oceniające były postrzegane jako jedno ze źródeł informacji dla decydentów. Wcześniej jeszcze kojarzono je dość powszechnie z testowaniem osiągnięć szkolnych lub konfrontowaniem celów programów interwencji społecznych z ich rezultatami po to, aby odpowiedzieć na pytanie, które z tych pierwszych — i w jakim stopniu — udało się osiągnąć (House, 1973a, s. 3).

Przełom polegający na poszerzeniu funkcji ewaluacji tak, aby mogła wspomagać podejmowanie decyzji, zawdzięczamy Cronbachowi. Pisał on: „Aby ogarnąć wszystkie funkcje «ewaluacji», możemy zdefiniować ją jako gromadzenie i wykorzystywanie informacji pozwalających na podejmowanie decyzji dotyczących programu edukacyjnego" (1963, s. 672).

Podobne podejście do ewaluacji, różniące się od poprzedniego jedynie nieco odmiennym widzeniem jej roli w procesie *podejmowania* decyzji, reprezentują Stufflebeam i jego współpracownicy: „Ewaluacja to proces określania, gromadzenia i prezentowania użytecznych informacji, pozwalających na ocenę alternatyw, jakie ma przed sobą osoba podejmująca decyzję" (Stufflebeam i współaut., 1971, s. XXV).

Model proponowany przez Stufflebeama bezpośrednio nawiązywał do idei wspierania administratorów w ich postępowaniu zmierzającym do podjęcia decyzji w trybie racjonalnie wyznaczonych, kolejno następujących po sobie etapów, i odwoływał się do teorii racjonalnego podejmowania decyzji. W Wielkiej Brytanii idea gromadzenia informacji wspierających procesy podejmowania decyzji była najczęściej rozumiana jako idea „wykorzystania" określonej wiedzy i akcentowała to, że administratorzy stanowią jedynie część potencjalnych użytkowników informacji powstających w toku studiów ewaluacyjnych (MacDonald, 1970; 1973).

Informacyjne wspieranie osób podejmujących decyzje opiera się na założeniu, że podejmowanie decyzji dotyczących konkretnego programu zmiany społecznej nie należy do osób gromadzących informacje,

na decyzję bowiem ma wpływ wiele czynników, których ewaluator nie kontroluje. Najważniejsze jednak jest to, że byłoby wielce niewłaściwe, gdyby uzurpował on sobie uprawnienia przysługujące w demokratycznym społeczeństwie ludziom pełniącym określone role zawodowe i ponoszących z tego tytułu polityczną odpowiedzialność. Tych ról ewaluatorzy nie powinni brać na siebie. Zasadę tę jasno formułował Cronbach: „Opisujemy tu rolę zawodową, w którą wbudowane jest założenie, że ewaluator uważa się za człowieka odpowiedzialnego za realizację ogólnego interesu zbiorowego. Do niego należy niezależna ocena, nie powinien jednak dążyć do tego, aby jego ocena zastępowała proces polityczny" (Cronbach i współaut., 1980, s. 72).

Późniejsze analizy funkcji studiów ewaluacyjnych dalej akcentowały proces gromadzenia informacji wspierających procesy podejmowania decyzji, jednakże z biegiem czasu coraz częściej mówiło się o działaniu nie tyle na rzecz konkretnych decyzji, ile na rzecz określonej polityki społecznej czy edukacyjnej (Cronbach i współaut., 1980; MacDonald, Norris, 1981). Coraz częściej też kładziono nacisk na to, że komunikowanie zdobywanych doświadczeń stanowi istotny element ewaluacji (Stake, 1975; Cronbach i współaut., 1980).

Konkretne decyzje mogą odwoływać się do wyników studiów ewaluacyjnych, ale nie zawsze się tak zdarza, nie jest to też konieczne. Cronbach i jego współpracownicy (1980) utrzymywali, że „w sprawach publicznych z reguły nie dysponujemy zbiorami dobrze określonych opcji decyzyjnych. Rzadko podejmowane są decyzje typu «tak – nie»" (s. 62), a House (1980) wskazywał, że są takie formy studiów oceniających, które kładą większy nacisk na „rozumienie" aniżeli na ocenę (s. 19). Nie jest wykluczone, że uwagi dotyczące mniej bezpośredniego wykorzystania wyników ewaluacji w procesach decyzyjnych podyktowane były świadomością, że (a) udokumentowanie zasadności konkretnej decyzji nie jest łatwe, a wyniki ewaluacji stanowią jedynie część informacji branych pod uwagę przez decydentów; (b) model procesu decyzyjnego ma niewiele wspólnego z rzeczywistością; (c) idea odpowiedzialności badaczy wykonujących ewaluację wymaga głębokiego przemyślenia.

Analizy wskazujące na to, że wyniki studiów oceniających nie są wykorzystywane (Weiss, 1977; Patton, 1978), oraz niezadowolenie zarówno ewaluatorów, jak i ich klientów z wzajemnych kontaktów skłoniło Cronbach i jego współpracowników (1980) do przekonania, że należy zmodyfikować oczekiwania i uznać, iż wkład ewaluacji w procesy decyzyjne jest znacznie skromniejszy, choć nie mniej ważny, i polega głównie na funkcji edukacyjnej. Zadaniem studiów oceniających nie powinno być

przesądzanie o wartości ocenianego programu. „Ewaluacja nie może i nie powinna konkludować politycznej debaty redukowaniem wad i zalet określonych propozycji do «faktów»" (s. 61), lecz skłaniać do zastanowienia się nad tym, czego możemy się nauczyć w toku realizacji danego programu (s. 35). Mówiąc nieco inaczej, celem ewaluacji jest nie tyle wspomaganie wyboru określonej opcji decyzyjnej, ile przyczynianie sie do lepszego rozumienia programów zmiany społecznej i pobudzanie publicznej dyskusji w sprawach o znacznej doniosłości publicznej: „Nie chodzi o wybór jednej z opcji, lecz o negocjacje prowadzące do postąpienia kroku naprzód" (s. 66). Najkrócej rzecz ujmując, rola studiów ewaluacyjnych „nie polega na formułowaniu ostatecznych prawd, lecz na wyjaśnianiu, gromadzeniu dokumentacji, stawianiu nowych pytań i proponowaniu nowych punktów widzenia" (Lindbloom, Cohen, 1979, cyt. za Cronbach i współaut., 1980, s. 53).

Uświadamiając sobie te funkcje badań ewaluacyjnych i dochodząc do wniosku, że „powinny one służyć tym, którzy zastanawiają się nad pokierowaniem działaniami publicznymi i mają na nie wpływ" (Cronbach i współaut., 1980, s. 54), powinniśmy jednocześnie zdawać sobie sprawę, że upodobniają się one do innych badań służących polityce społecznej i oddalają od modelu badań eksperymentalnych, który dał im początek.

Ewaluacja „naturalistyczna"

Mój przegląd teoretycznych ustaleń dotyczących ewaluacji jest dość wybiórczy, skupiony wokół kilku autorów. Kierując się własnymi opiniami dotyczącymi ich wkładu w kształtowanie współczesnych koncepcji ewaluacji, zwracałam uwagę na takich autorów amerykańskich, jak Stake, House i Cronbach, oraz na MacDonalda z Wielkiej Brytanii. Chciałabym wierzyć, że nie jest to kwestia przypadku, że właśnie ci autorzy wpłynęli na nasze myślenie o ewaluacji, dlatego że silniej aniżeli inni ewaluatorzy akcentowali swój krytyczny stosunek wobec technokratycznego menedżeryzmu wbudowanego w powojenny ruch reform społecznych, uznając jednocześnie, że orientacja tego rodzaju zagraża najcenniejszym wartościom liberalnych, demokratycznych społeczeństw — dialogowi, poszanowaniu różnorodności i prawu do własnego sądu. Sądzę też, że właśnie im zawdzięczamy nieco teoretycznej przestrzeni, w której mogą rozwijać się bardziej wspólnotowe formy studiów ewaluacyjnych. Tytuł mojej książki, *Getting to Know Schools in a Democracy*, ma wyrazić mój osobisty związek z tymi wartościami, wskazując zarazem

na interesujące mnie zagadnienie, tj. pytanie, jak obmyślać studia oceniające, aby służyły upowszechnieniu wiedzy o systemie edukacyjnym, a jednocześnie nie zatraciły swego profesjonalizmu. Sądzę, że pojawianie się takich form życia zbiorowego, które Campbell nazwał „społeczeństwem eksperymentującym" (Campbell, 1977), to jest społeczeństwem stanowiącym jedno wielkie laboratorium, w którym badacze społeczni sugerują władzom publicznym najlepsze programy zmiany społecznej, odzwierciedla rozczarowania, jakie przyniosła idea „systemowej kontroli jakości" obecna w powojennym myśleniu o zmianie społecznej (por. Apple, 1979). W ciągu ostatnich dwudziestu lat historia programów zmiany społecznej i historia ewaluacji to dzieje kooptacji wielu badaczy społecznych, którzy wstępowali na służbę technokratycznych elit jako ewaluatorzy. Koncentrowałam się wprawdzie na omawianiu koncepcji ludzi, którzy odrzucali taką interpretację społecznej roli ewaluatora, ale zdaję sobie sprawę, że ich wysiłków nie uwieńczył sukces. Uwagi te mają służyć jako wstęp do nieco bardziej szczegółowego omówienia nietechnokratycznych koncepcji ewaluacji, które zazwyczaj określa się mianem „szkoły naturalistycznej" (Guba, 1978b; Guba, Lincoln, 1981; 1985).

Charakterystyka ewaluacji naturalistycznej

Studia naturalistyczne to ogólny termin stosowany na określenie tych alternatywnych podejść do ewaluacji, które pojawiły się w latach siedemdziesiątych jako reakcja na tradycyjne studia ewaluacyjne uznawane za niewystarczające dla ogarnięcia całej złożoności zmian w programach szkolnych. Te alternatywne podejścia obejmują takie propozycje, jak ewaluacja holistyczna (MacDonald, 1970; 1973 i 1978b), ewaluacja wyjaśniająca (Parlett, Hamilton, 1977), ewaluacja demokratyczna (MacDonald, 1974); ewaluacja dialogiczna (Stake, 1975), ewaluacja typu krytyki artystycznej (Kelly, 1975), ewaluacja transakcyjna (Rippey, 1973), ewaluacja *quasi*-sądowa (Wolf, 1979), ekspertyza edukacyjna (Eisner, 1975). Wypada też zauważyć, że dwie ostatnie z wymienionych propozycji zaliczano także — odpowiednio — do paradygmatu jurorskiego i paradygmatu konfrontacyjnego (Guba, Lincoln, 1981).

Źródłem tych wszystkich alternatywnych podejść do studiów ewaluacyjnych było przekonanie, że tradycyjny, eksperymentalny model ewaluacji nie wystarcza do ogarnięcia złożonych celów reform, które mają swoją własną dynamikę, zmieniają się w toku realizacji i często przynoszą nieoczekiwane rezultaty spowodowane zmianą kontekstu sy-

luacyjnego. Hamilton i House zaliczyli te wszystkie modele do szerszej kategorii ewaluacji pluralistycznej (to jest biorącej pod uwagę wartości wyznawane przez różnych adresatów), a House (1980) wskazał też argumenty polityczne, do których odwoływały się te modele. Hamilton proponuje następującą charakterystykę alternatywnych modeli[3] ewaluacyjnych:

„W porównaniu z modelami klasycznymi są one zazwyczaj szersze (nie koncentrują się wyłącznie na danych ilościowych), bardziej sytuacyjne (koncentrują sie na działaniu, a nie na intencjach przyświecających projektowanej zmianie), bardziej elastyczne (wychodzą poza schemat eksperymentu lub inne formalne schematy badawcze). Jednocześnie są bardziej podatne na wpływ wartości wyznawanych przez uczestników programów zmiany edukacyjnej i w większej mierze uwzględniają ich reakcje, posługując się przy tym ich językiem naturalnym, częściej korzystają z metod badawczych należących do arsenału etnografów prowadzących badania terenowe. Generalnie rzecz biorąc, mają one tendencję do przerzucania ciężaru oceny z ewaluatora na ludzi uczestniczących w programie poddawanym analizie" (Hamilton, 1978, s. 30).

Jak widać, dopiero przyjęcie bardzo rozległej perspektywy ujawnia to, co łączy tę zróżnicowaną gamę podejść alternatywnych — mimo przezornego „zazwyczaj", jakim Hamilton obwarowuje określone cechy przysługujące tym podejściom. Określenie „ewaluacja naturalistyczna" pojawiło się jako możliwy do zaakceptowania wspólny mianownik dopiero w końcu lat siedemdziesiątych, ale nawet wtedy poszczególni badacze mniej lub bardziej akcentowali metodologiczną, epistemlogiczną lub polityczną specyfikę swoich propozycji. Moim zdaniem, naturalistyczna koncepcja studiów ewaluacyjnych jest równoznaczna z przeświadczeniem, że programy zmian należy analizować biorąc pod uwagę ich społeczny kontekst, że najlepiej służą temu metody jakościowe, takie na przykład jak swobodny wywiad, bezpośrednia obserwacja czy historyczna rekonstrukcja. W ślad za tym idą takie formy prezentacji wyników, które skłaniają odbiorców do ich samodzielnego uogólniania, co Stake nazywał „uogólnieniem naturalistycznym" (Stake, 1980, s. 6). W swoim podsumowaniu House (1980) również kieruje naszą uwagę na

[3]Stake podkreślał, że wprawdzie w literaturze dotyczącej ewaluacji często pisze się o ogólnych modelach reprezentujących rozmaite podejścia badawcze, to jednak termin ten jest używany niewłaściwie. W zgodzie z House'em (1980) uważa on, że termin „model" z jego jednoznaczną zawartością znaczeniową lepiej jest zastąpić terminem „perswazja". Skłonna jestem przyjąć ten punkt widzenia. Jednakże posługuję się terminem „model", bo jest najczęściej stosowany w omawianych tu pracach.

naturalistyczne uogólnianie, przez które rozumie zwracanie się do niewyspecjalizowanych odbiorców, posługiwanie się językiem potocznym, nieformalnymi sposobami wnioskowania, odwołującymi sie do reguł myślenia potocznego, oraz argumentami strukturalizującymi (*attempting to establish*) rzeczywistość (s. 279).

Argumenty jakościowe, które zdaniem Hause'a mogą pomóc strukturalizować rzeczywistość, to przykład, metafora, ilustracja i analogia, a zatem apele nie popadające — jego zdaniem — w sprzeczność ani z logiką myślenia naukowego, ani z formą argumentacji rozwijanej także w dziedzinie ewaluacji. „Według Kuhna, uczeni nabywają biegłości w swoich specjalistycznych dziedzinach wzorując się na przykładach — konkretnych problemach rozwiązywanych w sposób pozwalający nowicjuszom na skonfrontowanie ich z innymi problemami. Te doświadczenia, a nie tylko formalne reguły warsztatu badawczego, są przekaźnikami grupowo podzielanych sensów" (tamże, s. 282).

Guba i Lincoln (1981) opowiadają się za zaproponowaną przez Wolfa i Tymitza charakterystyką podejścia naturalistycznego, w myśl której ewaluacje naturalistyczne zmierzają do zrozumienia „rzeczy w ich naturalnej postaci, rzeczywistości społecznej i tego, jak widzą ją ludzie, których percepcji nie deformują formalne metody pomiaru lub zawczasu obmyślane pytania. [...] Naturalistyczne podejście badawcze usiłuje prezentować fragmenty bieżących realiów, dokumentując je w języku naturalnym i możliwie dokładnie odzwierciedlając ludzkie uczucia i ludzką wiedzę, biorąc pod uwagę to, jak ludzie świat postrzegają i przeżywają, to, co o nim myślą, i to, co stanowi przedmiot ich troski" (s. 78).

MacDonald i jego współpracownicy (1975) pozostali przy prostszej metaforze: „Nasz cel przypomina zadanie przyrodnika usiłującego przeniknąć zarówno różnorodność, jak i wzajemną zależność form życia w stawie oraz jego powiązania ze światem zewnętrznym" (s. 2).

Jednakże ważniejsze od definicji są przekonania badaczy i przyjmowane przez nich założenia. Guba i Lincoln koncentrują się w swojej analizie na porównaniu naturalistycznego paradygmatu badawczego z paradygmatem nauki, na którym osadziła się już „patyna ortodoksji" — nawet w naukach społecznych (Guba, Lincoln, 1981, s. 56). Każdy z tych paradygmatów, utrzymują wspomniani autorzy, odwołuje się do określonych założeń dotyczących rzeczywistości, stosunku między badaczem a przedmiotem badania oraz koncepcji zdań prawdziwych (patrz tab. na stronie obok).

Podstawowe założenia naturalistycznego i naukowego paradygmatu badań

Założenia dotyczące:	Paradygmaty	
	scjentystyczny	naturalistyczny
Rzeczywistości	jest jednorodna, sfragmentaryzowana	jest różnorodna, wewnętrznie powiązana
Relacji badacz – przedmiot badania	są niezależni	są współzależni
Zdań prawdziwych	ogólnienia, twierdzenia nomotetyczne, nastawione na podobieństwa	hipotezy robocze, twierdzenia idiograficzne, nastawione na różnice

Źródło: Guba, Lincoln, 1981, s. 75[4].

Charakterystyka ewaluacji dialogicznej (Stake, 1975), studiów przypadku (MacDonald, Walker, 1975) a także ewaluacji wyjaśniającej, (Parlett, Hamilton, 1977), tj. trzech form ewaluacji, które mieszczą się w paradygmacie naturalistycznym, wspiera tezę Guby i Lincoln mówiącą, że badanie naturalistyczne obrazuje wielość rzeczywistości i wielość prawd, jakie są w związku z nimi wypowiadane, oraz że badanie to stanowi proces interaktywny prowadzący do „opisów zgrubnych" (Geertz, 1973) lub „hipotez roboczych" (Cronbach, 1975), nie zaś do generalizacji (Guba, Lincoln, 1981, s. 58).

House, wspierając się pracą Denzina (1971), przyrównuje naturalistyczną ewaluację do fenomenologicznej socjologii, ale dostrzega też różnice w zakresie podstawowych celów i funkcji: „Socjolog jest, oczywiście, zainteresowany budową ogólnej teorii. Ewaluator-naturalista koncentruje się jedynie na przedmiocie ewaluacji. Socjolog będzie adresował swoje konkluzje do ogółu. Wyniki naturalistycznych studiów ewaluacyjnych muszą być adresowane do konkretnych odbiorców, którymi mogą być też uczestnicy ewaluowanego programu. Zarówno socjologia, jak naturalistyczna ewaluacja odwołują się często do przykładów i doświadczeń zgromadzonych w toku badań terenowych. Ewaluator może traktować przykłady, którymi się posługuje, jako przesłanki swoich konkluzji. Jeśli przykłady są gromadzone i prezentowane systematycznie

[4]Guba i Lincoln są świadomi uproszczeń swojej analizy, jednakże posługują się jej wynikami po to, aby wskazać na podstawowe cechy badań naturalistycznych i ich ukrytych założeń epistemologicznych. W późniejszej książce wzbogacili tę analizę, a termin „naukowy" zastąpili terminem „pozywistyczny" (Guba, Lincoln, 1985).

logika ich przywoływania będzie przypominała logikę wnioskowania indukcyjnego. Jednakże w przypadku ewaluacji odbiorcy zawsze mają możliwość swobodnej oceny wiarygodności przykładów i interpretowania ich na swój sposób" (House, 1980, s. 280).

Trafność ewaluacji naturalistycznej

„Nie wystarcza, że studia ewaluacyjne są prawdziwe i wiarygodne, muszą one być także zgodne ze standardami warsztatowymi" (House, 1980, s. 250).

Problem trafności jest istotny dla większości badaczy, ale dla ewaluatorów ma on znaczenie zgoła podstawowe, m.in. dlatego, że wyniki ich pracy mogą być weryfikowane na wiele różnych sposobów. Na przykład nie zawsze i nie wszystkie formalne kryteria obowiązujące w badaniach naukowych mają zastosowanie w ewaluacji — chociaż w początkowej fazie rozwoju studiów ewaluacyjnych przywiązano do nich zasadnicze znaczenie, a różne podejścia do ewaluacji ustanawiały swoje własne kryteria trafności. W miarę upowszechniania się praktyki ewaluacyjnej kryteria o charakterze praktycznym stały się równie ważne jak kryteria warsztatu naukowego. Stufflebeam i jego współpracownicy (1971) proponują obszerną listę kryteriów: trafność wewnętrzna i zewnętrzna, rzetelność, obiektywizm (kryteria naukowe); wiarygodność, doniosłość, relewancja, terminowość, przejrzystość, efektywność (kryteria praktyczne) (s. 27–30).

House jest zdania, że trafność oznacza tyle, co „zasługiwanie na uznanie", ale nie pomija także definicji słownikowej: „[...] dobre osadzenie w faktach lub uznawanych zasadach mających odniesienie do badanego przypadku bądź sytuacji; poprawność i siła argumentacji, uzasadnienia, źródła itp." (House, 1980, s. 249).

Jednakże, jak sugeruje cytat, od którego rozpoczęłam swoje omówienie, trafność oznacza coś więcej niż poprawność techniczną, wokół której koncentrowała się większość dyskusji, trafność ma także swój wymiar etyczny. Współczesne pojmowanie studiów ewaluacyjnych jako procedury publicznego podejmowania decyzji implikuje odpowiedzialność wobec szerszego społeczeństwa, a „pojęcie trafności trzeba poszerzyć tak, aby dało się zastosować do tej ważnej społecznej praktyki" (tamże, s. 257).

House nie wpada w pułapkę uproszczonej trafności postulatywnej. Dzieląc podejścia do studiów ewaluacyjnych na takie, które odwołują się do epistemologii obiektywistycznej (analiza systemowa, podejście zoperacjonalizowanych zadań, procesy podejmowania decyzji), oraz

podejścia subiektywistyczne (krytyka literacka, inspekcja kwalifikacji, studium przypadku i podejścia *quasi*-sądowe), stwierdza jednocześnie, że każde zabiega o trafność na swój sposób. Jego podstawowe uwagi krytyczne dotyczą podejść obiektywistycznych, które w dążeniu do trafności odwołują się przede wszystkim do wyjaśnienia, kładą nacisk na powtarzalność uzyskiwanych rezultatów oraz rzetelność pomiarów, koncentrują się na prawdzie jako jednym z aspektów trafności, pomijając wiarygodność i aspekty normatywne. „Jeśli w badaniach naukowych jest to słabość — konkluduje House — to w przypadku ewaluacji oznacza to zasadniczą niedoskonałość" (s. 251). Takie studia oceniające są wiarygodne tylko dla tych, którzy wierzą w metodologię. Ci, którzy są przedmiotem studiów ewaluacyjnych, najczęściej do nich nie należą. Jego drugi komentarz krytyczny wobec podejść obiektywistycznych wychodzi od tego, że podejścia obiektywistyczne posługują się narzędziami badawczymi, w które wbudowane są wartości, z czego sami badacze najczęściej nie zdają sobie sprawy, tymczasem wartości te oddziałują na określanie dążeń i potrzeb badawczych. Są one najczęściej uznawane za oczywiste (tamże, s. 252).

MacDonald i Norris (1981) sformułowali podobny pogląd dotyczący obiektywizacji badań terenowych w naukach społecznych: „Lektura tomów podających przepisy, jak należy pitrasić badania społeczne, nie pozostawia wątpliwości, że w toku kolejnych historycznych przekształceń zagadnienia etyczne i polityczne zostały tu zredukowane do kwestii technicznych. Oznacza to, że polityczny dyskurs, który w demokratycznych społeczeństwach powinien towarzyszyć tym zagadnieniom, został zamknięty jeszcze zanim się rozpoczął" (s. 21).

Te podejścia badawcze, które House określa jako subiektywistyczne, źródeł swojej trafności szukają raczej w odwoływaniu się do doświadczenia niż do metody naukowej, ale starając się przezwyciężyć ograniczenia tej ostatniej — są również problematyczne ze względu na kryteria trafności. Na przykład o jednym z typowych podejść tego rodzaju, tj. o studium przypadku, House pisze: „Trafność jest uzależniona od dopasowania ewaluacji do doświadczenia uczestników programu oraz jej adresatów. Subiektywiści uważają, oczywiście, że rezultaty ich ewaluacji są również prawdziwe, ale usiłują uzyskać wgląd w rzeczywistość analizowanego programu nie wykraczając poza struktury poznawcze adresatów i uczestników. Ważna jest zrozumiałość. Studium ewaluacyjne jest dobre tylko wtedy, gdy jego wyniki są rozumiane przez odbiorców" (House, 1980, s. 253).

ZROZUMIENIE

Ewaluacja — konkluduje House — jeśli ma być trafna, musi być równocześnie zgodna z prawdą, wiarygodną i respektująca obowiązujące normy. Wprawdzie najczęściej trafność utożsamia się z prawdziwością (i oczywiście wyniki ewaluacji zostaną uznane za nietrafne, jeśli okażą się nieprawdziwe), ale to za mało w przypadku ewaluacji programów zmiany społecznej. Ich wyniki trzeba zakomunikować odbiorcom, oni z kolei uznają je (lub nie) za wiarygodne, a zatem uwierzą w ich prawdziwość. Wiarygodność, utrzymuje House, jest w równym stopniu uzależniona od ewaluatora (który może być uznany za osobę godną zaufania, a wtedy jego intencje i jego zdolność spójnego rozumowania i działania nie będą kwestionowane) co i od sposobu przeprowadzenia samej ewaluacji (o której opinie budowane są na podstawie takich cech, jak przejrzystość czy spójność raportu, jego autentyzm itp.). Dalsze zobowiązania ewaluatora wynikają z tego, że adresuje on wyniki swojej pracy do wielu odbiorców. Studia ewaluacyjne dotyczące programów publicznych nie mogą więc uwzględniać potrzeb jedynie niektórych z nich. „Dlatego też ewaluacja programów zmiany społecznej musi respektować obowiązujące normy" (tamże, s. 256).

Kolejnym elementem wkładu House'a w dyskusję dotyczącą trafności ewaluacji jest jego koncepcja bezstronności. Ewaluator zarówno w ramach paradygmatu pozytywistycznego, jak w ramach paradygmatu fenomenologicznego stara się pełnić rolę „bezstronnego obserwatora", osoby nie uwikłanej w problemy codzienności. House uważa jednak, że jest to rola „wątpliwa moralnie": „Troszcząc się o to, aby nie zostać uwikłanym w jakiekolwiek interesy, ewaluator pozostaje — paradoksalnie — obojętny na dobro wspólne" (tamże, s. 255).

Powołując się na swoją krytykę *Follow Through Ewaluation* (1979b), House wyjaśnia, co miał na myśli twierdząc, że podejście nawiązujące do analizy systemowej zastosowane do ewaluacji myli obiektywność z bezstronnością: „Analitycy sądzą, że obiektywność wystarcza, aby zapewnić im przewagę i wpływ. Najczęściej oznacza ona obojętny stosunek do zagadnienia (*irrelevance*). Obiektywność usiłuje uporać się z interesami zaangażowanymi w ewaluację poprzez ich ignorowanie. To, czego potrzebujemy, to bezstronność, która uwzględnia różne interesy i stara się je równoważyć" (House, 1980, s. 224).

Inni autorzy starali się pokazać różnice między sposobem traktowania kryteriów trafności w badaniach naturalistycznych i w badaniach scjentystycznych (por. np. Dawson, 1977; Guba, Lincoln, 1981; 1985). Nowsze rozeznanie związków między ewaluacją i pluralistycznymi badaniami na rzecz polityki społecznej (Cronbach i współaut.,

1980) pociągnęło za sobą zmiany tradycyjnych kryteriów trafności. Najwyraźniej zmiany te można dostrzec obserwując zmniejszenie nacisku na kryteria trafności wewnętrznej i zwiększenie na kryteria trafności zewnętrznej. Cronbach i jego współpracownicy (1980, s. 74) wiążą te zmiany ze znaczeniem, jakie zyskują kryteria trafności zewnętrznej w pracach części ewaluatorów (por. np. Cook, Campbell, 1979). Dostrzeżone przez Campbella znaczenie kryteriów trafności zewnętrznej w naukach stosowanych może okazać się równie doniosłe dla obrony i uprawomocnienia studiów ewaluacyjnych, co sformułowana przez Cronbacha w 1975 r. teza o postępującym zmierzchu generalizacji (Hamilton, 1980c). W następnych latach (Cronbach i współaut., 1980) myśl ta została rozwinięta poprzez wskazanie, że trafność wewnętrzna nie ma zasadniczego znaczenia w przypadku studiów ewaluacyjnych (s. 314). „Trafność zewnętrzna — trafność wniosków wykraczających poza dostępne materiały empiryczne — stanowi zagadnienie kluczowe dla interwencji społecznych, trafność wewnętrzna ma dla nich znaczenie uboczne" (s. 231). Jest to pogląd zgodny z opisanymi wyżej tendencjami zmian w określaniu funkcji studiów ewaluacyjnych.

Z postulatem Cronbacha mówiącym, że ewaluator jest w istocie „badaczem w służbie publicznej", to jest że powinien służyć pluralistycznej opinii publicznej, zbieżna jest ze względu na uwikłane w nią wartości „ewaluacja demokratyczna". Do koncepcji ewaluacji „wypranej z wartości" (*value-free*) krytycznie odnoszą się Nilsson i Hogben (1983) przypisując podobne poglądy Cronbachowi i MacDonaldowi. Jednakże ani Cronbach, ani MacDonald nie odrzuciliby zapewne poglądu House'a sformułowanego w toku dyskusji nad problemem trafności: „Publiczne studia oceniające powinny być demokratyczne, uczciwe i oparte na moralnych wartościach takich, jak równość, autonomia, bezstronność i wzajemność" (House, 1980, s. 256).

Należy jednak odnotować także, że wśród teoretyków opowiadających się za subiektywistycznymi epistemologiami studiów oceniających MacDonald jest tym, który broni obiektywistycznej koncepcji trafności wewnętrznej i utrzymuje, że między nią i trafnością zewnętrzną istnieją silne powiązania. Jego model procedury właściwej studiom ewaluacyjnym bierze pod uwagę zarówno kwestię technicznej trafności wyników, jak wartości wymienione w deklaracji House'a.

Ponieważ ważnym zagadnieniem w koncepcji ewaluacji demokratycznej staje się ponowne przemyślenie sposobów przekroczenia bariery dzielącej obiekty podlegające ewaluacji i adresatów jej wyników (MacDonald, Norris, 1981, s. 24), to — logicznie rzecz biorąc — dzielenie

kryteriów trafności na kryteria trafności wewnętrznej i kryteria trafności zewnętrznej musi się okazać co najmniej możliwe do zakwestionowania. Nie jest zatem bezsensowne utrzymywanie, że w demokratycznym procesie gromadzenia danych o tym, jak funkcjonują programy zmiany społecznej, szkoły czy klasy poddane rozmaitym interwencjom społecznym, kwestie wewnętrznej trafności wyniku ponownie nabierają znaczenia. MacDonald i Sanger (1982) odwołują się do literatury po to, aby wyrazić swoją opinię w tej kwestii: „Czy dosłowna prawda ma jakiekolwiek znaczenie? Z pewnością liczy się dla ludzi, których dotyczy" (s. 176).

W dalszej części omawianego artykułu jego autorzy uzasadniają tezę mówiącą, że w toku demokratycznego procesu, jakim powinna być ewaluacja, procesu, w którym kontrolę sprawuje zarówno badacz, jak ludzie podlegający ewaluacji, występuje napięcie między tymi ostatnimi, przywiązującymi znaczenie do trafności wewnętrznej, a zobowiązaniami ewaluatora wobec tego układu odniesienia, który Cronbach (Cronbach i współaut., 1980) nazywa „społecznością kształtującą politykę społeczną" (*policy-shaping community*).

Podsumowanie

Starałam się przedstawić najważniejsze wątki dyskusji dotyczącej studiów ewaluacyjnych prowadzonej w minionym ćwierćwieczu. Należy odnotować to, że z biegiem czasu koncepcja ewaluacji coraz bardziej oddalała się od modelu podporządkowującego ją osiąganiu określonych celów i w coraz większym stopniu brała pod uwagę potrzeby informacyjne decydentów po to, aby w następnej fazie skoncentrować się na szerszych celach politycznych, polegających głównie na informacyjnym wsparciu opinii publicznej i przyczynianiu sie w ten sposób do lepszego funkcjonowania systemu społecznego. Analizowałam też szereg założeń dotyczących ewaluacji, zwłaszcza tych, które mówiły, że jest to działalność o charakterze politycznym, konkretnym i perswazyjnym.

Powstanie naturalistycznego podejścia badawczego wiązało się z określonymi aspiracjami politycznymi; problem trafności jest tu traktowany, zgodnie z propozycjami House'a, przede wszystkim jako zagadnienie o charakterze moralnym. Właśnie takie podejście do ewaluacji stwarza możliwość demokratyzacji wiedzy i doskonalenia systemów edukacji szkolnej.

Rozważania te koncentrują się nie tyle na rozstrzygnięciu toczącej się obecnie dyskusji, ile na przedstawieniu jej podstawowych wątków.

Mam nadzieję, że ten krótki przegląd stanowisk ujawnia przynajmniej jedno, to mianowicie, że prowokują one więcej pytań niż dają odpowiedzi. Na przykład stosunkowo łatwo można zauważyć, że nacisk na konkret, którego dotyczy ewaluacja, i dążenie ewaluatorów do wniesienia swego wkładu w poziom wiedzy ogólnej, jaką posługuje się szeroka publiczność, rodzi pewne napięcia, jeśli zgoła nie sprzeczności. Nietrudno dostrzec także i to, że — nawiązując do dyskusji dotyczącej problemu trafności — nacisk na respekt dla standardów moralnych kłóci się z ideą pluralistycznej opinii publicznej. Wreszcie nie jest jasne, jak pogodzić respekt ewaluatora dla prawomocnej władzy politycznej z krytyką stosunków władzy, nie jest też oczywiste, że rozróżnienie to ma sens teoretyczny — nie mówiąc już o jego znaczeniu praktycznym. Są to zagadnienia o kapitalnym znaczeniu i dotyczą nie tylko ewaluacji. Warto jednak pamiętać, że prowadząc studia ewaluacyjne nie da się od tych zagadnień uciec. Chcąc nie chcąc, ewaluator jakoś je rozstrzyga — nawet jeśli nieświadomie.

tłumaczył *Kazimierz Frieske*

Rozdział II

Elementy teorii

David Nevo

Konceptualizacja ewaluacji edukacyjnej
Analityczny przegląd literatury

David Nevo, *The Conceptualization of Educational Evaluation: An Analytical Review of the Literature*, w: *New Directions in Educational Evaluation*, Ernest R. House (red.), The Falmer Press, London - New Philadelphia 1986.

W ostatnich latach podejmowano liczne próby precyzyjnego objaśnienia, czym jest ewaluacja, oraz wyłożenia różnic między ewaluacją a takimi zbliżonymi pojęciami, jak pomiar lub badanie. W literaturze znaleźć można wiele podejść dotyczących zarówno konceptualizacji ewaluacji, jak i jej miejsca w edukacji. Znaczną część z nich uznawano niesłusznie za „modele", mimo iż nie reprezentowały dostatecznego stopnia kompleksowości i kompletności (dotyczy to np. „modelu CIPP", „modelu niezgodności, „modelu dialogicznego" lub „modelu *goal-free*"). Wydaje się, że rację ma Stake (1981), iż są to nie tyle modele, co perswazje (*persuasions*).

Wielu z nas nie potrafiło odnaleźć się wśród różnych modeli, podejść i perswazji ewaluacyjnych, stąd próby pewnego uporządkowania rosnącej literatury przedmiotu. Ich celem była najczęściej klasyfikacja podejść ewaluacyjnych. Ważną zaletą tych klasyfikacji (np. Guba, Lincoln, 1981; House, 1980; Popham, 1975; Stake, 1976; Stufflebeam, Webster, 1980; Worthen, Sanders, 1973) jest krytyczny przegląd literatury przedmiotu z zaznaczeniem podobieństw i różnic między różnymi podejściami. Trzeba przy tym zauważyć, iż proponowane klasyfikacje miały — przynajmniej do pewnego stopnia — charakter holistyczny. Ujmowały one każdy model całościowo, lokując (wraz z innymi) w którejś z uprzednio wyróżnionych kategorii. Jednakże w tej próbie „całościowego oddania sprawiedliwości" każdemu modelowi ewaluacji pominięto często zagadnienia stanowiące istotę zgodności i niezgodności między różnymi podejściami.

Stufflebeam (1974) przedstawił listę ośmiu pytań, które — jego zdaniem — powinny stanowić punkt wyjścia każdej próby konceptualizacji ewaluacji. Nevo (1981) rozszerzył tę listę do dziesięciu głównych wymiarów ewaluacji. Dotyczą one podstawowych zagadnień, które poruszane są w najbardziej znaczących podejściach do ewaluacji w edukacji. Wymiary te zostaną wykorzystane jako oś organizująca analityczny przegląd literatury dotyczącej ewaluacji edukacyjnej. Można je przedstawić za pomocą następujących pytań:

1. Jak ewaluacja jest definiowana?
2. Jakie są funkcje ewaluacji?
3. Co jest obiektem ewaluacji?
4. Jakie rodzaje informacji o każdym obiekcie należy gromadzić?
5. Jakie kryteria powinny stanowić punkt wyjścia przy ocenie zalet i wartości ewaluowanego obiektu?
6. Komu ewaluacja powinna służyć?
7. Na czym polega proces ewaluacji?
8. Jakie metody badawcze należy stosować w ewaluacji?
9. Kto może przeprowadzać ewaluację?
10. Jakie standardy należy przyjąć za podstawę oceny przebiegu ewaluacji?

Niżej dokonamy przeglądu literatury podejmującej próby odpowiedzi na te pytania w kontekście propozycji zawartych w różnych modelach, podejściach i perswazjach ewaluacyjnych. Na zakończenie artykułu wskażemy potencjalne znaczenie, jakie analiza ta może mieć zarówno dla praktyków, jak teoretyków oraz badaczy ewaluacji.

Jak ewaluacja jest definiowana?

W literaturze można znaleźć wiele definicji ewaluacji. Jedna z najbardziej znanych, zaproponowana przez Ralpha Tylera, ujmuje ewaluację jako „proces zmierzający do stwierdzenia, w jakim stopniu założone cele edukacyjne są rzeczywiście realizowane" (Tyler 1950, s. 69). Inna szeroko akceptowana definicja ewaluacji koncentruje się na dostarczaniu informacji potrzebnych do podjęcia decyzji. Jej zwolennikami było kilku znanych ewaluatorów, takich jak Cronbach (1963), Stufflebeam (Stufflebeam i współaut., 1971) i Alkin (1969). Z kolei w ostatnich latach wielu ewaluatorów zaakceptowało pogląd, że ewaluacja jest szacowaniem (*assessment*) zalet lub wartości (Eisner, 1979; Glass, 1969; House, 1980; Scriven, 1967; Stufflebeam, 1974) bądź działalnością, na którą składa się zarówno opis, jak i ocenianie (Guba, Lincoln, 1981; Stake, 1967a). Niedawno powołano komitet, który podjął próbę ustalenia stan-

dardów ewaluacji. W jego skład weszło siedemnaście osób reprezentujących dwanaście organizacji zainteresowanych ewaluacją edukacyjną. W przyjętej przez komitet definicji ewaluacji określono ją jako „systematyczne badanie wartości lub zalet jakiegoś obiektu" (Joint Committee, 1981, s. 12).

Zgoda co do eksponowania w definicji ewaluacji funkcji osądzających została zakwestionowana przez grupę działającą w ramach Stanford Evaluation Consortium. Zdefiniowała ona ewaluację jako „systematyczne badanie zdarzeń, które mają miejsce w ramach aktualnie realizowanego programu bądź stanowią jego konsekwencję — badania te mają przyczynić się do usprawnienia zarówno tego programu, jak i innych, stawiających te same cele ogólne" (Cronbach i współaut., 1980, s. 14). Cronbach i jego współpracownicy (1980) w oczywisty sposób odrzucają osądzający charakter ewaluacji. Są oni natomiast zwolennikami podejścia postrzegającego ewaluatora jako „edukatora, [którego] sukces należy szacować według tego, czego nauczą się inni" (s. 11), a nie „arbitra sportowego", który został wynajęty po to, aby zadecydować, „kto ma rację", a „kto jej nie ma".

Trzeba zwrócić uwagę na fakt, że definicja eksponująca osądzający charakter ewaluacji wywoływałaby zapewne duży niepokój wśród tych, którzy mogą być ewaluowani, jak również wzmagałaby opór przeciwników ewaluacji. Z kolei definicja ewaluacji, która koncentruje się nie na osądzaniu, lecz na „dostarczaniu informacji potrzebnych do podjęcia decyzji" byłaby zapewne łatwiej akceptowana przez ewaluowanych i klientów. Z drugiej jednak strony, próby kształtowania pozytywnego stosunku do ewaluacji za pomocą pomijania jednej z jej głównych cech są pozbawione realizmu. Wydaje się, że alternatywą mogłoby tu być demonstrowanie pozytywnych funkcji ewaluacji w różnych dziedzinach edukacji.

Jakie są funkcje ewaluacji?

Scriven (1967) jako pierwszy zaproponował rozróżnianie ewaluacji „formatywnej" i ewaluacji „konkluzywnej" (chociaż nie był pierwszym mającym świadomość znaczenia takiego rozróżnienia) przypisując jej w ten sposób dwie główne role lub funkcje. Z kolei Stufflebeam (1972), nawiązując do tych samych funkcji, wyróżnił ewaluację proaktywną (*proactive*), która służy podejmowaniu decyzji, i retroaktywną (*retroactive*), która służy sprawozdawczości (*accountability*). Tak więc ewaluacja może spełniać dwie funkcje: formatywną i konkluzywną. W funkcji formatywnej wykorzystuje się ją do usprawniania i rozwijania aktualnie

trwających działań (lub wspomagania programów, osób, czy rezultatów itp.). W swojej funkcji konkluzywnej ewaluacja służy sprawozdawczości, nadawaniu certyfikatów lub selekcjonowaniu.

W tym miejscu należy wziąć pod uwagę trzecią funkcję ewaluacji — psychologiczną lub społeczno-polityczną, nie często rozważaną w literaturze ewaluacyjnej (Cronbach i współaut., 1980; House 1974; Patton, 1978). Jest oczywiste, że w wielu przypadkach ewaluacja nie służy ani celowi formatywnemu, ani sprawozdawczości czy innym celom konkluzywnym. Jest natomiast wykorzystywana do kształtowania świadomości dotyczącej poszczególnych działań, motywowania pożądanych zachowań tych, którzy są ewaluowani, lub promowania określonych stosunków społecznych. Tej funkcji ewaluacji nie można ignorować — i to niezależnie od osobistych odczuć w kwestii używania (lub nadużywania) ewaluacji do wymienionych wyżej celów.

Inna, w pewnym stopniu „niepopularna" funkcja ewaluacji polega na wykorzystywaniu jej do sprawowania władzy (Dornbusch, Scott, 1975). W organizacjach formalnych zwierzchnik posiada przywilej ewaluowania swoich podwładnych, natomiast sytuacja odwrotna zaistnieć nie może. W wielu przypadkach osoba na stanowisku kierowniczym może przeprowadzać ewaluację innej osoby w celu zademonstrowania swojej nad nią władzy. Tę funkcję ewaluacji nazwiemy „administracyjną".

W podsumowaniu trzeba stwierdzić, że ewaluacja może spełniać wiele funkcji: (a) formatywną — dla usprawniania; (b) konkluzywną — dla selekcjonowania, nadawania certyfikatów lub dla sprawozdawczości; (c) psychologiczną lub społeczno-polityczną — dla motywowania i kształtowania świadomości; (d) administracyjną — dla sprawowania władzy.

Część ewaluatorów (Alkin, Daillak, White, 1979; Cronbach i współaut., 1980) w oczywisty sposób preferuje formatywną funkcję ewaluacji. Jednak generalnie uznaje się, że nie istnieje coś takiego, jak „właściwe" lub „niewłaściwe" funkcje ewaluacji i że może ona w sposób zamierzony pełnić więcej niż jedną funkcję. Różne funkcje wymagają różnych sposobów i różnych metod ewaluacyjnych. Należy zatem uświadomić sobie ich istnienie i określić już na wczesnym etapie planowania ewaluacji rodzaj funkcji, jaką ma ona pełnić.

Co jest obiektem ewaluacji?

Popularnymi obiektami ewaluacji w dziedzinie edukacji byli zawsze uczniowie i nauczyciele. Do połowy lat sześćdziesiątych niemal cała literatura dotycząca pomiarów i ewaluacji w edukacji dotyczyła uczenia się. Trudno było znaleźć jakieś bardziej konkretne wskazówki na temat ewaluacji innych obiektów, takich jak projekty względnie programy edukacyjne, programy szkolne lub instytucje edukacyjne. Różne innowacje w systemie edukacyjnym Stanów Zjednoczonych (np. *Elementary and Secondary Education Act* z roku 1965) przesunęły punkt ciężkości w tej dziedzinie z uczniów na projekty, programy i materiały nauczania. To one stały się odtąd najczęstszym przedmiotem analiz w pracach znaczących autorów ewaluacji edukacyjnej (Alkin 1969; Provus 1971; Scriven, 1967; Stake, 1967a; Stufflebeam, 1969; Stufflebeam i współaut., 1971).

Z analizy współczesnej literatury ewaluacyjnej można wyciągnąć dwa wnioski: (a) niemal wszystko może stać się obiektem ewaluacji, stąd nie należy jej ograniczać do uczniów lub pracowników szkoły; (b) dokładne określenie obiektu ewaluacji stanowi ważny składnik każdego projektu ewaluacyjnego.

W planowaniu ewaluacji niezbędne jest określenie, czym są „rzeczy" (lub — aby użyć zwrotu Scrivena — „ewaluowani"), które mają być poddane ewaluacji. Dzięki temu można stwierdzić, jaki rodzaj informacji należy gromadzić i w jaki sposób należy ją analizować. Dokładne określenie obiektu „ogniskuje" ewaluację. Pozwala także określić i przeanalizować wartości, na jakie ewaluacja ma wpływ (Guba, Lincoln, 1981).

Jakie rodzaje informacji o każdym obiekcie należy gromadzić?

Wybrawszy obiekt ewaluacji trzeba się zdecydować, które jego aspekty i wymiary mają być poddane ewaluacji. To ich muszą dotyczyć gromadzone informacje. Wcześniejsze podejścia do ewaluacji koncentrowały się głównie na rezultatach lub wynikach. W takim ujęciu ewaluacja obiektu edukacyjnego (np. nowego programu) byłaby równoznaczna z ewaluacją jakości jego rezultatów (np. osiągnięć uczniów). W ostatnich latach w różnych modelach ewaluacji podjęto interesujące próby poszerzenia pola zmiennych ewaluacyjnych (Alkin, 1969; Provus, 1971; Stake, 1967a; Stufflebeam, 1969; 1974; Stufflebeam i współaut., 1971). Stufflebeam w swoim Modelu CIPP twierdzi, że ewaluacja ogniskuje się wokół czterech zmiennych każdego obiektu: (a) jego celów;

(b) planu działania; (c) procesu działania; (d) wyników. Zgodnie z tym podejściem, ewaluacja na przykład projektu edukacyjnego oznaczałaby szacowanie: (a) merytorycznej wartości jego celów; (b) jakości planu; (c) stopnia, w jakim plan został zrealizowany; (d) wartości jego rezultatów.

Z kolei w swoim modelu kontenansu (Countenance Model) Stake (1967a) zalecał gromadzenie dwóch rodzajów informacji na temat ewaluowanego obiektu: opisowych i orzekających. Informacje opisowe powinny dotyczyć zamierzeń, warunków wyjściowych, które mogą mieć wpływ na rezultaty, operacji zmierzających do praktycznego urzeczywistnienia zamierzeń oraz rezultatów. Z kolei orzekający zestaw informacji składa się ze standardów i ocen dotyczących wymienionych warunków, operacji i rezultatów.

Guba i Lincoln (1981) zmodyfikowali inny model Stake'a, model ewaluacji dialogicznej (Stake, 1975). Odwołując się do paradygmatu naturalistycznego twierdzili, że wynikiem pracy ewaluatora jest pięć rodzajów informacji: (a) informacje opisowe dotyczące obiektu ewaluacji, jego nastawienia i warunków towarzyszących; (b) informacje interesujące (*responsive information*) liczących się (*relevant*) odbiorców; (c) informacje o istotnych zagadnieniach; (d) informacje o wartościach; (e) informacje o standardach istotnych dla oszacowania wartości i zalet.

Z literatury przedmiotu wynikałoby zatem, że należy gromadzić szerokie spektrum informacji na temat ewaluowanego obiektu. Nie powinny się one ograniczać do wąsko pojętej ewaluacji wyników lub rezultatów. Nie znaczy to jednak, że w każdej ewaluacji należy zbierać wszystkie możliwe rodzaje informacji — można koncentrować uwagę na niektórych z nich, w zależności od priorytetów ewaluacyjnych lub istniejących ograniczeń praktycznych.

Jakie kryteria powinny stanowić punkt wyjścia przy ocenie zalet i wartości ewaluowanego obiektu?

Wybór kryteriów oceny zalet ewaluowanego obiektu stanowi jedno z najtrudniejszych zadań w ewaluacji edukacyjnej. Autorzy, którzy uważają, że ewaluacja powinna jedynie stwierdzać, czy zostały osiągnięte cele (Provus, 1971; Tyler, 1950), ułatwiają sobie zadanie poprzez częściowe zignorowanie problemu kryteriów ewaluacyjnych. W praktyce wykorzystują oni jako kryterium ewaluacji „osiągnięcie celu". Nie próbują przy tym uzasadniać tego kryterium. Nie wypowiadają się w kwestii celów błahych, a także tych wszystkich „założonych celów", które nie zasługują na realizację. Czy powinny być wykorzystywane jako kryteria ewaluacji?

Innym sposobem uniknięcia problemu kryteriów ewaluacji jest ignorowanie osądzającego charakteru ewaluacji. Autorzy, którzy definiowali ewaluację jako działanie zorientowane na gromadzenie informacji potrzebnych przy podejmowaniu decyzji lub realizacji innych celów (Alkin, 1969; Cronbach, 1963; Stufflbeam, 1969), mogli pominąć problem wyboru kryteriów.

W świetle literatury, osiągnięcie celów uznanych za ważne stanowi jeden z możliwych punktów wyjścia formułowania kryteriów ewaluacji. Wśród proponowanych przez różnych autorów alternatyw w tej dziedzinie można wymienić: zidentyfikowane potrzeby aktualnych lub potencjalnych klientów (Joint Committee, 1981; Patton, 1978; Scriven, 1972a); ideały lub wartości społeczne (Guba, Lincoln, 1981; House 1980); standardy ustalone przez ekspertów lub inne właściwe grupy (Eisner, 1979; Guba, Lincoln, 1981; Stake 1967a); jakość obiektów alternatywnych wobec ewaluowanego (House, 1980; Scriven, 1967).

Większość ekspertów ewaluacyjnych wydaje się podzielać opinię, że przy określaniu kryterium (lub kryteriów) szacowania każdego obiektu trzeba wychodzić od specyficznego kontekstu tego obiektu oraz funkcji, jaką ma pełnić jego ewaluacja. W wielu przypadkach ewaluator nie ma formalnych uprawnień do tego, aby wybrać kryterium spośród różnych alternatywnych. Jednak powinien dopilnować, aby decyzja w tej sprawie została podjęta. Na nim też spoczywa odpowiedzialność za uzasadnienie tego wyboru i to niezależnie od tego, czy miał wpływ na tę decyzję, czy podejmowali ją inni.

Komu ewaluacja powinna służyć?

Autorzy, którzy definiują ewaluację jako działanie dostarczające informacji potrzebnych do podjęcia decyzji (Alkin, 1969; Cronbach, 1963; Stufflebeam i współaut., 1971), nie mają problemu z odpowiedzią na pytanie, komu ewaluacja powinna służyć. Ustalają, kto jest właściwym decydentem, i próbują określić, jakie informacje są mu potrzebne. Inni teoretycy (Cronbach i współaut., 1980; House, 1980) odrzucają ideę „służby decydentom". Obawiają się uzależnienia ewaluatora bądź upraszczania analiz procesów społecznych i organizacyjnych. Cronbach i jego współpracownicy (1980) uważają, iż ewaluacja powinna służyć „społeczności, która kształtuje politykę społeczną" (*policy-shaping community*), a nie decydentom na kierowniczych stanowiskach. Wielu autorów uznaje, iż ewaluacja powinna służyć swoim klientom lub adresatom. Guba i Lincoln (1981) proponują termin „udziałowcy" (*stakeholders*) na

określenie grupy osób zainteresowanych w dokonaniu ewaluacji. To im właśnie ewaluacja powinna służyć.

Ewaluacja jest użyteczna o tyle, o ile wykorzysta ją jakiś konkretny klient lub adresat. Na podstawie literatury ewaluacyjnej trudno jest stwierdzić jednoznacznie, kim mieliby być „najwłaściwsi adresaci" ewaluacji. W tekstach poświęconych tej problematyce znaleźć można trzy propozycje: (a) ewaluacja może mieć więcej niż jednego klienta lub odbiorcę; (b) różni odbiorcy mogą mieć różne potrzeby ewaluacyjne; (c) specyficznych odbiorców i ich potrzeby ewaluacyjne należy jednoznacznie zidentyfikować na wczesnym etapie planowania ewaluacji.

Różnice potrzeb ewaluacyjnych mogą dotyczyć na przykład rodzaju informacji, które należy zgromadzić; poziomu, na którym przeprowadza się analizę wykorzystywanych danych; formy przedstawienia rezultatów ewaluacji. Niekiedy nie można zaspokoić wszystkich zidentyfikowanych potrzeb ewaluacyjnych jednocześnie. Stąd należy ustalić priorytety w tym zakresie, do których ewaluacja będzie mogła się odwoływać.

Na czym polega proces ewaluacji?

Proces ewaluacji może różnie przebiegać w zależności od podstaw teoretycznych, stanowiących jej punkt wyjścia. Podejścia teoretyczne, które postrzegają ewaluację jako działania zmierzające do określenia, czy osiągnięto cele (Tyler, 1950), zakładają następujący przebieg ewaluacji: (a) przedstawienie celów (w kategoriach behawioralnych); (b) opracowanie narzędzi pomiaru; (c) zbieranie danych; (d) interpretacja uzyskanych wyników; (e) sformułowanie rekomendacji.

Z kolei według modelu kontenansu (Stake, 1967a) proces ewaluacji powinien uwzględniać: (a) opis programu; (b) przedstawienie opisu właściwym adresatom; (c) uzyskanie i analizę ich opinii; (d) przekazanie wyników analizy ponownie adresatom. W swoim późniejszym modelu ewaluacji dialogicznej Stake (1975) postulował nieustanną „konwersację" między ewaluatorem i wszystkimi innymi związanymi z ewaluaowanym obiektem stronami. Wymienił dwanaście etapów dynamicznej interakcji między ewaluatorem i adresatami ewaluacji, które powinny mieć miejsce w trakcie jej przeprowadzania.

Provus (1971) zaproponował ewaluację składającą się z pięciu etapów: (a) sprecyzowanie projektu programu; (b) oszacowanie sposobów jego realizacji; (c) oszacowanie jego natychmiastowych rezultatów; (d) oszacowanie jego długofalowych rezultatów; (e) oszacowanie jego kosztów i korzyści.

Komitet Phi Delta Kappa do Spraw Badań (The Phi Delta Kappa Study Committee) (Stufflebeam i współaut., 1971) zaproponował trzyetapowy proces ewaluacji: (a) sprecyzowanie wymaganych informacji w interakcji z podejmującymi decyzje odbiorcami; (b) zdobywanie potrzebnych informacji poprzez formalne gromadzenie danych i procedury analityczne; (c) przekazanie informacji decydentom.

Scriven (1972b) w swoim modelu porównań ścieżkowych zaproponował dziewięć etapów. Z kolei Guba i Lincon (1981) twierdzą, iż ewaluacja o charakterze naturalistyczno-dialogicznym może być realizowana w procesie czteroetapowym: (a) inicjowanie i organizowanie ewaluacji; (b) identyfikowanie kluczowych kwestii i zagadnień; (c) gromadzenie potrzebnych informacji; (d) przedstawienie rezultatów i rekomendacji.

Eksperci ewaluacyjni nie zdołali ustalić kształtu „najlepszego" z możliwych procesów ewaluacji. Większość z nich zgodziłaby się jednak zapewne, że każda ewaluacja wymaga pewnej sumy interakcji między ewaluatorami i ich klientami — na początku ewaluacji w celu identyfikacji potrzeb ewaluacyjnych i na zakończenie w celu przekazania wyników. Ewaluacja nie powinna ograniczać się do działalności technicznej — zbierania danych i ich analizowania.

Jakie metody badawcze należy stosować w ewaluacji?

W ostatnich latach zakwestionowano użyteczność niektórych metod badawczych dla badań ewaluacyjnych (Guba, 1969; Stufflebeam i współaut., 1971). Jednocześnie jednak wprowadzono do ewaluacji edukacyjnej różne nowe metody. Tradycyjne metody eksperymentalne i *quasi*-eksperymentalne (Campbell, 1969; Stanley, 1972; Cook, Campbell, 1976) uzupełniono o metody naturaliztyczne (Guba, Lincoln, 1981; Patton, 1980), rozprawy parasądowe (*jury trials*) (Wolf, 1979), studium przypadku (Stake, 1980), metodę typu „krytyka artystyczna" (Eisner, 1977; 1979), metody dziennikarskie (Guba, 1978a), metodę typu *modus operandi* (Scriven, 1976c) i wiele innych. Wszystkie stały się uprawnionymi metodami ewaluacji. Wszakże część metodologów jest nadal przekonana o wyższości niektórych metod: z jednej strony — metod eksperymentalnych (Boruch, Cordray, 1980; Rossi, Freeman, Wright, 1979), z drugiej strony — metod naturalistycznych (Guba, Lincoln, 1981; House, 1980; Patton, 1980). Ogólnie jednak można zauważyć coraz większą akceptację eklektycznego podejścia do metodologii ewaluacyjnej. Na obecnym etapie rozwoju ewaluacji wydaje się zasadne stwierdzenie, że „ewalutor będzie na tyle mądry, aby nie obstawać przy wyłączności ani metodologii ilościowo-scjentystycznie-konkluzywnej, ani jakościowo-na-

turalistyczno-opisowej" (Cronbach i współaut., 1980, s. 7). Trudno też nie zauważyć, że dla realizacji tego skomplikowanego celu, jakim jest przeprowadzanie ewaluacji, trzeba znaleźć metodę (lub układ metod), która najlepiej pozwoli odpowiedzieć na szczegółowe pytanie ewaluacyjne. Stąd założenie, iż jest jedna metoda najlepsza dla wszystkich celów, nie wydaje się zasadne.

Kto może przeprowadzać ewaluację?

Wyodrębniając się jako grupa profesjonalna, ewaluatorzy poświęcili dużo uwagi zagadnieniu cech, jakie powinni posiadać „dobrzy" ewaluatorzy, oraz sposobom kształcenia tych osób (Boruch, Cordray, 1980; Cronbach i współaut., 1980; Guba, Lincoln, 1981; Stufflebeam i współaut., 1971; Worthen 1975). Kompetentny i godny zaufania ewalutor powinien łączyć w sobie techniczne kompetencje w dziedzinie dokonywania pomiarów, znajomość metod badawczych, zdolność rozumienia społecznego kontekstu i zawartości treściowej ewaluowanego obiektu, umiejętność nawiązywania dobrych stosunków z ludźmi, integralność osobowościową, obiektywizm, zdolności organizacyjne i odpowiedzialność. Ponieważ trudno znaleźć osobę, która posiadałaby wszystkie wymienione cechy, często trzeba powołać zespół do przeprowadzania ewaluacji lub do realizacji określonego zadania ewaluacyjnego wybierać osoby z najbardziej odpowiednimi cechami.

W literaturze ewaluacyjnej znaleźć można także dwa ważne rozróżnienia, które powinny być brane pod uwagę przy podejmowaniu decyzji dotyczącej doboru ewaluatorów. Pierwsze dotyczy podziału na ewaluatora wewnętrznego i zewnętrznego (Scriven, 1967; 1976b; Stake, Gjerde, 1974; Stufflebeam i współaut., 1971). Wewnętrzny ewaluator projektu jest zwykle zatrudniony przez projekt i przekazuje raporty bezpośrednio jego kierownictwu. Wydaje się oczywiste, że ewaluator taki jest mniej obiektywny i mniej wiarygodny dla osób spoza projektu niż ewaluator zewnętrzny, który jako nie zatrudniony przez projekt jest bardziej niezależny.

Drugie rozróżnienie, dokonane przez Scrivena (1967), dzieli ewaluatorów na zawodowców i amatorów. Wiąże się ono nie tyle z jakością przeprowadzanej ewaluacji, ile z typem wykształcenia i specjalizacji. Wykształcenie ewaluatora-amatora na ogół nie jest związane z ewaluacją, a działania ewaluacyjne stanowią jedynie część jego obowiązków zawodowych (np. jeden z dyrektorów projektu nowego programu matematyki, z wykształcenia matematyk, który kwalifikacje ewaluacyjne uzyskał w toku praktycznego udziału w ewaluacjach, przeprowadza

formatywną ewaluację tego projektu). Natomiast ewaluatorem-profesjonalistą jest ktoś, kto uzyskał wykształcenie w dziedzinie ewaluacji i dokonywanie ewaluacji jest jego głównym lub jedynym zajęciem (np. wewnętrzny ewaluator projektu pedagogiki specjalnej, który posiada tytuł magisterski w dziedzinie pomiaru i ewaluacji, jak również pięcioletnie doświadczenie w ewaluowaniu projektów pedagogiki specjalnej). Trzeba przy tym pamiętać, że o ile ewaluacyjne kwalifikacje ewaluatora-amatora w sensie technicznym mogą być mniejsze niż profesjonalisty, to jednak może on lepiej rozumieć jedyne w swoim rodzaju potrzeby ewaluacyjne określonego projektu i przygotować lepszy raport (w efekcie współpracy z uczestnikami ewaluowanego projektu).

Wymienione rozróżnienia są niezależne. Możemy mieć do czynienia z wewnętrznym ewaluatorem-amatorem, zewnętrznym ewaluatorem-amatorem, wewnętrznym ewaluatorem-profesjonalistą itd.

Jakie standardy należy przyjąć za podstawę oceny przebiegu ewaluacji?

W ostatnich latach podjęto kilka prób określenia standardów ewaluacji programów edukacyjnych i społecznych (Evaluation Research Society, 1980; Joint Committee, 1981; Stufflebeam i współaut., 1971; Tallmadge, 1977; US General Accounting Office, 1978). Niezależnie od faktu, że niektórzy autorzy (Cronbach i współaut., 1980; Stake, 1981) krytykowali — jako przedwczesny na obecnym etapie rozwoju ewaluacji — wysiłek na rzecz stworzenia wyczerpującego zestawu standardów, wydaje się, że osiągnięto konsens co do ich zakresu i treści.

Boruch i Cordray (1980) przeanalizowali sześć proponowanych zestawów takich standardów i doszli do wniosku, iż są one w dużym stopniu zbieżne. Najbardziej precyzyjny i wszechstronny zestaw (który jednocześnie uzyskał największą akceptację) został przygotowany i opublikowany przez Joint Committee on Standards for Educational Evaluation (1981). Standardy opracowywał siedemnastoosobowy komitet pod kierunkiem dra Daniela Stufflebeama. Reprezentował on dwanaście profesjonalnych organizacji związanych z ewaluacją edukacyjną. Komitet zaproponował trzydzieści standardów, które podzielono następnie na cztery grupy: standardy przydatności (aby ewaluacja zaspokajała zapotrzebowanie praktyki na informacje); standardy wykonalności (aby ewaluacja była realistyczna i rozważna); standardy przyzwoitości (aby ewaluacja była prowadzona w sposób zgodny z prawem i etyczny); standardy rzetelności (aby ewaluacja ujawniała i przekazywała informacje technicznie adekwatne).

Podsumowanie

Podejmując ryzyko uproszczenia, można podsumować dokonany wyżej przegląd literatury przedstawiając skrótowo najczęstsze odpowiedzi na wymienione dziesięć pytań. Dzięki temu uzyskamy zwięzły opis stanu naszej wiedzy na temat konceptualizacji ewaluacji edukacyjnej.

1. Jak ewaluacja jest definiowana?

Ewaluacja edukacyjna stanowi systematyczny opis obiektów edukacyjnych i/lub szacowanie ich zalet względnie wartości.

2. Jakie są funkcje ewaluacji?

Możliwe są cztery różne funkcje ewaluacji edukacyjnej: (a) formatywna (zapewnianie postępu); (b) konkluzywna (selekcja lub sprawozdawczość); (c) społeczno-polityczna (kształtowanie motywacji i uzyskiwanie poparcia społecznego); (d) administracyjna (sprawowanie władzy).

3. Co jest obiektem ewaluacji?

Wszystko, co istnieje realnie, może być obiektem ewaluacji. Typowymi obiektami ewaluacji w edukacji są uczniowie, personel edukacyjny i administracyjny, programy szkolne, materiały nauczania, programy, projekty i instytucje.

4. Jakie rodzaje informacji o każdym obiekcie należy gromadzić?

W przypadku każdego obiektu należy brać pod uwagę cztery zespoły zmiennych. Ogniskują się one wokół: (a) celów obiektu; (b) jego strategii i planów; (c) procesu wcielania ich w życie; (d) rezultatów i oddziaływania.

5. Jakie kryteria powinny stanowić punkt wyjścia przy ocenie zalet i wartości ewaluowanego obiektu?

Następujące kryteria powinny być brane pod uwagę przy ocenianiu zalet lub wartości obiektu edukacyjnego: (a) zdolność wychodzenia naprzeciw rozpoznanym potrzebom aktualnych i potencjalnych klientów; (b) realizowanie narodowych celów, ideałów lub wartości społecznych; (c) zgodność z przyjętymi standardami i normami; (d) przewyższanie jakością obiektów alternatywnych; (e) realizowanie (ważnych) własnych, uprzednio określonych celów. Kryteria powinny być zawsze dostosowane do obiektu.

6. Komu ewaluacja powinna służyć?

Ewaluacja powinna służyć potrzebom informacyjnym aktualnych i potencjalnych grup zainteresowanych obiektem ewaluacji. To właśnie na ewaluatorze (lub ewaluatorach) spoczywa odpowiedzialność za określenie, kim są zainteresowani ewaluacją oraz zidentyfikowanie lub zaprojektowanie ich potrzeb informacyjnych.

7. Na czym polega proces ewaluacji?

Niezależnie od metody badawczej w procesie ewaluacji należy uwzględnić trzy następujące działania: (a) sformułowanie problemu ewaluacji; (b) gromadzenie i analizowanie danych empirycznych; (c) komunikowanie rezultatów odbiorcom ewaluacji. Możliwa jest przy tym zawsze więcej niż jedna właściwa sekwencja realizacji tych działań. Każda taka sekwencja może (i czasami powinna) być powtórzona kilka razy w trakcie przeprowadzania badań ewaluacyjnych.

8. Jakie metody badawcze należy stosować w ewaluacji?

Ewaluacja jako zadanie kompleksowe wymaga posłużenia się wieloma alternatywnymi metodami typowymi dla nauk o zachowaniu (*behavioral sciences*) i związanych z nimi dziedzin. Sposób ich spożytkowania wynikać powinien z natury specyficznego problemu ewaluacji. Na obecnym etapie rozwoju ewaluacji jako dziedziny brak podstaw do apriorycznego preferowania jakiejś szczególnej metody.

9. Kto może przeprowadzać ewaluację?

Ewaluację mogą prowadzić osoby lub zespoły posiadające: (a) rozległe kompetencje w zakresie metodologii badawczej oraz technik analizy danych; (b) umiejętność rozumienia społecznego kontekstu ewaluacji i jego jedynej w swoim rodzaju materii; (c) umiejętność nawiązywania dobrych stosunków z ludźmi i porozumienia się z osobami i grupami włączonymi w ewaluację; (d) strukturę mentalną pozwalającą zintegrować wymienione wyżej umiejętności.

10. Jakie standardy należy przyjąć za podstawę oceny ewaluacji?

W ewaluacji powinno się dążyć do uzyskania optymalnej równowagi między: (a) przydatnością (możliwość wykorzystania i praktyczność); (b) rzetelnością (adekwatność techniczna); (c) wykonalnością (realizm i rozwaga); (d) przyzwoistością (przeprowadzenie ewaluacji w sposób zgodny z prawem i etyczny).

Wnioski

Krytyczna analiza różnych teoretycznych podejść do ewaluacji edukacyjnej ma ważne implikacje zarówno dla praktyków ewaluacji, jak teoretyków i badaczy zainteresowanych rozwojem nowych koncepcji i coraz lepszych metod. Im wszystkim może się przydać schemat analityczny złożony z przedstawionych wyżej dziesięciu pytań, jak również dokonany przegląd zawartych w literaturze ewaluacyjnej odpowiedzi na te pytania.

Ewaluatorzy mogliby wykorzystać te dziesięć pytań dla przemyślenia własnego podejścia do ewaluacji oraz posłużyć się literaturą przedmiotu dla opracowania spójnego zestawu własnych odpowiedzi na te pytania. Byłoby to równoznaczne z rezygnacją z bezkrytycznego adaptowania tego czy innego modelu ewaluacji. Ponadto owe dziesięć pytań pomaga zrozumieć, o czym myślą inni, gdy odwołują się do pojęcia ewaluacji. Ewaluatorzy mogą napotkać znaczne trudności postrzegając konkretną ewaluację odmiennie od swoich klientów lub odbiorców. Ważne jest więc, aby jeszcze przed przystąpieniem do planowania ewaluacji lub nawet przed podjęciem decyzji w tej sprawie odpowiedzieć sobie na następujące pytania: Jak pojmują ewaluację różne uczestniczące w niej strony? Jakim celom ma (w założeniu) służyć? Co ma być ewaluowane? Jakie istnieją (wykonalne) alternatywy wobec niej? Jakie standardy powinny stanowić podstawę oceny ewaluacji samej w sobie (jeśli w ogóle mamy ją przeprowadzać)? Mówiąc inaczej, odwołanie się do dziesięciu pytań dyskutowanych w tym rozdziale mogłoby wyjaśnić problemy ewaluacyjne, zanim staną się one częścią, pełnej niepewności, przygody ewaluacyjnej.

Dyskusje prowadzone przez teoretyków ewaluacji mogą przyczynić się do rozwoju teorii i praktyki ewaluacyjnej. Ich rezultaty mogłyby być jeszcze bardziej znaczące, gdyby koncentrowały się one kwestiach spornych, a nie na konkurujących ze sobą modelach i paradygmatach. Byłoby przy tym bardzo korzystne, gdyby różnorodne propozycje teoretyczne uzasadniano określonymi rezultatami badawczymi. Dziesięć pytań przeanalizowanych w tym rozdziale mogłoby stanowić schemat wyjściowy dla określenia empirycznych zmiennych ewaluacji. Dane o rzeczywistych zależnościach między tymi zmiennymi, jak również o zależnościach między nimi a np. praktycznym wykorzystaniem ewaluacji lub zmiennymi odzwierciedlającymi kontekst ewaluacji byłyby z pewnością bardzo cenne dla społeczności ewaluatorów.

tłumaczył *Zbyszko Melosik*

Helen Simons

Polityczne implikacje teorii ewaluacyjnych
Przybliżenie problemu

Helen Simons, „The Political Implications of Evaluation Theory: A Closer Look", rozdział z książki tejże autorki *Getting to Know Schools in a Democracy*, The Falmer Press, London – New York – Philadelphia 1987.

Chcę zastanowić się nad pewnymi szczegółowymi kwestiami związanymi z politycznym uwikłaniem teorii ewaluacyjnych. Omawiane dalej poglądy formułowane były nie tyle w kontekście wielkich programów zmiany społecznej, ile w odniesieniu do poszczególnych szkół, ich potrzeb rozwojowych i oczekiwań związanych z ich publiczną odpowiedzialnością. Sądzę, że przed podjęciem analizy powinnam przypomnieć trzy podstawowe pytania i podtrzymać tezę mówiącą, że od odpowiedzi na te pytania zależy jakość kształcenia. Oto pytania:

- Kto powinien prowadzić studia ewaluacyjne w dziedzinie edukacji?
- Kto powinien ustalać kryteria oceny programów szkolnych, rozwiązań instytucjonalnych i organizacyjnych?
- Kto powinien mieć dostęp do wyników studiów ewaluacyjnych?

Jedna z trudności, jakie nastręcza tego rodzaju teoretyczny przegląd stanowisk, polega na tym, że racje stron są prezentowane poza kontekstem, w którym zostały sformułowane. Ze znaczenia tej trudności praktyk zajmujący się studiami ewaluacyjnymi poszczególnych przypadków — a sama kimś takim jestem — nie może nie zdawać sobie sprawy. W nadziei przynajmniej częściowego rozwiązania tego problemu chcę omówić najpierw poglądy moich amerykańskich kolegów, potem zaś przedstawić stanowiska teoretyków brytyjskich. Za taką strukturą prezentacji przemawia to, że zestawy analizowanych poglądów funkcjonują w odrębnych kontekstach, ponadto zaś będę mogła pokazać charakterystyczną odmienność praktyki ewaluacyjnej w Wiel-

kiej Brytanii i w Stanach Zjednoczonych: otóż w Wielkiej Brytanii koncentruje się ona przede wszystkim na samoewaluacji szkół, w Stanach zaś na ewaluacji poszczególnych programów edukacyjnych. Do grona teoretyków amerykańskich, których poglądy analizuję, włączyłam także Ralpha Tylera. Jest on dość powszechnie uznawany za jednego z prekursorów problematyki, która mnie tu interesuje. „Wszystkie współczesne podejścia badawcze charakterystyczne dla studiów ewaluacyjnych programów szkolnych albo stanowią rozwinięcie propozycji Tylera, albo pośrednio do nich nawiązują" (Hamilton i współaut., 1977, s. 26).

Wpisanie Tylera na listę autorów, których poglądy chcę przedstawić, skłania mnie do poczynienia jeszcze jednej uwagi. Otóż nie jest wykluczone, że rozwój studiów ewaluacyjnych traktowanych jako osobna praktyka stanowiąca zarazem w ostatnim ćwierćwieczu przedmiot refleksji teoretycznej nie byłby się dokonał lub miałby znacznie skromniejszy wymiar, gdyby nie wprowadzenie w Stanach Zjednoczonych ustawowego obowiązku dokonywania ewaluacji wszystkich programów społecznych finansowanych z funduszów federalnych. *The Elementary and Secondary Education Act* z 1965 r. nadał rozwojowi studiów ewaluacyjnych znaczną dynamikę[1]. Pracę Tylera i pierwsze prace Cronbacha wniosły liczący się wkład w tę problematykę jeszcze przed wprowadzeniem w życie tych przepisów, natomiast prace prawie wszystkich autorów, o których wspominam, powstały później. Czytelnicy zechcą może zauważyć, że przed 1965 r. studia ewaluacyjne były finansowane na znacznie mniejszą skalę niż w następnych latach. Można też przedstawić argumenty wskazujące na to, że skromność środków finansowych, jakie przeznaczane były na studia ewaluacyjne przed 1965 r., stwarzała okoliczności podobne do tych, jakie w wiele lat później i w innym kraju skłoniły część z nas do uznania, że studia ewaluacyjne mogą stać się przesłanką rozwoju szkolnictwa. Oczywiście, wypada zdawać sobie sprawę z tego, że historia studiów ewaluacyjnych jest znacznie dłuższa aniżeli okres omawiany w tej książce, nie zamierzam więc sugerować, że należy ignorować wszystko, co okres ten poprzedzało, jednakże wczesna historia studiów ewaluacyjnych została już dobrze udokumentowana (por. Hamilton, 1976c; 1978; Cronbach i współaut., 1980; Guba, Lincoln, 1981; Norris, 1984) i nie widzę powodów, dla których należałoby przedstawiać ją jeszcze raz.

[1] Znakomicie udokumentowany przegląd skutków *The Elementary and Secondary Education Act* i losów towarzyszących mu studiów ewaluacyjnych znaleźć można w pracy M.W. McLaughlina *Evaluation and Reform: The Elementary and Secondary Education Act of 1965, Title I*, Rand Educational Policy Study, Ballinger Publishing Company, Cambridge Mass. 1975.

Przegląd koncepcji amerykańskich

W pracach kilku teoretyków wyraźnie formułowane jest założenie, że tylko osoby o najwyższych kompetencjach poznawczych, o najwyższym poziomie umiejętności i w najwyższym stopniu zdolne do wyzbycia się personalnych uprzedzeń — to jest właśnie specjaliści-ewaluatorzy — powinni prowadzić studia ewaluacyjne. Ten nacisk na specjalizację jest funkcją trzech czynników: szybkiego rozwoju „nowej" profesji; liczby studiów ewaluacyjnych prowadzonych w Stanach Zjednoczonych; rodzaju specjalistycznego przygotowania tych, którzy z biegiem czasu stali się zawodowymi ewaluatorami. W kontekście ogromnych nakładów finansowych przeznaczanych na doskonalenie programów szkolnych stało się oczywiste, że potrzebne są ich niezależne oceny ze strony obserwatorów zewnętrznych. Ośrodki finansujące zmiany w systemie edukacyjnym nie mogły takich niezależnych ocen formułować, bo stanowiły stronę, zaangażowanie innowatorów wyznaczających kierunek tych zmian uniemożliwiało zachowanie obiektywizmu (Scriven, 1967, s. 45; 1974, s. 36), natomiast nauczycielom realizującym nowe programy szkolne zwykle brakło umiejętności pozwalających na prowadzenie profesjonalnych studiów ewaluacyjnych (Scriven, 1967, s. 47). Przepisy z 1965 r., nakazujące uzależnienie decyzji finansowych od wyników studiów ewaluacyjnych, stworzyły rynek dla ewaluacji, wyznaczyły jego dynamikę i umocniły pozycję ekspertów. Kompleksowość różnych modeli studiów ewaluacyjnych osiągnęła taki poziom, że stały się one zajęciem dostępnym *tylko* dla specjalistów. Praktycy podzielali i podzielają nadal to przekonanie, a w Stanach Zjednoczonych status profesjonalnego ewaluatora nigdy nie był poważnie kwestionowanym, choć moim zdaniem stanowisko Stake'a w tej sprawie nie jest do końca jasne.

Najbardziej jednoznacznie o specjalistycznych kompetencjach ewaluatorów wypowiedział się Scriven: „Przekonanie, że każdy system szkolny lub każdy nauczyciel może dziś prowadzić dojrzałe studia ewaluacyjne dotyczące swoich własnych prac, jest równie absurdalne, co przekonanie, że każdy współczesny psychoterapeuta jest zdolny samodzielnie oceniać swoją pracę z pacjentami. Mówiąc najprościej, każdy z nich może się wiele nauczyć uważnie analizując swoją pracę, może dostrzec jej mankamenty lub cechy warte utrwalenia jednakże ktoś, kto oczekuje odpowiedzi na ważne pytania dotyczące efektów ich pracy lub źródeł ich sukcesów i porażek, będzie musiał odwołać się do wiedzy i umiejętności, jakie spotyka się dość rzadko, nawet w skali całego kraju" (Scriven, 1967, s. 53).

Z nieco innego punktu widzenia Eisner (1984), odwołując się do takich określeń, jak krytyka i ekspertyza edukacyjna, argumentuje tezę, że studia ewaluacyjne powinny pozostać domeną wyspecjalizowanych obserwatorów zewnętrznych. Podobne opinie wyrażał także Stufflebeam (i współaut., 1971) uznając, na przykład, że jego model studiów ewaluacyjnych znany jako Context, Input, Process Product Model (CIPP), nawiązujący do teorii racjonalnego podejmowania decyzji, nadaje się do realizacji jedynie przez wyspecjalizowanych ewaluatorów.

Stanowisko Tylera w tej sprawie jest znacznie bardziej złożone. W swojej prezentacji klasycznego modelu behavioralnego kładł on nacisk na opis planowania programu szkolnego uznając jednocześnie, że ewaluacja stanowi część tego procesu (Tyler, 1950). Jego dobrze ustrukturowany model teoretyczny sugeruje, że studium ewaluacyjne powinno stanowić zadanie dla specjalistów. Jednakże Guba i Lincoln (1981, s. 6) utrzymują, że praca Tylera skierowana była przede wszystkim do nauczycieli, a więc do osób, które ponoszą odpowiedzialność za rozwój i zmianę programów szkolnych.

Tyler kierował swoją uwagę głównie na szkołę. Interesowało go doskonalenie programów szkolnych, a jego opis procesu doskonalenia programu szkolnego, choć bez wątpienia bardzo szczegółowy, nie wyłącza nauczycieli (por. np. Tyler, 1950, s. 107 i 109). Jednakże terminy, jakimi się posługiwał, takie np. jak „konstruktor programu" (s. 114), czy „pracownik programu" (s. 111), a także budowa specjalistycznych narzędzi ewaluacji (s. 118) oraz znaczna skala przedsięwzięć polegających na „konstruowaniu programów" pozwala przypuszczać, że pisząc o studiach ewaluacyjnych miał na myśli ekspertów nie należących do personelu szkolnego.

Moda na takie formy rozwoju programów szkolnych, jakie zaczęły dominować w latach sześćdziesiątych, sprawiła, że model Tylera okazał się przydatnym instrumentem ich rozwoju i ewaluacji, i zaczął być wykorzystywany przez ludzi, których później Hamilton nazwał „programistami" (Hamilton, 1978).

W okresie dzielącym powstanie modelu Tylera od pojawienia się pierwszych prac Cronbacha (1963) zmienił się zasadniczo kontekst rozwoju systemów edukacji. W dużej mierze rozwój ten został przyspieszony przez rywalizację wywołaną wystrzeleniem sputnika, a zmiany edukacyjne zaczęły koncentrować się na modernizacji programów nauczania poszczególnych przedmiotów w szkołach średnich i objęły cały kraj (Guba, Lincoln, 1981). W tym kontekście Tylerowski model studiów ewaluacyjnych, zbudowany z myślą o zdecentralizowanych procesach

tworzenia nowych programów szkolnych i koncentrujący się na formach pracy nauczycieli z uczniami, był mało odpowiedni. Jednakże wzmocniło to nacisk na udział eksperta-ewaluatora w procesie reformowania programów szkolnych. Pojawienie się pierwszych prac Stake'a (1967a) i House'a (1973a) oznaczało wprawdzie stopniowe odchodzenie od konsensualnego modelu studiów ewaluacyjnych, ale nie podważało roli eksperta-ewaluatora. Specjalistyczna wiedza i umiejętności są ciągle w cenie.

Odpowiedź na moje drugie pytanie, kto powinien ustalać kryteria oceny nowych programów szkolnych i nowych instytucji systemu edukacyjnego, nie jest jednoznaczna. Kilku teoretyków studiów ewaluacyjnych odpowiada na nie tak samo jak na pierwsze z postawionych przeze mnie pytań, należą do nich w pierwszej kolejności Scriven i Eisner. Scriven nie ma tu żadnych wątpliwości: ocena wartości programu należy do ewaluatora. Pogląd ten odnajdujemy zarówno w jego stosunkowo wczesnej pracy (Scriven, 1967), jak w późniejszej koncepcji ewaluacji nie skoncentrowanych na konkretnych celach (*goal-free evaluation*) (Scriven, 1974). Eisner z kolei nie podejmuje interesującego nas tutaj zagadnienia bezpośrednio, jednakże promowana przezeń idea znawstwa systemów edukacyjnych implikuje pogląd, że ocena należy do eksperta. Odnotowując „elityzm wbudowany *implicite* w koncepcje zakładające istnienie «kultury wyższej», a więc także znawstwo systemów edukacyjnych", (Hamilton i współaut., 1977), a także to, że „ewaluacja nie jest dla mas" (Guba, Lincoln, 1981), inni ugruntowują ten pogląd.

Ci teoretycy ewaluacji, którzy swą podstawową rolę widzą w gromadzeniu informacji użytecznych w procesach podejmowania decyzji (Stake, Stufflebeam, Tyler, Cronbach), uznają wprawdzie potrzebę analizowania opinii dotyczących przydatności czy celowości programu, ale nie biorą na siebie zadania formułowania jego autorytatywnej oceny. Uznają, że dokonywanie tego rodzaju oceny należy do osób pełniących inne role w systemie, choć poglądy dotyczące tego, o jakie tu role chodzi, są dość zróżnicowane.

Stufflebeam uznaje na przykład, że podejmowanie decyzji należy do administratorów (Stufflebeam i współaut., 1971, s. 96). „Ewaluator nie powinien ani podejmować decyzji dotyczących programu, ani stawać się ich wykonawcą" (s. 93). W pracach Tylera odnajdujemy to samo napięcie, jakie rejestrowałam już wcześniej. W zasadzie zwraca się on do nauczycieli, uznając ich za osoby, od których ostatecznie zależy wybór celów stawianych programowi zmiany (Tyler, 1950, s. 4), ale ich wstępnej selekcji dokonują eksperci (s. 86) lub — co wynika ze

sposobu, w jaki jego model został wdrożony w latach sześćdziesiątych — specjaliści zajmujący się rozwojem programów edukacyjnych.

W pierwszej pracy Cronbacha (1963) znajdujemy uwagi świadczące, że i on podstawowych adresatów widział w tych właśnie specjalistach. Jednakże w następnych pracach, które charakteryzowałam wcześniej jako zbliżone do nurtu pluralistycznych studiów ewaluacyjnych, adresatami wyników ewaluacji okazują się już nie specjaliści, lecz wszyscy zainteresowani oraz uprawnieni do wywierania wpływu na programy szkolne (Cronbach i współaut., 1980, s. 66). Celem studiów ewaluacyjnych jest przygotowanie informacji podlegającej analizie i niezależnej interpretacji (s. 17), „ułatwianie podjęcia decyzji, nie jej narzucanie" (s. 155). Cronbach i jego współpracownicy zwracają uwagę, że ich punkt widzenia bliższy jest tradycji badań ewaluacyjnych w Anglii (MacDonald, Parlett, 1973) aniżeli koncepcji, jaką prezentował Scriven, sugerując, że to właśnie ewaluator powinien informować publiczność, czy program edukacyjny jest wystarczająco dobry (Cronbach i współaut., 1980, s. 154).

Stosunkowo najmniej rygorystycznie do pytania, kto powinien oceniać programy szkolne, podchodzi Stake. Przyjmuje, że wiele jest takich zainteresowanych grup, które powinny mieć prawo do wypowiedzenia opinii: rodzice, nauczyciele, a także podatnicy. Nie kwestionuje on wprawdzie tego, że specjalistyczne szkolenie przygotowuje ewaluatora do formułowania ocen, jednak zwraca uwagę, że byłoby rzeczą nierozsądną przeciwstawianie specjalistycznych opinii ocenom formułowanym na innej podstawie, toteż, jego zdaniem, studia ewaluacyjne powinny służyć wypracowaniu zrównoważonego kompromisu: „Odpowiedzialność za doprowadzenie do sformułowania oceny jest znacznie bardziej właściwym rozumieniem roli ewaluatora aniżeli uznanie, że jego zadaniem jest samodzielne jej formułowanie" (Stake, 1967a, s. 149). Jednocześnie jednak nie lekceważy on doniosłości specjalistycznych ocen. Zarówno w jego wczesnej (Stake, 1967a), jak i późniejszej pracy (Stake, Easley, 1978) odnajdujemy takie np. koncepcje, jak posłużenie się grupową oceną ekspertów. Niemniej należy odnotować, że z biegiem czasu poglądy Stake'a ulegały zmianie zbliżając się do demokratycznego pluralizmu.

Trzecie pytanie, kto powinien mieć dostęp do wyników ewaluacji, interesujący mnie badacze — poza Stake'em, House'm i Cronbachem — podejmują jedynie pośrednio, niektórzy zaś (np. Eisner i Scriven) nie podejmują go w ogóle, choć biorąc pod uwagę znaczenie, jakie przywiązują do roli specjalistów, można sądzić, że skłonni byliby uznawać,

że rezultaty ewaluacji powinny być udostępniane przede wszystkim zleceniodawcom, specjalistom, którzy opracowują i wdrażają nowe programy szkolne, oraz zainteresowanym kolegom. Stufflebeam, pozostając w zgodzie ze swoim modelem racjonalnej pętli decyzyjnej, wierzy, że wystarczy, jeśli adresatami będą administratorzy, co więcej — że wyrażają oni także interesy innych grup odniesienia (Stufflebeam i współaut., 1971, s. 112–113). Jest to stanowisko zgodne ze stanowiskiem Scrivena, gdy mówi o „korzyści, jaką odnosi całe społeczeństwo" (Scriven, 1967, s. 81).

Wśród teoretyków zainteresowanych tą problematyką najbardziej stanowczo wypowiadają się Stake i House. Stake podnosił to zagadnienie kilka razy, m.in. wtedy, gdy omawiał swoją koncepcję dialogicznych studiów ewaluacyjnych. Akcentował przy tej okazji m.in. to, że na ewaluatora spada odpowiedzialność za prezentowanie swoich ocen tak, aby uwzględniały one rozmaite punkty widzenia i wyrażone były w języku zrozumiałym dla szerokiego grona odbiorców — m.in. po to, aby zwiększyć uczestnictwo w systemie edukacyjnym lub przynajmniej pomóc w jego lepszym zrozumieniu (Stake, 1972; 1975; 1977). Rekapitulacje tego stanowiska znajdujemy u Hamiltona: „Dialogiczne studia ewaluacyjne są «odpowiedzią» na pytanie o nieprofesjonalnych odbiorców. Są one konsekwentnie demokratyczne, unikają specjalistycznego żargonu i usiłują podporządkować się regułom «naturalnego» komunikowania się ludzi. Koncentrują się bardziej na działaniach aniżeli na intencjach, proponują interpretacje i opisy. Akcentują problemy, traktując je jako punkt wyjścia do zrozumienia złożonych zjawisk" (Hamilton i współaut., 1977, s. 145).

Jak już wskazywałam, House jest tym amerykańskim teoretykiem studiów ewaluacyjnych, który najwyraźniej angażuje się w problematykę polityczną. Jednakże, m.in. z uwagi na jego szerokie zainteresowania, rekonstrukcja jego poglądów nie przychodzi łatwo. Jest on nie tylko ewaluatorem-praktykiem (House i współaut., 1971; House, Gjerde, 1973), lecz także teoretykiem studiów ewaluacyjnych (House, 1973b; 1976; 1977; 1979a) oraz metaewaluatorem (House i współaut., 1978) i metateoretykiem (House, 1978a; 1980). Dwie ostatnie z jego ról powodują, że niekiedy trudno ustalić, czy House jedynie analizuje, czy też opowiada się za określoną teorią.

Jednakże dwie jego idee mają związek z odpowiedzią na pytanie o to, komu powinny służyć wyniki studiów ewaluacyjnych. Pierwsza z nich to stosunkowo wczesna wypowiedź, w której stwierdza, że warunkiem wstępnym wykorzystania rezultatów ewaluacji jest wyra-

żenie ich w „języku działania", jakim posługują się decydenci (House, 1973c, s. 135). Druga zawarta jest w znacznie późniejszym omówieniu warunków, jakie powinny zostać spełnione po to, aby ewaluację zdemokratyzować (House, 1980, s. 139–157)*. Tu prezentuje on swoje własne stanowisko dotyczące polityki wyboru, które — jego zdaniem — jest znacząco odmienne od liberalnych założeń, na jakich opiera się wiele mechanizmów podejmowania decyzji i wiele studiów ewaluacyjnych: „Właściwa liberalnej demokracji zasada równych szans domaga się nie tylko maksymalnych możliwości wyboru, ale i tego, aby były one w jakimś stopniu równo dystrybuowane. Liberalna demokracja stwarza zatem warunki maksymalizacji nie tylko asortymentu, ale również jego dystrybucji" (tamże, s. 141).

Jak wynika z tego krótkiego przeglądu prac teoretyków amerykańskich, może poza pracami Stake'a, moje trzy pytania, które uznaję za ważne, nie były jako takie rozpoznawane w czasach, gdy prace te powstawały.

Trzeba też zwrócić uwagę, że w dyskusjach dotyczących studiów ewaluacyjnych rzadko pojawia się kwestia ich celów. Hamilton (1978) składa to na karb konsensu w kwestii celów, charakterystycznego dla wielu wczesnych modeli ewaluacji: „Wizja konsensu została stosunkowo dobrze sformułowana, była w oczywisty sposób racjonalna i nie kwestionowano jej" (s. 327).

Ów konsens umacniała skłonność „ekspertów", tj. „kompetentnych zawodowych ewaluatorów" — jak nazywał ich Scriven (1967, s. 53) — lub „konstruktorów programu", mówiąc językiem Tylera (1950, s. 114), do mówienia o „korzyściach", jakie odnosi całe społeczeństwo (Scriven, 1967, s. 81). W późnych latach sześćdziesiątych konsens wartości i interesów zaczęto kwestionować (Atkin, 1967–68; Stake, 1967a), ale dopiero w latach siedemdziesiątych w modelach studiów ewaluacyjnych zaczęto uwzględniać problemy, jakie nastręcza pluralizm wartości.

Ewaluatorzy brytyjscy

W Wielkiej Brytanii rozwój studiów ewaluacyjnych postępował inną drogą. Wprawdzie formalnie reformom programów szkolnych towarzyszyły systematyczne prace ewaluacyjne, ale zapoczątkowane zostały nieco później aniżeli w Stanach Zjednoczonych i nie były zdominowane tradycją behawiorystyczną (Hamilton i współaut., 1977). Inspirowała je

*Chodzi o rozdział książki *Evaluating with Validity*, który pod tytułem *Demokratyzacja ewaluacji* publikujemy niżej na stronach 225–243 — *przyp. red.*

głównie na wpół niezależna Schools Council[2], organizująca ewaluacje ogólnokrajowych programów szkolnych; ewaluatorami byli ludzie działający w systemie edukacyjnym — w szkołach różnego szczebla, na uniwersytetach, w kuratoriach — mający bardzo różne przygotowanie zawodowe. Na ogół byli to absolwenci różnych wydziałów uniwersyteckich, niektórzy mieli stopnie naukowe, nie istniały natomiast specjalistyczne ośrodki szkoleniowe dla ewaluatorów, a tradycja tego rodzaju prac badawczych praktycznie w Wielkiej Brytanii nie istniała[3]. Studia ewaluacyjne prowadzili ludzie uczący się w trakcie pracy, formułujący pytania teoretyczne stosownie do bieżących trudności, jakie musieli pokonywać. Inaczej aniżeli ich amerykańscy koledzy, większość teoretyków brytyjskich rozpoczynała *od praktyki* ewaluacyjnej. Brytyjska teoria studiów ewaluacyjnych rozwijała się w ścisłych związkach z praktyką.

Różne drogi rozwoju studiów ewaluacyjnych w Stanach Zjednoczonych i w Wielkiej Brytanii, które w końcu doprowadziły do wyłonienia się nowej profesji, mogą po części tłumaczyć, dlaczego podstawowy model studiów ewaluacyjnych, powszechny w Stanach Zjednoczonych w latach pięćdziesiątych i sześćdziesiątych, oparty przede wszystkim na ewaluacjach dotyczących innowacji w programach szkolnych (MacDonald, 1974; Hamilton i współaut., 1977), nigdy nie zyskał wielkiej popularności w Wielkiej Brytanii. Nieistnienie określonej rutyny badawczej powodowało tu, że chętniej dyskutowano różne alternatywne propozycje w tej dziedzinie. W pracy Hamiltona i jego współpracowników (1977) znajdujemy też tezę mówiącą, że różnice dzielące amerykańskie i brytyjskie podejście do studiów ewaluacyjnych są zakorzenione w kulturowych różnicach między tymi krajami: amerykańskiej smykalce do

[2]The Schools Council for Curriculum Development and Examinations in England and Wales była *quasi*-rządową, ogólnokrajową agencją finansowaną wspólnie przez Department of Education and Science i Local Education Authority, zorganizowaną w 1964 r. po to, aby wspierać rozwój programów szkolnych i reformę systemu egzaminacyjnego w szkołach. Przez pierwsze dziesięć lat swego istnienia wspierała ona centralne programy zmian edukacyjnych realizowane w wielu szkołach. Na skutek fali krytyki i w obliczu zmniejszającego się budżetu wysiłek tej agencji w późnych latach siedemdziesiątych koncentrował się na programach badawczych służących wspomożeniu zmian edukacyjnych wprowadzanych na małą skalę oraz na studiach ewaluacyjnych. Dalsza krytyka doprowadziła w 1984 r. do jej rozwiązania, w jej miejsce zaś powstały dwie inne agencje, Secondary Examination Council i School Curriculum Development Committee. Opis powstania i upadku School Council patrz *Life and Death of the Schools Council*, M. Plaskow (ed.), Falmer Press, Lewes 1985.

[3]Można, oczywiście, twierdzić, że Her Majesty's Inspectorate (HMI) przez całe lata prowadził studia ewaluacyjne systemu edukacyjnego poprzez inspekcje poszczególnych szkół, ale aż do 1983 r. raporty z tych inspekcji były poufne, a sam proces inspekcji pozostaje poufny do dzisiaj. Większość pracowników HMI nie ma żadnego doświadczenia w prowadzeniu empirycznych badań społecznych (por. Walker, 1982).

racjonalistycznej inżynierii społecznej — z jednej strony — i brytyjskim zaufaniu do doświadczania praktycznego i sceptycznym podchodzeniu do tezy o sprawczym potencjale racjonalnego myślenia — z drugiej.

W konsekwencji, gdy podejmowałam swoje prace badawcze, których wyniki zawarte są w tej książce, niewiele można było powiedzieć o systematycznej refleksji nad politycznym uwikłaniem studiów ewaluacyjnych w Wielkiej Brytanii, a to, co mieli do powiedzenia w tej sprawie Amerykanie, nie miało zastosowania do brytyjskich realiów. Pierwsze próby ewaluacji projektów reformy programów szkolnych wskazywały na to, że wiele uwagi należy poświęcić kontekstowi wdrażania tych reform na poziomie poszczególnych klas i szkół. Stąd nacisk na studia przypadku i ewaluacje całościowe (holistyczne) (Simons, 1971; MacDonald, 1970). Inaczej aniżeli w Stanach Zjednoczonych, w stosunkowo wczesnym okresie rozwoju studiów ewaluacyjnych podstawowym obiektem ewaluacji stają się szkoły, a nie programy systemu edukacyjnego. Z tego powodu włączyłam do dyskutowanych dalej materiałów prace takich badaczy, jak Eraut czy Elliott, którzy — nie nawiązując do idei ewaluacji programów reform edukacyjnych — zajmowali się zagadnieniami bliskimi problematyce efektywności programów kształcenia ustawicznego nauczycieli. Odwołanie się do tego kontekstu spowodowało, że poglądy ewaluatorów w odniesieniu do trzech interesujących mnie zagadnień różniły się dość istotnie od stanowiska amerykańskich teoretyków ewaluacji.

Charakterystyczne są w tej mierze dwa istotne teksty brytyjskich ewaluatorów sprzed 1976 r. W 1972 r. Parlett i Hamilton jako pierwsi przedstawili zainteresowanemu środowisku tekst o „ewaluacji wyjaśniającej", w którym nawoływali do zerwania z modelem studiów ewaluacyjnych opartym na technikach statystycznych i eksperymentalnych* i położenia nacisku na społeczne i antropologiczne aspekty ewaluacji (Parlett, Hamilton, 1977). Ta praca niemalże natychmiast uzyskała dość znaczny rozgłos. Była szeroko czytana i komentowana przez specjalistów zajmujących się programami szkolnymi, została też uznana za interesującą alternatywę studiów ewaluacyjnych skoncentrowanych na pomiarze efektów reform edukacyjnych.

*Model ten określano metaforycznie jako „rolno-botaniczny", odwołujący się do metodologii opartej na myśleniu hipotetyczno-dedukcyjnym i, jak piszą Parlett i Hamilton, przejętej przez badania edukacyjne z wczesnych studiów nad eksperymentami rolno-botanicznymi. Patrz: M. Parlett, *Evaluating Innovations in Teaching*, w: H.J. Butcher i E. Rudd, *Contemporary Problems in Higher Education*, McGraw-Hill, London 1972 — *przyp. red.*

Parlett i Hamilton nie zajmowali się we wspomnianej pracy odpowiedzią na pytanie o to, kto powinien prowadzić studia ewaluacyjne, a przynajmniej nie zajmowali się tym zagadnieniem bezpośrednio. Chociaż proponowane przez nich metody badawcze są — w zasadzie — dostępne również dla niespecjalistów, to jednak, podobnie zresztą jak teoretycy amerykańscy, zakładali oni, że badania ewaluacyjne powinni prowadzić wyspecjalizowani fachowcy (s. 20). Co do drugiej kwestii, poglądy tych autorów zbieżne są z opiniami Stake'a. Wybór kryteriów oceny programu edukacyjnego nie powinien leżeć w gestii jednej tylko z zainteresowanych grup: „Nie ulega najmniejszej wątpliwości, że ewaluator, który uwzględnia interesy każdej z tych grup, nie może — nawet jeśli się tego od niego oczekuje — dać prostej odpowiedzi na pytanie o przyszłość innowacji. Decyzja w tej sprawie oparta na kryteriach oceny przyjętych przez jedną z zainteresowanych grup może prawie na pewno zostać zakwestionowana przez inną grupę, która ma swoje, odmienne priorytety" (Parlett, Hamilton, 1977, s. 19).

Co do trzeciej z interesujących mnie kwestii, omawiani autorzy wyraźnie opowiadają się za pogladem, że wyniki studiów ewaluacyjnych powinny być dostępne wielu zainteresowanym ośrodkom uczestniczącym w podejmowaniu decyzji dotyczących ewaluowanych programów (s. 19). Interpretacje tę wspiera znaczenie, jakie przywiązują w omawianej pracy do prezentacyjnej warstwy raportów badawczych i postulatu mówiącego, że powinny być one dostępne dla wszystkich zainteresowanych (s. 20).

Podobnie jak Stake (1975), Parlett i Hamilton nie zajmują się politycznymi aspektami studiów ewaluacyjnych bezpośrednio, nie zwracają też uwagi na polityczny kontekst ich realizacji. Tu ich stanowisko jest zbliżone do poglądów wielu teoretyków amerykańskich, utrzymujących, że istota ewaluacji polega przede wszystkim na określonej metodologii. To, że właśnie metodologia studiów ewaluacyjnych skoncentrowanych na wyjaśnianiu przyciąga uwagę komentatorów, potwierdzają inne prace Parletta, w których proponuje on podejście wyjaśniające jako nowy paradygmat badawczy (Parlett, 1975).

W swoich kolejnych pracach poświęconych ewaluacji wyjaśniającej Parlett[4] zaczął zajmować się jej dwoma aspektami politycznymi: po pierwsze, akcentował, że podejście to jest nie tylko istotną zmianą orientacji metodologicznej, lecz inaczej definiuje samą ewaluację, kładąc

[4] W tym czasie zainteresowania Hamiltona przeniosły się na badania klas szkolnych (1976a i 1976b), teorię studiów oceniających (1976c, 1978 i 1980c) oraz na historię teoretycznych przesłanek i założeń ewaluacji programów szkolnych i badań edukacyjnych (1978, 1980a i 1980b).

nacisk na odpowiedzialność ewaluatora wobec ludzi w ten czy inny sposób uwikłanych w program stanowiący przedmiot ewaluacji (Parlett, Dearden, 1977, s. 33); po drugie, dostrzegał także konieczność znalezienia kompromisu między „prawem zainteresowanej publiczności do informacji" a „prawem jednostek do wyboru tych informacji, którymi chcą się dzielić z innymi" (tamże, s. 34) — choć nie wskazywał jak w określonych realiach politycznych problemy te miałyby być rozwiązywane. W jakimś stopniu to zainteresowanie politycznymi aspektami można uznać za konsekwencję nabywania doświadczeń praktycznych (dla nikogo nie ulega wątpliwości, że polityczne aspekty studiów ewaluacyjnych należy brać pod uwagę), jednakże na sposób formułowania i analizowania tej problematyki wyraźnie oddziałała teoria polityczna MacDonalda. Dalej omówię ją bardziej szczegółowo.

Drugą pracą, która miała istotny wpływ na teorię i praktykę studiów ewaluacyjnych w Wielkiej Brytanii, był tekst MacDonalda „Ewaluacja i kontrola w dziedzinie edukacji" (1974), który bezpośrednio dotyczy kwestii politycznych. Omawiając polityczny wymiar ewaluacji, MacDonald proponuje jej klasyfikację w kategoriach typów idealnych i wyróżnia ewaluację autokratyczną, biurokratyczną i demokratyczną. Koncepcja demokratycznych studiów ewaluacyjnych powstała, przynajmniej po części, jako negatywna reakcja na technokratyczne formy ewaluacji. W odróżnieniu od nich, demokratyczne studia ewaluacyjne biorą pod uwagę pluralistyczny charakter społeczeństwa i nakładają na ewaluatora „obowiązek demokratyzowania jego wiedzy" (tamże, s. 12).

Również inne prace MacDonalda świadczą o tym, że dostrzega on polityczne aspekty studiów ewaluacyjnych. Z uwagi na problematykę prace te można podzielić na cztery podstawowe grupy: (a) konceptualizacja i wykład założeń ewaluacji demokratycznej (MacDonald, 1974; 1984; i współaut., 1975; MacDonald, Norris, 1981); (b) teoria i metodologia studium przypadku (MacDonald, Parlett, 1973; MacDonald, 1970; 1977; 1980a; 1981; MacDonald, Walker, 1975); (c) rola ewaluacji w kształtowaniu i egzekwowaniu odpowiedzialności (MacDonald, 1978a; 1979); (d) problemy, jakie nastręczają próby studiów ewaluacyjnych na szczeblu ogólnokrajowym i międzynarodowym (MacDonald i współaut., 1979; 1981; 1982; MacDonald, Jenkins, 1980; MacDonald, 1982b).

Co do pierwszego z moich pytań, tj. pytania o to, kto powinien prowadzić studia ewaluacyjne — stanowisko MacDonalda jest zbliżone do poglądów innych teoretyków: ewaluacje powinny być domeną ewaluatorów profesjonalnych. Argumentacja MacDonalda jest tu jednak

nieco odmienna. Reguły demokratycznych studiów ewaluacyjnych wymagają odwołania się do ewaluatora zewnętrznego dlatego, że tylko on może podjąć się roli bezstronnego pośrednika w wymianie informacji ważnych dla zainteresowanych stron (MacDonald, 1974). Trzeba wszakże zauważyć, że w innych pracach MacDonald kładzie nacisk na rolę praktyków: widać to najwyraźniej wtedy, gdy wspiera koncepcję systematycznej samooceny nauczycieli jako narzędzie sprzyjające zwiększeniu publicznej odpowiedzialności poszczególnych szkół (MacDonald, 1978a).

Co do drugiego i trzeciego z moich pytań, to jest, pytań dotyczących kryteriów ocen i dostępu do wyników ewaluacji, MacDonald jest zdania, że w pluralistycznym społeczeństwie „ewaluator nie ma prawa wykorzystywać swojej roli do narzucania własnych wartości lub ideologii, które uznaje za prawomocne. Jego zadanie polega na wskazaniu tych, którzy będą musieli dokonać oceny programu i podjąć określone decyzje, oraz na przedstawieniu im faktów, które uznają oni za istotne z punktu widzenia stawianych sobie pytań" (MacDonald i współaut., 1975, s. 74).

Harlen, Eraut i Elliott, trzej następni ewaluatorzy, koncentrują się przede wszystkim na problematyce studiów ewaluacyjnych dotyczących szkół i m.in. z tego powodu rozwiązują problemy, jakie nastręczają postawione przeze mnie pytania, inaczej aniżeli teoretycy amerykańscy. W pierwszej z interesujących mnie kwestii uważają oni, że najlepszym ewaluatorem jest nauczyciel — co różni ich zasadniczo od ich amerykańskich kolegów. Wprawdzie Harlen rozpoczęła swoją zawodową biografię jako ewaluator profesjonalny (Harlen 1973; 1975), to jednak jej późniejsze prace dotyczące oceny postępów uczniów (Harlen 1976; 1977a; 1977b) oraz samooceny i publicznej odpowiedzialności poszczególnych szkół (Harlen 1979a; 1979b) wyraźnie ujawniają jej zainteresowanie doskonaleniem nauczycieli jako ewaluatorów i przekonanie, że to właśnie oni powinni być kluczowymi aktorami procesu ewaluacji. Ten punkt widzenia ma swoje konsekwencje w dążeniu do prezentowania celów i metod ewaluacji w języku, jakim posługują się nauczyciele.

Eraut nie zajmuje się bezpośrednio kwestią, kto powinien prowadzić studia ewaluacyjne, jednakże sądząc po znaczeniu, jakie przywiązuje do wspomagania w tym zakresie nauczycieli (Eraut, 1976; 1978; 1984b), wypada sądzić, że jest przekonany, iż — przynajmniej niekiedy — społeczna rola szkoły obejmuje również ewaluację. Jednakże teoretykiem, który kładł największy nacisk na zaangażowanie nauczycieli w realizację studiów ewaluacyjnych, był John Elliott. Początkowo (1976) —

w ślad za koncepcją „nauczyciela-badacza" proponowaną przez Stenhouse'a (1975) — postulował, aby nauczyciele sami analizowali własną pracę, ale z biegiem czasu propozycję tę rozszerzył także na analizę programów szkolnych. Na tym, między innymi, miała polegać publiczna odpowiedzialność szkół (Elliott, 1981a).

Cała ta trójka autorów kładzie nacisk na wspieranie *pojedynczych nauczycieli* w ich ewaluacji *procesów zachodzących w klasie*. W pewnych przypadkach dostrzega się tu wprawdzie konieczność łączenia studiów nad tymi procesami ze strukturą szkoły (Elliott, 1980; 1982; Eraut, 1984b), ale — ogólnie rzecz biorąc — szkoła jako swoista całość nie jest przedmiotem badania. Gdy Elliott przekonuje, że „samoocena w kontekście tego, co dzieje się w klasie, powinna stać się podstawą ogólnego procesu ewaluacji szkoły" (1979, s. 6), to widać, jaki nacisk kładł na rolę nauczyciela jako ewaluatora. Eraut i Harlen dochodzą do podobnych konkluzji mając na uwadze przede wszystkim korzyści, jakie z samoewaluacji nauczycieli mogą wynikać dla uczniów (Harlen, 1979b; Eraut, 1984b). Nie oznacza to, że omawiani tu autorzy wykluczają udział specjalistów w procesach ewaluacji, jednakże inaczej widzą ich rolę. Zadaniem specjalistów-ewaluatorów ma być wspomaganie nauczycieli w nabywaniu odpowiednich umiejętności (Harlen, 1979b), w prowadzeniu samodzielnych ewaluacji (Eraut, 1984b) i formułowaniu kryteriów ewaluacji (Elliott, 1976). Niemniej trzeba zdawać sobie sprawę, że są to poglądy bardzo odległe od tych opinii, według których studia ewaluacyjne powinny pozostać obszarem specjalistycznych analiz prowadzonych przez profesjonalistów.

Odpowiedź na pytanie o to, kto powinien ustalać kryteria ocen, nie ogranicza się już od nauczycieli. Oczywiście, stanowią oni jedną z grup oceniających wartość programów szkolnych, ale jednocześnie dostrzega się, że szkoła powinna komunikować się również z rodzicami, podatnikami czy pracodawcami i przekazywać im wszystkie informacje, jakich mogą potrzebować po to, aby wyrobić sobie niezależny osąd tych programów (Harlen, 1979b).

Stanowisko Eraута w tej sprawie jest bardziej złożone. Odrzuca on pogląd Scrivena, uznaje też, że alternatywa w postaci „ewaluacji wyjaśniającej" nie jest zadowalająca, bo nie rozwiązuje problemu konfliktu ocen, a sama koncepcja ma charakter konserwatywny (Eraut, 1984a, s. 31). Odrzucając wymienione stanowiska, Eraut zgłasza propozycję kompromisową: uznaje wprawdzie odpowiedzialność ewaluatora wobec bezpośredniego zleceniodawcy, a także szerszej publiczności, ale *jednocześnie* akcentuje konieczność respektowania wartości i standardów,

jakie reprezentują specjaliści od spraw edukacji: „W proces, który chciałbym nazywać *ewaluacją zróżnicowaną*, wbudowane jest przewidywanie, że najprawdopodobniej trzeba będzie liczyć się z kilkoma różnymi systemami wartości, a niektóre z nich mogą się nawzajem wykluczać. To powoduje, że materiał empiryczny powinien być organizowany wokół każdego z tych aksjologicznych punktów widzenia bez przyznawania któremuś z nich pierwszeństwa przed pozostałymi" (tamże, s. 32).

Ma to pomóc ewaluatorowi uniknąć formułowania uproszczonych rekomendacji i stawić czoło konfliktowi wartości, jaki ujawnia bardzo wiele studiów ewaluacyjnych. W ten sposób pytanie, kto powinien ustalać kryteria oceny, zostaje przerzucone na odbiorców, którym przedstawia się materiał empiryczny zgromadzony w toku ewaluacji, z tym wszakże zastrzeżeniem, że materiał ten jest organizowany wokół dostrzeganych przez ewaluatora opcji aksjologicznych. Eraut kładzie nacisk na to, że żadna z grup reprezentujących określone systemy wartości nie powinna mieć monopolu na formułowanie ostatecznych ocen dotyczących programów nauczania. Ten pogląd jest zbieżny ze stanowiskiem MacDonalda, z tą jednak różnicą, że Eraut proponuje wliczanie do zbioru opcji aksjologicznych także i tę, która wyraża stanowisko samego ewaluatora. Pod tym względem jest to pogląd zbliżony do propozycji wniesionych przez Stake'a.

Reakcja Elliotta na interesujące mnie teraz pytanie jest prosta i nie pozostawia żadnych wątpliwości. Ocena należy do nauczyciela lub — jeśli informacje o wynikach ewaluacji stanowią przedmiot dyskusji — do jego kolegów (Elliott, 1979). Zewnętrzni ewaluatorzy mają jedynie formułować swoiste metaoceny. W opracowaniach rozliczających publiczną odpowiedzialność szkoły prawo do formułowania sądów oceniających mają także rodzice, przedstawiciele instytucji edukacyjnych oraz reprezentanci społeczności lokalnej (Elliott, 1981b; Elliott i współaut., 1981).

Biorąc pod uwagę znaczenie, jakie przywiązują wskazani autorzy do ewaluacji prowadzonych przez nauczycieli, do dynamiki funkcjonowania klasy szkolnej oraz do ocen formułowanych przez uczniów, nie powinno nas zaskakiwać, że stosunkowo mało uwagi poświęcają oni kwestii dostępu do wyników studiów ewaluacyjnych. Z ich punktu widzenia nie jest to zagadnienie, które można byłoby rozpatrywać niezależnie od odpowiedzi na pytanie, kto ma prowadzić ewaluację i formułować kryteria oceny. Ewaluacja programów nauczania dokonywana w ramach szkoły ma swoich lokalnych adresatów, przed którymi szkoła ponosi odpowiedzialność, jednakże ewaluacja jest tu postrze-

gana przede wszystkim jako aktywność wewnątrz-organizacyjna, służąca przede wszystkim zawodowemu samodoskonaleniu się nauczycieli.

Elliott postuluje ograniczenie kręgu odbiorców rezultatów ewaluacji do środowiska lokalnego uzasadniając to tym, że dla lokalnych grup interesu zasadnicze znaczenie mają ewaluacje o charakterze transakcyjnym, przy czym „transakcje między tymi grupami i nauczycielami są regulowane przez wzajemne zaufanie i szacunek" (Elliott, 1981a, s. 242). Idzie on dalej i powiada, że „po to, aby nie pomniejszać przydatności rezultatów ewaluacji dla odbiorców lokalnych ewaluator może być zmuszony odmówić dostępu do wyników swojej pracy rozmaitym agendom administracji" (tamże). Ten radykalny pogląd wynika z przekonania, że „proceduralna neutralność studiów ewaluacyjnych w dziedzinie edukacji nie jest możliwa, skoro ich przydatność jest jednakowa dla wszystkich, a konsekwencje dla różnych grup odbiorców mogą być skrajnie przeciwstawne" (tamże).

W tym miejscu mojej prezentacji wrócę do okresu poprzedzającego rok 1975 po to, aby sformułować dwie konkluzje. Po pierwsze, sądzę, że udało mi się pokazać, iż w teorii studiów ewaluacyjnych z biegiem czasu punkt ciężkości przesuwał się po myśli Tylera z analizy tego, co dzieje się w klasie szkolnej, na ewaluację dużych programów zmian w systemach edukacyjnych, czemu towarzyszył rozwój społecznego ruchu reform oświatowych, a wreszcie na ewaluację instytucji, gdy próby zrozumienia oporu przed zmianami zaowocowały studiami przypadku oraz innymi formami ewaluacji uwzględniającymi procesy rozwojowe i całą złożoność dokonującej się zmiany społecznej. Studium *The Humanities Curriculum Project* (HCP), zrealizowane w latach 1968–1972, było jedną z pierwszych ewaluacji brytyjskich, w których dostrzeżono potrzebę zmiany orientacji metodologicznej, tj. przejścia od „logiki dowodzenia" do „logiki odkrycia" (Nisbet, 1976). W Stanach Zjednoczonych tę zmianę uwidacznia wprowadzona przez Stake'a koncepcja portretowania (Stake, 1977). Metodologiczna reorientacja tego rodzaju zwróciła uwagę na nowe zagadnienia i przyniosła wzrost świadomości politycznych i etycznych konsekwencji ewaluacji. Odroczone efekty tych zmian pojawiły się jednak znacznie później.

Po drugie, nie ulega wątpliwości, że przed rokiem 1975 teoretycy ewaluacji, z wyjątkiem House'a i MacDonalda, nie zajmowali się bezpośrednio politycznymi aspektami studiów ewaluacyjnych. Praktyczne przedsięwzięcia w tej dziedzinie nastręczały wystarczająco dużo problemów metodologicznych i metodycznych, toteż na nich przede wszystkim koncentrowała się uwaga badaczy. Wprawdzie świadomość na-

rastającego pluralizmu wartości zaczęła znajdować odbicie w różnych modelach studiów ewaluacyjnych proponowanych w obu krajach w początku lat siedemdziesiątych, podjęte też zostały pierwsze próby bardziej demokratycznej dystrybucji wyników ewaluacji, jednakże polityczne konsekwencje zmiany orientacji metodologicznej, a także odmienne stanowiska polityczne związane z alternatywnymi metodologiami jeszcze się nie ujawniły. Teoretyków ewaluacji przywiązanie do idei pluralizmu łączyło na tyle mocno, że dzielące ich różnice wydawały się stosunkowo mało istotne. To poczucie wspólnoty podzielanych poglądów dominowało zarówno na pierwszej (MacDonald, Parlett, 1973), jak i na drugiej (Adelman, Kemmis, Jankins, 1976) konferencji brytyjskich i amerykańskich ewaluatorów w Cambridge.

Jednakże już w 1979 r. kontekst dyskusji uległ zmianie i na trzeciej konferencji ewaluatorów, poświęconej omówieniu wyników, jakie mogli zaprezentować promujący i popierający koncepcję ewaluacji naturalistycznych, wcześniej osiągnięty konsens, praktycznie rzecz biorąc, załamał się. Konflikty, jakie niosły ze sobą próby praktycznego posługiwania się pluralistycznymi modelami studiów ewaluacyjnych, ujawniły nie tylko polityczne różnice między ewaluatorami „nowej fali", lecz także i to, że przetrwanie studiów ewaluacyjnych wymaga wzmocnienia ich podstaw metodologicznych i wypracowania systematycznych procedur badawczych (Jenkins i współaut., 1981; MacDonald, Norris, 1981).

Na początek lat osiemdziesiątych przypadła w Wielkiej Brytanii następna zmiana w ramach „naturalistycznej szkoły" studiów ewaluacyjnych. Upolitycznienie atmosfery wokół zmian edukacyjnych spowodowało, że nauczyciele zaczęli odczuwać potrzebę ewaluacji swojej działalności, a wcześniejszy nacisk na specjalistyczne studia ewaluacyjne zaczął słabnąć wobec postulatów ewaluacji prowadzonych przez praktyków. W tym czasie metodologiczne i polityczne postulaty „szkoły naturalistycznej" były już dość powszechnie akceptowane i składały się na podejście pozwalające realizować jej zamysły.

Dalej chcę poświęcić nieco uwagi teorii ewaluacji demokratycznej. Mam nadzieję, że uda mi się w ten sposób osiągnąć dwa cele. Po pierwsze, przedstawić czytelnikom polityczne tło problemów i trudności, wokół których koncentrowała się moja własna praca. Po drugie, zwrócić uwagę na odmienność racji, dla których naturalistyczne studia ewaluacyjne odróżnia się od typowych badań prowadzonych w naukach społecznych.

Teoria ewaluacji demokratycznej

W 1974 r. MacDonald opublikował trzy ważne prace, z których pierwsza, nosząca tytuł *Evaluation and the Control of Education* (Ewaluacja i kontrola edukacji) była już przywoływana. Druga, *Case Study and the Social Philosophy of Educational Research* (Studium przypadku i społeczna filozofia badań edukacyjnych), napisana wspólnie z Robem Walkerem, umacniała tradycję ewaluacyjnych studiów przypadku budowaną w Center for Applied Research in Education oraz omawiała podstawowe zasady modelu ewaluacji demokratycznej i — już niekoniecznie w związku z problematyką studiów ewaluacyjnych — argumentowała postulat dzielenia kontroli nad badaniami edukacyjnymi między badaczy i badanych. Trzecia praca, *Confidentiality: Procedure and Principles of the UNCAL Evaluation with Respect to Information about Projects in the National Development Programme in Computer Assisted Learning* (Poufność. Procedura i zasady ewaluacji programu UNCAL [Understanding Computer Assisted Learning] z uwzględnieniem informacji o projekcie Narodowego Programu Rozwoju w Nauczaniu Wspomaganym Komputerowo), napisana wspólnie z Bobem Stake'em, proponowała zestaw procedur pozwalających na takie zarządzanie danymi gromadzonymi w toku studiów ewaluacyjnych, aby reguły ewaluacji demokratycznych dało się zastosować do programów zmian edukacyjnych budowanych zgodnie z zasadami biurokratycznej racjonalności.

Pierwsza z tych prac jest jedną z najczęściej cytowanych i jednocześnie najmniej rozumianych pozycji w literaturze przedmiotu. Dość powszechne niezrozumienie jej podstawowych idei bierze się stąd, że rzadko jest czytana w kontekście innych teoretycznych tez dotyczących okoliczności wdrażania programów zmian edukacyjnych. Jednym z moich celów jest podjęcie próby pełniejszego zrozumienia tego tekstu.

Rozpocznę od dobrze znanej typologii studiów ewaluacyjnych, jaką zaproponował MacDonald. Co ciekawe, typologia ta odwołuje się do analizy teorii i praktyki amerykańskich ewaluatorów i opracowana została wtedy, gdy brytyjska teoria ewaluacji dopiero raczkowała, a odpowiednia praktyka nie wyszła poza stadium wczesnych prób. Niemalże wszystkie amerykańskie studia ewaluacyjne, utrzymuje MacDonald, można włączyć do pierwszych dwóch z trzech zaproponowanych przez niego kategorii. Krótko mówiąc, definiuje on ewaluacje „biurokratyczne" jako studia, które mają stanowić instrument utrzymania bądź rozbudowania władzy menedżerów oświaty. „Rzeczywistość władzy" stanowi niejawną przesłankę ewaluacji biurokratycznej — tak samo jak studiów ewaluacyjnych „na zamówienie". Jako „autokratyczne" MacDonald defi-

niuje te ewaluacje, które podtrzymują bądź poszerzają władzę badaczy akademickich wspierających autorytetem nauki decyzje w sprawach publicznych w zamian za ich zgodność z rekomendacjami uzasadnionymi przez studia ewaluacyjne. W tym przypadku ukryta racja odwołuje się do „odpowiedzialności stanowiska", a ewaluator staje się gwarantem profesjonalnej kompetencji urzędnika. Nietrudno dostrzec, co zresztą MacDonald wielokrotnie sam potwierdził, że model ewaluacji zaproponowany przez Stufflebeama może uchodzić za prototyp ewaluacji biurokratycznej, natomiast Scriven ma wszelkie cechy ewaluatora autokratycznego. Demokratyczne studia ewaluacyjne opisuje MacDonald w sposób następujący:

„Demokratyczne studia ewaluacyjne to służba na rzecz społeczności, którą informuje się o właściwościach programu edukacyjnego. Uwzględnia się tu pluralizm wartości podzielanych przez członków społeczności, a w toku formułowania podstawowych pytań badawczych bierze się pod uwagę możliwie szeroki zakres ich interesów. Podstawową wartością dla ewaluatora jest poinformowana opinia publiczna, a jego rola polega przede wszystkim na pośredniczeniu w wymianie informacji pomiędzy różnymi grupami. Jego techniki gromadzenia materiałów empirycznych i prezentowania wniosków, jakie z tych materiałów wynikają, powinny być zrozumiałe dla nie mających specjalistycznego przygotowania odbiorców. Ewaluator powinien przede wszystkim dowiedzieć się, jak ludzie reagują na określony program zmiany społecznej i jak go definiują. Uzyskiwane informacje opatrzone są klauzulą poufności, a informatorzy powinni mieć pełną kontrolę nad tym, jak ewaluator je wykorzystuje. Raport konkludujący studium ewaluacyjne nie powinien koncentrować się na rekomendacjach, a informacje nie mogą być przedmiotem manipulacji. Ewaluator systematycznie negocjuje wzajemne relacje ze zleceniodawcami i uczestnikami programu. Podstawowym kryterium jego sukcesu jest to, jak liczne jest grono odbiorców zainteresowanych wynikami jego pracy. Jego ambicją powinno być, by raport końcowy był «bestsellerem». Kluczowe idee, wokół których organizowana jest ewaluacja demokratyczna, do «poufność», «negocjacje» oraz «dostępność». Ich najważniejszą racją jest «prawo do informacji»" (MacDonald, 1974, s. 18).

Ta koncepcja ewaluacji demokratycznej sprowokowała wiele komentarzy ze strony teoretyków studiów ewaluacyjnych i wiele uwag krytycznych. House wskazywał, że w swoich założeniach politycznych nawiązuje do „klasycznego liberalizmu", bo wbudowany w nią obraz społeczeństwa przedstawia je jako zrzeszenie niezależnych jednostek

współpracujących ze sobą po to, aby łatwiej osiągać swoje egoistyczne cele (House, 1980, s. 62). Inni komentatorzy pisali, że jest to koncepcja „nieszczera" (Eraut, 1984a, s. 27), „manipulatorska" (Jankins, 1980), „humanistyczna" (Cronbach, 1982), a także „aksjologicznie indyferentna" i „pokorna" (Nilsson, Hogben, 1983).

Najostrzejszą krytykę wzbudziły zalecenia proceduralne (MacDonald, Stake, 1974; MacDonald, Walker, 1974), to jest przede wszystkim teza mówiąca, że „publiczność ma wprawdzie «prawo do informacji», ale powinno je moderować «prawo jednostek do poufności» [...] a granice publicznie dostępnych informacji należy ustalać w drodze negocjacji" (MacDonald i współaut., 1975, s. 12). Kwestie i problemy, jakie nastręczają tego rodzaju negocjacje, omówię później odwołując się do moich własnych doświadczeń praktycznych. Chcę natomiast zatrzymać się jeszcze chwilę przy zagadnieniu „politycznego liberalizmu" MacDonalda, jaki w pełni ujawnił się w jego następnych pracach. „Wystarczający warunek demokracji jest spełniony, jeśli władza i sposób jej dystrybucji w systemie społecznym uzależniony jest od obywateli. W tego rodzaju systemach społecznych podstawową zasadą jest funkcjonowanie poinformowanej opinii publicznej" (Jenkins i współaut., 1977, s. 25). W końcowych partiach pracy, potwierdzającej pogląd House'a, że rola ewaluatora jako pośrednika jest związana z demokratycznym sposobem sprawowania władzy, MacDonald i współautorzy cytują z aprobatą opinię Smitha z 1976 r. o roli władzy w demokratycznym społeczeństwie, sprowadzającej się do tego, że „zestaw neutralnych instytucji wspomaga poszukiwanie kompromisu między sprzecznymi interesami społecznymi i ekonomicznymi" (Jenkins i współaut., 1977, s. 35). Dalej pojawia się, moim zdaniem, kluczowa dla interesującego nas tu zagadnienia teza: „Do ewaluatora należy znalezienie sposobów realizacji tych «uzasadnionych» aspiracji nie tyle przez założenie neutralności państwa, ile przez potraktowanie jego retoryki poważnie i stosowanie jej jako kryterium oceny" (s. 35). Nieco później, w odczycie wygłoszonym na Uniwersytecie Stanu Illinois, MacDonald stwierdził:

„Rozpoczynam od postawy wobec społeczeństwa, w którym żyję i pracuję. Jest to społeczeństwo, w którym ideały pluralizmu i demokracji nie są spełnione w sposób doskonały, ale zarazem jest to społeczeństwo, które w imię politycznego przetrwania podejmuje widoczne wysiłki, aby wyeliminować praktyki dyskwalifikujące jego członków. Retoryka tego społeczeństwa (jego podstawowe wartości) jest dla mnie źródłem siły w zmaganiach z rzeczywistością, z którą nie mogę się pogodzić. Liberalizm rozumiem jako postulat maksymalizowania wła-

dzy jednostek, natomiast demokrację jako sprawowanie powierzonej im władzy przez poinformowane zbiorowości" (MacDonald, 1980a, s. 1).

Ponieważ MacDonald i jego współpracownicy z Centre for Applied Research in Education mieli wiele okazji do praktycznego sprawdzenia swojej orientacji politycznej w kontaktach z wpływowymi funkcjonariuszami administracji centralnej, można rozsądnie postawić pytanie, w jaki sposób poglądy te realizowane są w praktyce. Przytaczane dalej fragmenty referatu wygłoszonego w 1982 r. podsumowują pierwsze wysiłki MacDonalda dotyczące wdrożenia zasad ewaluacji demokratycznej. Kontekst realizacji tych zasad stanowiło studium ewaluacyjne dotyczące *National Development Programme in Computer Assisted Learning*, które zamówił u niego komitet składający się z przedstawicieli siedmiu agend administracji publicznej (MacDonald i współaut. 1975; MacDonald, Jenkins, 1980). Jest to podsumowanie trwającego blisko pięć lat sporu MacDonalda z jego zleceniodawcami.

1. Im zależało na ewaluacji prowadzonej wedle modelu cel – poziom realizacji. Rozpatrywane projekty musiały wyraźnie określać cele jeszcze przed przyznaniem na nie jakichkolwiek pieniędzy. Oni chcieli wiedzieć, czy cele te zostały osiągnięte.

My powiedzieliśmy, że to jest krzywdzące. Żaden z projektów nie mógł osiągnąć wszystkich wyznaczonych celów, co, według tego kryterium, musiałoby oznaczać, że każdy z nich zawiódł. Skłonni byliśmy przedstawić starania podejmowane w ramach każdego z projektów, aby komitet mógł ocenić, czy ich realizatorzy, uwzględniając wszystkie ograniczenia i okoliczności, robili co do nich należało.

2. Oni oczekiwali od nas rekomendacji mówiących, który z tych projektów należy kontynuować, któremu zaś odmówić dalszego finansowania.

My powiedzieliśmy, że taka decyzja należy do was i wy będziecie ponosić za nią odpowiedzialność. To nie jest nasz problem. Przeczytajcie nasze raporty i wypracujcie sobie w tej sprawie własną opinię. My jedynie pośredniczymy w przekazywaniu informacji — ułatwiamy jej wymianę między ludźmi, którzy chcą dowiedzieć się czegoś o sobie nawzajem.

3. Oni chcieli, abyśmy komentowali raporty składane na piśmie, abyśmy mówili o tym, co nie mogło znaleźć się w raportach.

My powiedzieliśmy: Nie, żadnych poufnych informacji. Nasze raporty były dyskutowane z ludźmi, których pracy dotyczyły, i zwlekaliśmy z ich przedstawianiem komitetowi do czasu, aż ludzie uznali je za uczciwe, istotne i dobrze pokazujące stan rzeczy. Odmówiliśmy dodania do nich czegokolwiek.

4. Oni twierdzili, że nasze raporty są zbyt długie i zbyt złożone, aby członkowie komitetu byli w stanie się z nimi zapoznawać. Oczekiwali, że będziemy je streszczać.

My powiedzieliśmy, że nasze raporty są na tyle zwięzłe, na ile dopuszczały to negocjacje z ludźmi, których praca była ewaluowana. Traktowaliśmy naszą pracę jako świadczenie na rzecz trafnego osądu, nie zaś na rzecz ludzi, którzy mieli tego osądu dokonywać.

Oni nalegali twierdząc, że nie mogą sobie dać rady ze złożonością naszych raportów, uniemożliwiającą podjęcie decyzji. My replikowaliśmy, że to zależy już wyłącznie od członków komitetu.

5. Oni określili to jako próbę ewaluowania pracy komitetu. Macie ewaluować nie nas, powiadali, lecz poszczególne projekty.

My odpowiadaliśmy, że — choć niechętnie — nie możemy tego uniknąć. Poprzestanie na ewaluacji projektów byłoby nieuczciwe. Realizatorzy poszczególnych projektów — mówiliśmy — chcą wiedzieć, jakim trybem toczy się wasza praca, czy porządnie ją wykonujecie i czy wywiązujecie się z postawionego przed wami zadania.

6. Oni sugerowali, że przewróciło się nam w głowie i że — skoro bierzemy od nich pieniądze — powinniśmy wykonywać to, co cię nam każe.

My odpowiadaliśmy, że jesteśmy niezależnymi, demokratycznymi ewaluatorami, a z tego, że opłacają nasz wysiłek, nie wynika, że mają jakieś szczególne prawo do jego rezultatów lub do narzucania kierunków naszej działalności.

Nie można kupić ewaluacji, można ją jedynie zlecić. Każdy ma prawo proponować zagadnienia, jakie mają zostać objęte jej zakresem, natomiast nikt nie ma prawa domagać się informacji, jeśli nie jest skłonny dzielić się informacją z innymi.

Te spory z biurokratami były bardzo trudne i zabierały wiele czasu. Toczyły się przez pięć lat. Mówię o tym tutaj po to, aby pokazać kierunek argumentacji przeciwstawiającej się założeniom przyjmowanym przez biurokratów, uzasadnionym bardziej przez zwyczaj i władzę, w jaką zostali wyposażeni, niż przez odpowiedzialność i rozsądek. (MacDonald, 1982a, s. 8-10).

Nawet jeśli uznać, że MacDonald z pewnym rozbawieniem traktował swoje boje z biurokracją, to jednak nie powinno umknąć naszej uwadze to, że negocjacyjna retoryka liberalnej demokracji posłużyła mu do wypracowania strategii pozwalającej na zrównoważenie stosunków władzy, w które uwikłana była realizacja ewaluowanych programów. Następny cytat potwierdza tę opinię i łączy ją z ideą naturalistycznych studiów ewaluacyjnych. Przywoływana dalej myśl zawarta była w podsumowaniu referatu na temat demokratycznych studiów ewaluacyjnych, wygłosznym na konferencji w Australii:

„Nie ulega wątpliwości, że podstawowym zagadnieniem stojącym przed studiami ewaluacyjnymi jest ich wpływ na dystrybucję władzy i — co stanie się jasne pod koniec mojej wypowiedzi — kwestia języka. Ewaluatorzy, którzy wybierają technokratyczną formę dyskursu, wzmacniają imerium menedżerów. Ewaluatorzy wybierający język nauk społecznych wspierają samych siebie lub środowiska akademickie. Ewaluatorzy, którzy wybierają język potoczny, służą nam wszystkim. Ja jestem ewaluatorem operującym językiem potocznym, jakimi ewaluatorami jesteśmy wy?" (MacDonald, 1984).

Dotychczas analizowałam logikę demokratycznego modelu studiów ewaluacyjnych, zastanawiając się nad tym, w jakiej mierze ewaluacje prowadzone zgodnie z jego założeniami mogą wpłynąć na stosunki władzy charakterystyczne dla centralnie finansowanych interwencji społecznych. Z tego punktu widzenia problem sprowadza się do politycznej nierówności między zleceniodawcami i administratorami programu

zmiany społecznej a wszystkimi pozostałymi jego uczestnikami, uzależnionymi od aprobaty tych, którzy kontrolują decyzje finansowe. Kontekst tego problemu stanowi zatem władza wyrażająca się w decyzjach finansowych. Demokratyczny ewaluator funkcjonujący w takim kontekście stara się o uzyskanie niezależności, o odrzucenie instrumentalnej formuły swojej roli profesjonalnej z wbudowanym w nią posłuszeństwem wobec struktury władzy i zastąpienie jej rolą pośrednika w tym, co MacDonald nazywa „wymianą dóbr" (MacDonald, Sanger, 1982). Dobrem jest tu, oczywiście, wiedza o programie zmiany społecznej, znacznie bardziej egalitarnie dystrybuowana aniżeli władza. Zasada, na której opiera się ta koncepcja, mówi, że „ludzie są właścicielami faktów konstytuujących ich życie" (Walker, MacDonald, 1975). Jest to zasada etyczna przekształcona przez MacDonalda w strategię polityczną.

Jednakże chciałabym też zastanowić się nad takim kontekstem, w którym problem władzy w mniejszym stopniu wiąże się ze stosunkami władzy w strukturach politycznych, w większym zaś ze stosunkami między środowiskiem naukowym i społecznością stanowiącą przedmiot badań. Idea demokratycznych studiów ewaluacyjnych może być interpretowana również jako atak na autorytaryzm nauki i jako strategia zmierzająca do redukcji nierówności politycznych właściwych stosowanym naukom społecznym. Warto zwrócić uwagę na to, że część autorów uznających za możliwe do zaakceptowania „biurokratyczne" i „demokratyczne" modele ewaluacji kwestionuje jednocześnie sensowność modelu „autokratycznego". Na przykład Cronbach uznał, że lepszym terminem dla nazwania tego rodzaju studiów ewaluacyjnych byłoby określenie „autorytatywne" (Cronbach, 1982), Adelman i Alexander (1982) sugerowali termin „autonomiczne". Klucz do zrozumienia logiki demokratycznego modelu badań można znaleźć w drugiej z wyżej wspomianych prac MacDonalda (Studium przypadku i społeczna filozofia badań edukacyjnych). Przyjmując za punkt wyjścia model ewaluacji demokratycznej, Walker i MacDonald atakują to, co nazwali „autokratyczną tradycją" — tradycję badań zorientowanych na rozwój teorii naukowej, badań, których wyniki adresowane są przede wszystkim do środowisk akademickich, prowadzonych w oparciu o założenie, że badaczom przysługuje niekwestionowane prawo do interpretowania świata społecznego. Autorzy omawianej pracy proponują listę zagadnień, które mogą stanowić dobrą podstawę do kwestionowania tej tradycji:

— Czyje potrzeby i interesy stanowią bodziec do prowadzenia określonych badań naukowych?

— Kto jest właścicielem danych badawczych (badacz, przedmiot badań, zleceniodawca)?

— Kto ma dostęp do tych danych (komu odmawia się dostępu do danych)?

— Jaki status przysługuje proponowanej przez badacza interpretacji badanych faktów w zestawieniu z ich innymi interpretacjami (do kogo należy decyzja, kto mówi prawdę)?

— Na czym polegają zobowiązania badacza wobec badanych obiektów, wobec zleceniodawców, środowisk naukowych i szerszej publiczności?

— Dla kogo prowadzone są badania? (MacDonald, Walker, 1975, s. 6).

W dalszych partiach tej pracy autorzy formułują propozycję charakterystyczną dla demokratycznej alternatywy wobec „autokratycznego" modelu badań naukowych: „Uważamy za możliwe zastanowienie się nad taką formą badań edukacyjnych, która wiązałaby je z praktyką inaczej aniżeli dotychczas. Idzie nam tu o badania nawiązujące bezpośrednio do formułowanych przez praktyków definicji sytuacji, struktur pojęciowych i języka. Ograniczany przez nie — badacz stawałby się reprezentantem różnych grup składających się na badaną społeczność, jego zadanie polegałoby na testowaniu ich hipotez, formułowanych w ich języku i w ich strukturach pojęciowych, traktując je jako punkt wyjścia, ale też jako stały układ odniesienia. Celem takich badań byłoby nie tyle kreowanie rzeczywistości alternatywnej dla badanych, ile poszukiwanie sposobów zachęcenia ich do wglądu w ich istniejące rzeczywistości i rozumienia rzeczywistości innych mieszkańców społecznego świata" (tamże, s. 9).

Konkluzja tej argumentacji brzmi następująco: „Rzeczywista korzyść polega na szansach stworzenia technik i procedur, które znajdą zastosowanie w szkołach i instytucjach wspierających system edukacyjny. Wyspecjalizowani, profesjonalni badacze nigdy nie będą mogli zastąpić samo-obserwującej się uważnie społeczności nauczycieli" (s. 11).

Jedna z zasadniczych propozycji formułowanych w tej pracy pod adresem społeczności naukowej dotyczy poufności: „Informatorom należy zagwarantować poufność w toku badania, a później negocjować z nimi zgodę na opublikowanie zebranych materiałów. [...] To dzielenie się prawem do kontroli nad materiałem empirycznym z uczestnikami badań oznacza, że badacz musi liczyć się z utratą swoich danych lub poprzestaniem na ich archiwizacji" (tamże, s. 10).

Według jednego z obserwatorów (Elliott, 1984), myśl ta, przedstawiona na dorocznej konferencji British Educational Research Association w 1975 r., spotkała się z bardzo negatywną reakcją audytorium. Na konferencji w Cambridge w 1975 r. Becker zauważył, że trudno byłoby respektować taką procedurę w przypadku badań dotyczących skorumpowanego komisariatu. MacDonald, replikując, utrzymywał, że to zależy od tego, jak zakreśli się granice badanej sytuacji. Przywołując współcześnie tę kontrowersję warto zauważyć, że zgoła niedawno MacDonald zakończył studium ewaluacyjne poświęcone szkoleniu policjantów (MacDonald i współaut., 1987).

Nietrudno dostrzec, że logika działania ufundowana na demokratycznym modelu ewaluacji zmienia sie w zależności od warunków, w jakich ma być zastosowana. Trudno się temu dziwić, pamiętając, że jest to model oparty na ewaluacjach stanowiących studia przypadku. Zarazem jednak pozwala to zrozumieć te racje krytyczne, które elastyczność tego modelu i sprzęgniętych z nim strategii badawczych traktowały jako stwarzanie przesłanek do manipulacji. Powstają tu również i inne interesujące pytania, m.in. o relacje między demokratycznymi studiami przypadku a edukacyjnymi badaniami etnograficznymi czy antropologicznymi, które również rozwijały się bardzo żywo w latach siedemdziesiątych.

Z analizy prac MacDonalda jasno wynika, że — jego zdaniem — zmiana orientacji metodologicznej bez zmiany struktury stosunków władzy w istocie niewiele zmieni. Odnajdujemy tę myśl, na przykład, w referacie napisanym wspólnie ze współpracownikami: „Stworzenie dającej się obronić alternatywy studiów ewaluacyjnych to nie jest kwestia wyboru pomiędzy, powiedzmy, zaletami orientacji psychometrycznej i orientacji antropologicznej — choć właśnie tego rodzaju wybory koncentrowały na sobie uwagę wielu badaczy [...] relacje między studiami ewaluacyjnymi a systemami edukacyjnymi nie są kwestią wyboru modelu badawczego którejś z dyscyplin. Wybór narzędzi badawczych jest tylko konsekwencją wcześniejszego wyboru politycznego" (Jenkins i współaut., 1977, s. 24).

MacDonald krytykował również socjologiczne studia przypadku, m.in. badania tego typu prowadzone przez Nasha (1973) oraz Sharpa i Greena (1975), uznając je za przykłady „społecznej i etycznej nieadekwatności" tradycji badawczej w naukach społecznych (MacDonald, 1980b). Najogólniej mówiąc, krytykował on tradycję badań terenowych, które — jak utrzymywał w swojej pracy napisanej wspólnie z Norrisem — tylko obiecują, że w niczym nie zaszkodzą badanym,

gdy tymczasem traktują ich jedynie jako pozbawione podmiotowości „obiekty badania". „Badacz terenowy jako ewaluator pośrednio lub bezpośrednio bierze na siebie odpowiedzialność za korzyści lub szkody, jakich przysporzy ludziom stanowiącym przedmiot jego badania" (MacDonald, Norris, 1981, s. 19). Stąd wniosek: „[...] jeśli ludzie, których badamy, są pozbawieni wpływu na to, jak sformułowany zostaje problem badawczy, kto kontroluje dane empiryczne i wykorzystuje informacje zgromadzone w toku badań [...] przyczyniamy się do utrwalania obiektywnych warunków podtrzymujących przekonanie, że ludzie pozbawieni władzy są posłusznymi marionetkami w rękach polityków" (tamże, s. 23).

Podsumowanie uwag dotyczących politycznych aspektów teorii ewaluacyjnych

Konteksty mają swoje znaczenie. W połowie lat sześćdziesiątych wpływ władz federalnych na programy amerykańskich szkół publicznych spowodował przesunięcie akcentów: dotychczasowy nacisk na osiągnięcia najzdolniejszych uczniów (wymuszony przez „rosyjski syndrom" z lat sześćdziesiątych) ustąpił miejsca zwiększaniu szans edukacyjnych nieuprzywilejowanych grup społecznych (dążeniu do edukacyjnych kompensacji). Ta koncentracja publicznej uwagi na nieuprzywilejowanych warstwach społeczeństwa coraz wyraźniej odciskała ślad na świadomości teoretyków ewaluacji, zwłaszcza że kompensacyjne wysiłki bardziej ujawniały niż zmieniały relacje między tymi, którzy zyskują, i tymi, którzy tracą w tzw. „społeczeństwie otwartym". Taki był sens argumentów House'a (1974), gdy pisał o systemie edukacyjnym jako „instytucji zamrożonej" (s. 5), i krytyk MacDonalda dotyczących „funkcjonalnej dyskryminacji w przemądrzałych demokracjach" (MacDonald i współaut., 1982, s. 2). W Wielkiej Brytanii w latach sześćdziesiątych zmiany w systemie edukacyjnym wiązały się z podniesieniem wieku uczniów opuszczających szkoły. Wprowadzenie tych zmian planowano na rok 1970, a udało się je zrealizować w roku 1972; uczniowie zaczęli opuszczać szkoły w wieku lat 16, a nie — jak poprzednio — 15. W obu krajach szerszy kontekst tych zmian stanowiła dynamika gospodarcza i stosunkowo łatwo dostępne fundusze dla tych, którzy chcieli projektować nowe programy szkolne przeznaczone dla uczniów osiągających słabe wyniki, głównie dzieci należących do mniejszości etnicznych w Stanach Zjednoczonych i robotniczej młodzieży w Wielkiej Brytanii. Nie powinno nas dziwić, że w obu krajach przynajmniej część

z tych, którzy wzięli na siebie obowiązek monitorowania i analizowania programów zmiany w systemie edukacyjnym, reagowała również na polityczne okoliczności powstawania i realizacji tych programów. I gdy — w obu krajach — „wojnę o zaspokojenie aspiracji zastąpiła wojna o dobrobyt" (tamże), a wzrost publicznych wydatków na edukację ustąpił polityce nakładów celowych na programy, w których uczestnictwo miało charakter obowiązkowy, pilnym problemem stała się kwestia właściwego określenia roli ewaluatora, między innymi dlatego, że coraz trudniej było programy socjalne definiować jako filantropię zmierzającą do redystrybucji części dochodu narodowego z korzyścią dla tych, którym powodziło się najgorzej.

W swoim przeglądzie zajmowałam się, w gruncie rzeczy, różnymi reakcjami ewaluatorów na te zagadnienia, choć muszę przyznać, że nie wspomniałam dotychczas o jednym z najważniejszych wymiarów, w jakich można je rozpatrywać. Tacy teoretycy studiów ewaluacyjnych, jak Cronbach, House, Stake i MacDonald, są również teoretykami planowej zmiany społecznej, uwikłanymi w analizy trzydziestoletniej historii reform programów szkolnych. Dalecy od obojętności wobec sukcesów i porażek tych reform, wiązali swoje propozycje dotyczące modeli studiów ewaluacyjnych z warunkami doskonalenia systemów edukacyjnych.

Zaproponowane przez MacDonalda trzy kategorie ewaluacji można analizować jako jego stanowisko wobec współzawodniczących ze sobą poglądów na sposób poprawy jakości kształcenia. Jego propozycje dotyczące modelu ewaluacji demokratycznej oznaczają zarazem, że jest skłonny wspierać te mechanizmy doskonalenia programów szkolnych, które odwołują się do opinii społeczności lokalnych, akceptują różnorodność i wzajemne zobowiązania. Uczciwość, rozsądek i szacunek dla jednostek to podstawowe wartości sprzęgnięte z takimi poglądami w takim samym stopniu, w jakim stosunki budowane w oparciu o dominację i instrumentalny sposób myślenia są z nimi sprzeczne. Nie ulega wątpliwości, że MacDonald chce widzieć w retoryce liberalnej demokracji te wartości właśnie, a nie „maszynę do produkcji uzasadnień" (McPherson, cyt. za MacDonald, Norris, 1981, s. 5), i szanując ideę demokracji szuka sposobu na zatrzymanie tej maszyny. Równie oczywiste jest, że usiłuje on uświadomić środowiskom naukowym ich moralną odpowiedzialność.

Moim zdaniem, atrakcyjność modelu demokratycznych studiów oceniających tkwi w ich logice edukacyjnej, nie zaś w jej opozycyjności wobec struktur politycznych. Zgadzam się z MacDonaldem w tym, co

pisze o zasadniczych celach ewaluatora, polegających na ujawnianiu „potencjału edukacyjnego", i o procesie ewaluacji jako wspólnym zadaniu. Co do studiów przypadku dotyczących poszczególnych szkół, na których skupiała się dotychczas moja działalność, skłonna jestem przyjąć jego tezy dotyczące procesu, obiektu zainteresowań i wartości:

„Gdy ubiegamy się o dostęp do ludzi [...] powiedzmy im, że zakładamy, iż wszystkie działania społeczne stanowią swoisty kompromis między wartościami, okolicznościami i interesami, zarówno naszym, jak i ich zadaniem jest nie tyle obrona czy atakowanie tego kompromisu, ile dokładne zrozumienie jego złożonej struktury. Nie jest to obietnica, lecz zaproszenie do powiązania aktu ewaluacji z ewaluowaną rzeczywistością" (MacDonald, 1981, s. 6).

Skłonna jestem interpretować tę myśl jako określenie edukacyjnej i zarazem emancypacyjnej roli studiów ewaluacyjnych. Dla MacDonalda oznacza to stymulowanie refleksyjności na wszystkich poziomach systemu edukacyjnego. Dla mnie stanowi układ odniesienia dla gromadzenia wiedzy o szkole w przekonaniu, że dzieląc się tą wiedzą wspieram jej uzasadnione dążenia do samodoskonalenia praktyki edukacyjnej.

tłumaczył *Kazimierz Frieske*

Rozdział III

Różnorodność podejść i ujęć

Ernest R. House

Ewaluacja i jej uprawomocnianie
Główne podejścia

Ernest R. House, „Approaches to Evaluation. The Major Approaches", rozdział z książki tegoż autora *Evaluating with Validity*, Sage, Beverly Hills - London 1980.

Istnieje wiele podejść do ewaluacji i każde ma swoich zwolenników. Jednak większość z nich można pogrupować, przyporządkowując kilku podstawowym typom. Typy te nazywane są niekiedy „modelami", co wynika z faktu, iż są one projektami lub typami o charakterze strukturalnym, za którymi opowiadają się wybitni teoretycy. Stąd uważa się, że są one wzorami, które warto naśladować. Można też nazywać je po prostu „podejściami".

Podstawowe modele mają swoich powszechnie znanych propagatorów oraz wyraźnie artykułowaną racjonalność. Są omawiane w literaturze. Ponadto identyfikują się z nimi grupy praktyków, którzy — co nie najmniej ważne — dokonują w oparciu o nie rzeczywistych ewaluacji. Modele same w sobie stanowią idealizacje określonych podejść do ewaluacji. Każda rzeczywista ewaluacja jest jednak wypadkową wielu różnych okoliczności; w konsekwencji, poszczególne ewaluacje mogą przybierać odmienne kształty, nawet jeśli ich punktem wyjścia jest ten sam partykularny typ. Innymi słowy, model stanowi typ idealny.

Warto także zauważyć, że wielu wybitnych teoretyków ewaluacji — na przykład Cronbach, Campbell i Glass — nie identyfikuje się z jakimś jednym partykularnym podejściem do ewaluacji. Z kolei inni, jak na przykład Guba, opowiadali się w swoich pracach za dwoma różnymi podejściami. W ewaluacjach, które przeprowadzałem osobiście, wykorzystywałem wszystkie główne podejścia. Modeli nie powinno się więc uznawać za „własność" jednej osoby ani traktować jako punkt wyjścia klasyfikowania badaczy.

Tabela obok przedstawia taksonomię głównych modeli ewaluacji. Definiując je, wykorzystałem w znaczącym stopniu wcześniejsze klasyfikacje, szczególnie Stake'a (1976), Popham (1975) oraz Worthena i Sandersa (1973). Bez wątpienia pewne typy zostały przeze mnie w sposób nie zamierzony pominięte. Inne wykluczyłem świadomie, jako że nie stały się jeszcze na tyle popularne, aby można było uznać, iż tworzą szkołę mającą wpływ na praktykę. Modele, które uwzględniam, uosabiają główne wykorzystywane obecnie w praktyce podejścia. W rozdziale tym zamieszczam też krótki opis każdego podejścia.

Analiza systemowa

Najpierw definiuje się kilka mierników uzyskiwanego wyniku (*output*), w rodzaju punktowanych testów wyników nauczania, a następnie próbuje się odnieść różnice w programach lub kierunkach prowadzonej polityki do obserwowanych zmian wskaźników. Dane mają charakter ilościowy i pomiary wyników odnosi się do programów za pomocą analiz korelacyjnych lub innych technik statystycznych. Ostatnio coraz częściej stosowano eksperymenty i większą wagę przywiązywano do „kontrolowanej zmiany" programu niż do zmiany „naturalnej" (Cooley, Lohnes, 1976).

Jednym z głównych źródeł tego podejścia były analizy systemowe prowadzone w Departamencie Obrony (Department of Defense, DOD) w okresie, gdy kierował nim McNamara. Od około roku 1965 podejście to uznane zostało za podstawową perspektywę ewaluacyjną w Departamencie Zdrowia, Edukacji i Opieki Społecznej (Department of Health, Education and Welfare, DHEW). W tym też roku prezydent Johnson utworzył w rządzie federalnym ośrodek Planowania, Programowania i Systemu Budżetowego (Planning, Programming, and Budgeting System, PPBS). Gdy Departamentem Zdrowia, Edukacji i Opieki Społecznej kierował John Gardner, stworzono stanowisko asystenta sekretarza do spraw ewaluacji programów (Assistant Secretary for Program Evaluation, ASPE). Do pracy w jego biurze przyjęto grupę ekonomistów i księgowych kierowanych przez Williama Gorhama, który wcześniej zdobył bogate doświadczenie we wspomnianym PPBS w Departamencie Obrony. Zastępcami Gorhama zostali Robert Grosse i Alice Rivlin, która następnie objęła stanowisko asystenta sekretarza oraz dyrektora Kongresowego Biura Budżetowego (Congressional Budget Office).

Taksonomia głównych podejść ewaluacyjnych

Model	Główni adresaci lub grupy odniesienia	Dziedzina wyjściowego konsensu	Metodologia	Rezultat	Typowe pytania
Analiza systemowa	ekonomiści, menedżerowie	cele, znane przyczyny i skutki, zmienne ilościowe	PPBS, programowanie, kontrolowana zmiana, analizy kosztów – korzyści	wydajność	czy osiąga się oczekiwane wyniki? czy można je osiągnąć bardziej ekonomicznym sposobem? które programy są *najbardziej* efektywne?
Zoperacjonalizowanych zadań	menedżerowie, psycholodzy	z góry określone cele, zmienne ilościowe wyników	cele behawioralne, testy osiągnięć	produktywność, przewidywalność	czy program osiąga założone cele? czy wywołuje zamierzone skutki?
Zorientowany na podjęcie decyzji	decydenci, zwłaszcza administratorzy	cele ogólne, kryteria	badania kwestionariuszowe, wywiady jakościowe, obserwacja zmian naturalnych	efektywność, kontrola jakości	czy program jest efektywny? które jego części są efektywne?
Goal-free	konsumenci	konsekwencje, kryteria	kontrola stronniczości, analizy logiczne, *modus operandi*	wybór konsumencki, użyteczność społeczna	dotyczą *wszelkich* skutków
Krytyka artystyczna	znawcy, konsumenci	krytycy, standardy	krytyczne omówienia	podniesienie standardów, wzrost świadomości	czy krytycy zaaprobowaliby program? czy wzrasta uznanie odbiorców?
Inspekcja kwalifikacji	profesjonaliści, społeczeństwo	kryteria, zespół oceniających, procedury	inspekcje, samoocena	akceptacja ze strony specjalistów	jak oceniliby program profesjonaliści?
Quasi-sądowy	jury	procedury i sędziowie	procedury *quasi*-sądowe	postanowienie	jakie są argumenty za, a jakie przeciw programowi?
Studium przypadku	klienci, praktycy	negocjacje, działania	studia przypadku, wywiady jakościowe, obserwacje	zrozumienie różnorodności	jak zapatrują się na program różni ludzie?

„Model menedżerski propagowany i dopracowany przez Roberta McNamarę w DOD i zastosowany przez Williama Gorhama w DHEW czerpał z teorii mikroekonomicznej. Podobne perspektywy teoretyczne analitycy ASPE próbowali wykorzystać w dziedzinie usług świadczonych ludności (*human service areas*). Projekt TEMPO Tytuł I zawierał *implicite* kilka założeń właściwych patrzeniu na mechanizm świata przez pryzmat ekonomii. Podstawowym założeniem jest, że jednostki i organizacje zachowują się w sposób umożliwiający maksymalizację pewnych identyfikowalnych rezultatów lub serii rezultatów. W związku z tym analityk powinien umieć odpowiednio modelować wybory organizacyjne oraz uzasadniać zarówno pożądane cele, jak i względną efektywność różnych strategii mogących służyć ich realizacji. Ta struktura analityczna zakłada stałą produktywność, regularną i wymierną zależność między wkładem w działanie a jego końcowymi rezultatami" (McLaughlin, 1975, s. 35).

Przebieg ewaluacji programu ESEA Tytuł I (dostarczającego środków dla dzieci upośledzonych) stanowi przykład wykorzystania podejścia opartego na analizie systemowej. Biuro Ewaluacji (The Office of Evaluation) zarządziło, aby stosowano je we wszystkich 30 000 projektów Tytułu I, służących pięciu milionom dzieci. Dwóch działaczy HEW wyjaśnia istotę tego podejścia w sposób następujący:

„Dla wypełnienia swego ustawowego mandatu Biuro Edukacji Stanów Zjednoczonych (U.S Office of Education) podpisało kontrakt z Korporacją Badawczą RMC w celu opracowania modeli ewaluowania rozwoju podstawowych umiejętności [dzieci] w wyniku działania lokalnych projektów Tytułu I. Każdy z trzech stworzonych przez RMC modeli zawierał: testowanie przed podjęciem działań i po ich zakończeniu; metodę szacowania konsekwencji, w przypadku gdy projekt Tytułu I nie funkcjonował; procedury transformacji wyników testów lokalnych w równoważniki „krzywej normalnej", co miało pozwolić na ich agregację (Barnes, Ginsburg, 1979,7 s. 7).

W książce „Systematyczne myślenie na rzecz działania społecznego" (*Systematic Thinking for Social Action*) Rivlin (1971) pisze, że problemem podstawowym była kwestia, co zrobić, aby decyzje rządu mogły być lepsze. Uznała ona, iż można postawić w tym kontekście następujące kluczowe pytania: (1) W jaki sposób definiuje się problemy społeczne i jak się one przejawiają? (2) Kto i w jakim stopniu skorzystałby z proponowanych rozwiązań? (3) Co przyniosłoby najwięcej korzyści i w jaki sposób można porównać korzyści wynikające z różnych podejść? (4) Jaki sposób realizacji usług jest najbardziej efektywny? Dwóch ostatnich py-

tań miała dotyczyć ewaluacja. Istotą tego podejścia są analizy kosztów i korzyści porównywanych programów. Jako przypadek paradygmatyczny wymienia się w tym kontekście ewaluację *Follow Through*.

Genralnie zakłada się konsens w sferze celów. Problem polega na wyborze sposobów mierzenia rezultatów programów rządowych. Określenie zależności przyczynowo-skutkowych między programami i rezultatami wymaga zaprojektowania dobrego eksperymentu. Warto przy tym mieć losową grupę kontrolną, jednak nie zawsze jest to możliwe. Pomiar rezultatów programów metodami psychometrycznymi, socjometrycznymi lub ekonometrycznymi pozwala na porównywanie ich kosztów w celu ustalenia, jakie rezultaty można uzyskać przy możliwie najmniejszych kosztach. W procesie tym wykorzystuje się w dużej mierze metodologię nauk społecznych.

Podejście opierające się na analizie systemowej opisują Rossi, Freeman i Wright w *Evaluation: A Systematic Approach* (1979). Uważają oni, że ewaluację podejmuje się w celach związanych z zarządzaniem, planowaniem, ustaleniem kierunku działań oraz zagadnieniami fiskalnymi. Wymieniają następujące pytania kluczowe (s. 20):

1) Czy działania interwencyjne objęły swoim zasięgiem założoną populację?
2) Czy przeprowadzano je we właściwy sposób?
3) Czy są one efektywne?
4) Jaki jest ich koszt?
5) Jak się mają koszty do uzyskanych efektów.

Ewaluacja musi być „możliwie najbardziej obiektywna i dostarczyć rzetelnej oceny — oceny, która powinna się potwierdzić, jeśli zostanie dokonana przez inną grupę lub powtórzona przez tych samych ewaluatorów" (s. 21). Ewaluatorzy zobowiązani są „w sposób uporządkowany ocenić kierunki rozwoju działań społecznych i ich wcielanie w życie". Podstawą systematycznych ewaluacji są techniki stosowane w naukach społecznych. Rossi i jego współpracownicy zakwestionowali zasadność przeprowadzania ewaluacji o charakterze zdroworozsądkowym. Prowadzi ona bowiem do formułowania błędnych wniosków. Autorzy ci odrzucali również ewaluacje oparte na standardach konkretnych profesji, ponieważ nie są poparte dowodami naukowymi. Podobnie orzeczenie kliniczne nie dostarcza „solidnego" punktu wyjścia do podjęcia decyzji.

Twierdzi się, że systematyczne ewaluacje dostarczają przekonywających i rzetelnych dowodów, które mogą uzyskać powtórnie również inni obserwatorzy. Informują także o efektywności wykorzystania funduszy. Celem wszechstronnej ewaluacji jest dostarczenie odpowiedzi na pyta-

nia dotyczące planowania i kontrolowania programu, oszacowania jego skutków i wydajności ekonomicznej. Informacje na temat planowania pomagają podjąć odpowiednie interwencje. Weryfikacja systemu informacji pozwala stwierdzić, czy program zgodny jest ze swoim projektem i czy spowodowana przez program zmiana następuje w pożądanym kierunku. Z kolei informacje odnoszące się do strony ekonomicznej umożliwiają określenie wydajności programu. Rossi, Freeman i Wright ostrzegają przy tym, że żadna ewaluacja nie spełni pokładanych w niej nadziei, jeśli kierunek działań nie został wytyczony jasno i wyraźnie.

Podejście oparte na analizie systemowej jest w znaczącym stopniu zorientowane na metodologię pozytywistycznych nauk społecznych (co wyklucza zwykle inne metodologie). Zostało ono zaakceptowane przez wiele osób zajmujących wysokie stanowiska w rządzie federalnym (zwłaszcza przez ekonomistów). Ponadto mieni się częścią nauk społecznych, co ma być potwierdzeniem jego autorytetu.

Jak już stwierdzono wyżej [patrz *Ewaluator w społeczeństwie*], podstawowy problem współczesnego ewaluatora wynika z faktu, że musi ewaluować programy publiczne. Rozwiązaniem tego problemu w przypadku podejścia opartego na analizie systemowej jest zaaprobowanie perspektywy decydentów na poziomie federalnym oraz przyjęcie założenia, że badany program stanowi funkcjonalną część struktury społecznej i rządowej. Uznanie takiej roli programu jest równoznaczne z akceptacją określonych kryteriów i funkcji, które program ten musi spełniać, w związku z czym ewaluator musi odpowiedzieć na pewne pytania. Jest to podejście funkcjonalne. Każdy program musi spełniać wymagania odpowiadające jego funkcji.

Podejście zoperacjonalizowanych zadań

Podejście zoperacjonalizowanych zadań rozwiązuje dylemat współczesnego ewaluatora w nieco odmienny sposób. Punktem wyjścia są tutaj cele programu. Następnie zbiera sie materiały, które mają pomóc stwierdzić, czy cele te zostały osiągnięte. Cele są wyłącznym wyznacznikiem standardów i kryteriów. Ewaluator identyfikuje zamierzenia twórców programu. Uznaje się, że program odniósł sukces, jeśli stopień zgodności między założonymi celami oraz rezultatami jest wystarczająco wysoki. Odwoływanie się do przyjętych w programie celów zwalnia ewaluatora od dociekania jego funkcji.

Na użytek edukacji model mający za punkt wyjścia cel opracował Tyler (1950), zwolennik definiowania celów edukacyjnych w katego-

riach zachowań uczniów. Jego zdaniem ewaluacja programu powinna określać zarówno jego rezultaty, jak i cele w kategoriach specyficznych zachowań jednostek. Zadaniem ewaluatora jest stwierdzenie, czy uczestniczący w programie uczniowie przejawiali w konsekwencji określone partykularne zachowania. Podejście to określane jest zwykle jako „podejście zoperacjonalizowanych zadań" lub jako model programu i ewaluacji Tylera.

Omawiane podejście rozwinął Mager (1962), który nie tylko kładł nacisk na określenie celów w kategoriach behavioralnych, ale podkreślał, że cele ogólne same w sobie zawierają kryteria i zadania operacyjne, które należy urzeczywistnić według twórców programu. Skwantyfikowane zmienne rezultatów, a zwłaszcza testów osiągnięć edukacyjnych, są miernikiem zachowań uczniów. Między innymi Bloom (1956) skonstruował taksonomię odpowiednich celów edukacyjnych. Bloom też wraz ze współpracownikami (1971) oraz Popham (1975) wykorzystali to podejście w praktyce ewaluacyjnej.

W ostatnich latach zwolennicy omawianego podejścia przenieśli punkt ciężkości swoich zainteresowań z problemu adekwatnego identyfikowania celów na problem pomiaru ich realizacji. Tradycyjne standaryzowane testy osiągnięć nie są dostosowane do mierzenia stopnia realizacji wcześniej określonych celów. Stąd pojawiło się duże zainteresowanie konstruowaniem testów mających „odniesienie do kryteriów" (*criterion reffered*). Te zaś są nierozerwalnie związane z celami. Tak więc dla twórców programu punktem wyjścia jest układ celów, z których następnie wyprowadza się zadania testowe.

Taki charakter miała prowadzona przez Tylera Narodowa Ocena Postępu Edukacyjnego (The National Assessment of Educational Progress). W ostatnim okresie w podobny sposób testowano kompetencje. Cele będące podstawą testów często przyjmuje się za minimum na poziomie klasy. Mówiąc inaczej, każde dziecko musi je opanować. Uznaje się, iż uzyskało ono „minimalne" kompetencje wówczas, gdy osiągnęło określoną liczbę punktów. Niemal każdy stan posiada taki — zorientowany na kompetencje — program testowania. W niektórych stanach (na przykład na Florydzie) promowanie uczniów do następnej klasy jest uwarunkowane wynikami testów. Podejście oparte na celach behavioralnych jest najbardziej popularne na poziomie stanu.

Przyjęło się ono również w innych, nie związanych z edukacją dziedzinach. Na przykład zarządzanie w biznesie i organizacjach rządowych bazuje na celach. Zarówno organizacje, jak i jednostki definiują swoje cele, a następnie oceniane są na podstawie stopnia ich realizacji. Nic

w tym dziwnego, biorąc pod uwagę fakt, iż omawiany — zorientowany na cele — trend ma swoje źródło w „analizie zadań" i psychologii przemysłowej popularnej na początku XX w.

Jedna z wcześniejszych książek Suchamana (1967), dotycząca ewaluacji zdrowia publicznego, oparta była na podejściu zorientowanym na cel. Autor przedstawia w niej proces ewaluacji jako: ustalenie celu działania; zoperacjonalizowanie celu działania; oszacowanie [oczekiwanego] wyniku operacji celowych, stworzenia wartości, osiągnięcia celu lub celów; dokonanie pomiaru (s. 34). „Najbardziej charakterystyczną cechą badań ewaluacyjnych — pisze Suchman — jest obecność pewnego celu, istotą zaś problemu badawczego jest dążenie do określenia stopnia jego realizacji" (s. 37). Jasno określony cel programu jest odpowiednikiem hipotezy w badaniach naukowych.

Widać, że podejście Schumana jest z gruntu różne od opisanej wyżej analizy systemowej. Analizą systemową posłużono się w ewaluacji *Follow Through*, której poddano programy edukacyjne dla trzynastoletnich dzieci. Ewaluacja składała się jedynie z czterech testów. Uznano przypuszczalnie, że to, co mierzyły te cztery testy, miało znaczenie decydujące. W rzeczywistości jednak każdy z tych programów miał dziesiątki jasno sprecyzowanych celów. Przy podejściu opartym na ewaluacji celów behavioralnych dążono by do stwierdzenia, czy każdy z tych celów został osiągnięty. W najgorszym wypadku próbowano by dokonać redukcji celów do jakiegoś mierzalnego ich układu.

Podejście zorientowane na podjęcie decyzji

We wszystkich współczesnych podejściach ewaluacyjnych podkreśla się związek między ewaluacją a podejmowaniem decyzji, chociaż różne bywają odpowiedzi na pytanie, kim są decydenci i w jaki sposób decyzje powinny być podejmowane. Podejście zorientowane na podejmowanie decyzji głosi, że struktura ewaluacji powinna być podporządkowana konieczności podjęcia decyzji. Dotyczy to zazwyczaj decyzji osób wysoko postawionych, menedżerów lub administratorów.

W dziedzinie edukacji głównym zwolennikiem tego podejścia był Stufflebeam (1973). „Ewaluacja to proces określania charakteru, uzyskiwania i dostarczania takich informacji, które pozwoliłyby na ocenę alternatywnych decyzji" (s. 129). Stufflebeam wyróżnił trzy zadania decyzji (utrzymanie homeostazy, uzyskanie „przyrostu" i pobudzenie nowej ruchliwości), cztery typy decyzji (planowanie, strukturyzowanie, wdrażanie i ponowne wprowadzanie w obieg [*recycling*]), trzy kroki ewaluacji

(„wyznaczanie zakresu", „uzyskiwanie" i „dostarczanie") i cztery typy ewaluacji (orientacja na kontekst, na wkład [*input*], na proces i na wynik).

Niezależnie od typu ewaluacji, w jej konstruowaniu najistotniejsze wydaje się: identyfkowanie poziomu decyzji, której ma służyć; projektowanie sytuacji decyzyjnych; definiowanie kryteriów dla każdej z tych sytuacji; określanie kierunków działania ewaluatora. Następnym krokiem jest gromadzenie informacji, ich porządkowanie, analizowanie i przedstawienie w formie raportu. Każda faza dzieli się na szereg zadań, które ewaluator musi wykonać (Stufflebeam, 1969).

Guttentag (1973; 1974) pokusił się o ilościową analizę struktury preferencji decydentów. Chodziło mu o ustalenie działań lub programów, które pozwoliłyby na maksymalizację różnych uznawanych przez decydentów wartości. Wykorzystał w tym celu metodę wielowymiarowego skalowania przydatności opracowaną przez Edwardsa. „Przydatność" dotyczy rezultatów działania w świetle wartości preferowanych przez decydentów. Celem jest ustalenie takiego przebiegu działania, który maksymalizowałby przydatność.

Omawiane podejście zakłada konieczność zdefiniowania wymiarów odnośnych wartości i uzyskania odpowiedzi na pytanie, jak na tej skali wartości decydenci lokują obiekty ewaluacji. Cała procedura oparta jest na subiektywnych ocenach decydenta. Użyteczność działań alternatywnych zsumowana według wymiarów różnych wartości pozwala określić, które działanie jest najbardziej użyteczne.

Patton (1978), kolejny przedstawiciel podejścia zorientowanego na podejmowanie decyzji, podkreśla znaczenie czynnika osobowego przy wykorzystywaniu wyników ewaluacji „Pierwszym krokiem w podejściu zorientowanym na wykorzystanie wyników ewaluacji jest identyfikacja i organizacja odpowiednich decydentów oraz użytkowników informacji" (s. 61). Patton utrzymywał, że wyniki ewaluacji są wykorzystywane wówczas, kiedy pewne jednostki przyjmują osobistą bezpośrednią odpowiedzialność za dostarczenie informacji właściwym ludziom. Odpowiedzialność za określenie, jaki rodzaj informacji jest potrzebny, spoczywa na decydentach.

Drugi krok w ewaluacji polega na ustaleniu odpowiednich, wartych uwagi, pytań. Właściwe pytanie charakteryzuje się określonymi cechami. Można wyjaśnić dane mające związek z pytaniem, a odpowiedzi na nie może być kilka. Konkretni decydenci mogą potrzebować informacji, które pomogą im odpowiedzieć na pytania. Decydenci pragną odpowiedzieć na pytania sami, przywiązują wagę do odpowiedzi. Wreszcie,

decydenci potrafią stwierdzić, jak chcieliby spożytkować odpowiedzi na pytania.

Mówiąc krótko, podejście zorientowane na podjęcie decyzji próbuje rozwiązać problemy, przed którymi stoi ewaluator, uznając decydenta za adresata ewaluacji, jego zaś zainteresowania i kryteria za istotne. Sposoby identyfikowania decydentów i formy przejawiania się ich decyzji bywają różne. Zakłada się jednak, że takie podejście zwiększy możliwość wykorzystania rezultatów ewaluacji.

Typową dla tego podejścia metodą badawczą są kwestionariusze i wywiady. Ewaluator nie ucieka się do eksperymentów, lecz śledzi naturalne zmiany w toku realizacji programu. Tym, co wymaga odpowiedzi, są pytania decydentów. Dotyczą one zwykle efektywności niektórych aspektów programów, a w szczególności stwierdzenia, które jego elementy zapewnią sukces w praktyce.

Podejście Goal-Free*

Podejście to było bezpośrednią reakcją na upowszechnienie się ewaluacji zorientowanej na cel. Scriven (1973) utrzymywał, iż ewaluator nie tylko nie powinien przyjmować za podstawę celów danego programu, ale wręcz nie powinien ich znać. Pozwoli mu to zachować bezstronność. Ewaluację *goal-free* należy rozpatrywać w szerszym kontekście zainteresowania Scrivena redukowaniem wpływu, jaki na ewaluację może wywierać brak bezstronności. Cel nie pozostawia ewaluatora obojętnym. Scriven pisze: „Niestety, nie mogę ewaluować ich [projektodawców] osiągnięć wystarczająco obiektywnie, jeżeli przez cały czas jestem przez nich indoktrynowany" (s. 320).

Model *goal-free* redukuje wpływ intencji twórcy programu na nastawienie ewaluatora dzięki temu, że ewaluator ich nie zna. Skutek jest taki, że ewaluator musi dociekać wszystkich możliwych rezultatów projektu. Wiele z nich to przypadkowe skutki uboczne, które mogą mieć charakter pozytywny lub negatywny. Scriven pisał o tym w sposób następujący:

„Tak zwany «skutek uboczny», tak korzystny, jak i niekorzystny, często przesądza o wynikach ewaluacji. Ewaluator, rzecz jasna, nie dzieli skutków na «uboczne» i «zasadnicze». Taki podział oddaje intencje twórców projektu, a ewaluator ocenia przecież nie intencje, lecz osiągnięcia.

*Wobec trudności przetłumaczenia angielskiego określenia modelu w sposób poprawny, a zarazem wystarczająco zwięzły, pozostawiamy je w brzmieniu oryginalnym — *przyp. red.*

W praktyce ryzykowne jest nawet ogólne zaznajamianie się z intencjami, odwraca bowiem uwagę od «skutków ubocznych», prowadzi do ich ignorowania bądź deprecjonowania" (s. 321).

Ze wszystkich przedstawianych podejść — ewaluacja *goal-free* znalazła najmniejsze zastosowanie w praktyce (w konsekwencji kwestionowany jest niekiedy jej status modelu). Zarówno ewaluatorzy, jak i twórcy programów społecznych nie potrafią sobie najczęściej wyobrazić sytuacji, w której kryteria ewaluacji nie wynikałyby z celów przyświecających tymże twórcom. Uważa się, że alternatywą są wtedy własne cele ewaluatorów.

Z drugiej jednak strony organizacja konsumentów regularnie ocenia towary konsumpcyjne bez uwzględniania celów producentów. Przy ewaluacji samochodów cele General Motors nie mają znaczenia. Punktem wyjścia dla organizacji konsumentów są standardy i kryteria, które — jak się sądzi — będą korzystne dla konsumentów, a nie dla producentów. Podobnie zwolennicy ewaluacji *goal-free* uznają za swoich adresatów (lub grupę odniesienia) konsumentów (a nie menedżerów) programów społecznych. Działania stowarzyszeń konsumenckich mogą stanowić dobre potwierdzenie sposobu myślenia preferowanego przez Scrivena.

Oczywiście, programy społeczne to nie to samo co samochody. Określenie zestawu standardów ewaluacji nie jest w ich przypadku takie proste, jak w odniesieniu do towarów konsumpcyjnych. W związku z tym Scriven przyjął za podstawę waluacji pojęcie „potrzeb". Potrzeby, w przeciwieństwie do wymagań i pragnień, ujawniane są w wyniku szacowania niedostatku. Ustalenie obiektywnej zasadności danych potrzeb stałoby się dla ewaluatora obowiązującym źródłem standardów. W konsekwencji więc ewaluacja *goal-free* sprowadza się do analizy potrzeb konsumentów, zamiast do celów producentów.

Metodologia ewaluacji *goal-free* nie jest tak sprecyzowana, jak w przypadku niektórych innych podejść. Najogólniej, Scriven jest przekonany, iż tylko niezaangażowane „spojrzenie z boku" może zapewnić obiektywność. Ewaluator postrzegany jest jako ekspert, który stara się eliminować i zapobiegać stronniczości zarówno u siebie samego, jak u innych. Jedną z proponowanych technik są eksperymenty, w których ani obiekt ewaluacji, ani ewaluator nie orientują się, co jest środkiem leczniczym, a co *placebo*. Inne techniki polegają na mieszaniu *ante* i *post factum* elementów ocenianych, skalowaniu sądów przy użyciu standardowych przykładów, konfrontowaniu ze sobą poszczególnych organizatorów programu, wy-

korzystywaniu kodeksów etycznych, replikacji zbieżnych, powtarzaniu analiz przez niezależne grupy itd. (Scriven, 1976a).

Scriven kładł duży nacisk na niezależność ewaluatora wobec organizatorów programu. Jego zasada „niezależnego sprzężenia zwrotnego" głosi, że nie należy zdawać się wyłącznie na tę część całego mechanizmu, której relacje zwrotne mają być obiektem ewaluacji. Z kolei zasada „niestabilności niezależności" wyraża przekonanie, iż w trakcie ewaluacji mają miejsce nieustanne próby „kooptowania" ewaluatorów, toteż muszą oni walczyć o swoją niezależność. Scriven przyrównuje ewaluatora do myśliwego lub detektywa, który odkrywa tropy i śledzi zdobycz. Niekiedy do koronera, który szuka powiązań przyczynowo-skutkowych, nie mogąc posłużyć się eksperymentem. Wykorzystuje wtedy metodę *modus operandi*. W taki właśnie sposób powinien działać ewaluator.

Techniki wykorzystywane w niezbyt licznych ewaluacjach *goal-free* były bardzo zróżnicowane. Na przykład ewaluując pewne innowacje w programie szkolnym, House i Hogben (1974) przeanalizowali wyniki testów wstępnych, zapoznali się z recenzjami ekspertów, przeprowadzili wizytacje w klasach szkolnych oraz wywiady z nauczycielami i uczniami, wreszcie — przeanalizowali nowe propozycje same w sobie. W innej ewaluacji programu szkolnego Welch (1978) wykorzystał grupę sędziów do opracowania punktacji zestawu narzędzi kontrolnych. Z kolei Harrington i Sanders (1979) podjęli próbę sformułowania pewnych ogólnych zasad ewaluacji *goal-free*.

Podejście typu „krytyka artystyczna"

Role krytyka sztuki, krytyka literackiego, krytyka teatralnego i krytyka filmowego są dobrze znane. Działania ich zawierają powszechnie akceptowane sposoby oceny dzieł sztuki. Eisner (1979) i inni badacze analizowali analogiczny model krytyka edukacji lub krytyka programu szkolnego. Wydaje sie, że takie samo podejście można by stosować również w innych dziedzinach działań społecznych.

Eisner wychodzi od tego, że krytyka, podobnie jak twórczość artysty, ma z istoty charakter jakościowy. „Krytyk sztuki staje przed trudnym zadaniem przełożenia jakości konstytuujących dzieło sztuki, z gruntu nie dających się wyrazić słowami, na język, który umożliwiłby innym ludziom pełniejszy odbiór dzieła" (s. 191). Eisner twierdzi, że tego typu krytyka ma charakter empiryczny, jako że jakości, do których odwołuje się krytyk, można odnaleźć w samym dziele. Ponadto zaś wszystko może

stanowić przedmiot krytyki. Krytyka nie jest równoznaczna z dążeniem do negatywnej oceny jej przedmiotu, ma natomiast rzucić światło na jego „jakości". Dzięki temu można ocenić jego wartość. Z kolei miarą instrumentalnego wpływu krytyki są preferencje odbiorców.

Eisner dokonuje rozróżnienia między ekspertyzą a krytyką. Ekspert ocenia wartość, krytyk natomiast ujawnia tajemnice (s. 193). Ekspert potrafi rozpoznać i ocenić jakość czegoś konkretnego. Nie wymaga to ani publicznych ocen, ani opisów. Stanowi natomiast konieczny wstęp do przeprowadzania krytyki.

Ekspert musi mieć ogromne doświadczenia, które pozwala mu na rozpoznanie jakości niezależnie od tego, czy chodzi o wino, przedstawienie czy klasę szkolną. Tym niemniej istnieje różnica między rozpoznawaniem a dostrzeganiem. Zwykłe rozpoznanie i klasyfikacje oparte na doświadczeniu to za mało. Trzeba również mieć umiejętność spostrzegania subtelności i zrozumienia ich roli w ukształtowaniu całości (na przykład życia klasy szkolnej).

Aby stać się ekspertem, trzeba mieć możliwość uczestniczenia w wydarzeniach i porównywania ich. Percepcja musi być wnikliwa, całościowa i szacująca. Konieczna jest także wrażliwość na wyłaniające się nowe tendencje w klasie szkolnej lub w jakimkolwiek ewaluowanym programie. To z kolei wymaga zestawu idei-teorii, modeli lub czegokolwiek, co pozwala odróżniać to, co znaczące, od tego, co trywialne. Kluczowe jest — w tym kontekście — uzyskanie właściwego wykształcenia oraz doświadczenia.

Z kolei „krytyka jest sztuką wydobywania na jaw tych swoistych jakości wydarzeń lub obiektów, które dostrzegają znawcy" (s. 197). Krytycy próbują „eksponować" sytuację w taki sposób, aby rzucić światło na znaczące aspekty sytuacji, obiektu lub programu. Eisner rozróżnia przy tym dyskursywne i niedyskursywne sposoby prezentacji. Krytyk próbuje wydobyć istotę tego, co partykularne. Język krytyki jest niedyskursywny, ponieważ wykorzystuje takie formy, które w większym stopniu prezentują uczucia niż reprezentują fakty.

We właściwy sposób dokonana krytyka powoduje wzrost świadomości i otwiera oczy na wartości. Zadaniem krytyka jest posłużenie się takimi kryteriami, aby sąd o danym wydarzeniu mógł być podbudowany oceną jego znaczenia. Takie analizy jakościowe pozwalają odbiorcy zastępczo uczestniczyć w opisywanych zjawiskach.

Podejmując problem trafności i rzetelności, Eisner używa pojęć „potwierdzenie strukturalne" i „adekwatność odniesienia". Dowód zyskuje potwierdzenie, gdy jego części składowe sobie nie przeczą. Z kolei ade-

kwalność odniesienia dotyczy możliwości odnalezienia w poddanym krytyce działaniu lub programie cech, które krytycy uznają za ważne. Podkreśla się przy tym konieczność pełnego rozeznania sytuacji. Jednak od krytyków nie oczekuje się jednomyślności. Według Eisnera, głównym celem jest poszerzenie percepcji, a nie definitywny osąd. Krytyka kształci znawstwo u innych ludzi. Jak dotąd, podejście typu „krytyka artystyczna" propagował w ewaluacji głównie Eisner i jego uczniowie np. McCutcheon (1978) i Vallance (1978). Ograniczało się ono do edukacji.

Trudno nie zauważyć, że istnieje wiele różnych podejść do samej krytyki. Kelly (1978) utrzymywał, że możliwa analogia występuje nie tyle w przypadku ewaluacji i krytyki artystycznej, co ewaluacji i krytyki literackiej. Autor ten przeniósł do ewaluacji z krytyki literackiej takie pojęcia, jak metafora, wypowiedź, fabuła i temat. W swojej analizie pokazuje, jak ewaluatorzy wykorzystują podobne pojęcia zarówno badając programy, jak sporządzając raporty. W ujęciu Kelly'ego opis przemieszany jest z ewaluacją. Ponadto problem wiarygodności krytyka ma dla niego większe znaczenie niż dla Eisnera. Kelly twierdzi, że wybieramy zwykle tego krytyka, z którym się zgadzamy. W konsekwencji charakter krytyki może być różny w zależności od materii jej przedmiotu.

Krytyka może być zróżnicowana również w obrębie tej samej dziedziny. Jenkins i O'Toole (1978) wykazali to na przykładzie krytyków sztuki. W placówkach uniwersyteckich spotkać można krytyków akademickich, którzy wszystko oceniają w oparciu o najwyższe standardy. Istnieją pisarze-krytycy, którzy uważają krytykę za „produkt uboczny" swojej działalności i próbują określić takie „standardy smaku", za pomocą których również i oni sami mogliby być ewaluowani. W tym przypadku krytyka nabywa pewnych cech twórczego pisarstwa. Wreszcie istnieją wolni strzelcy lub krytycy-dziennikarze, którzy odwołują się do szerokiego, „rynkowego" odbiorcy.

Najogólniej mówiąc, krytyk adresuje swoją wypowiedź albo do grupy konsumentów, albo do ekspertów. Odbiorca przekona się, czy program jest dobry; ponadto jego wrażliwość wzrośnie. Przyjmuje się przy tym, że odbiorca zaakceptuje zarówno krytyka, jak wykorzystywane przez niego standardy. Nie istnieje tutaj żadna standardowa metodologia, za wyjątkiem kilku sposobów krytycznego przygotowywania recenzji. Uważa się przy tym za niezbędne, aby krytyk „tkwił w temacie", aby był z nim doskonale obeznany. Punktem wyjścia jest należyte doświadczenie i wykształcenie. W swojej ostatecznej konsekwencji krytyka podnosi poziom świadomości, doskonali standardy i poprawia realizację założeń. Podejście typu „krytyka" rozwiązuje podstawowy

dylemat ewaluacyjny desygnując na ewaluatora osobę doświadczoną i właściwie wykształconą; mówiąc krótko — znawcę przedmiotu. Wykorzystując swoją wrażliwość, taka doświadczona osoba przekazuje swoje oceny w formie, która dostarcza informacji oraz kształci mniej wrażliwych.

Podejście typu „inspekcja kwalifikacji"

Przez przeszło pięćdziesiąt lat stowarzyszenia zawodowe organizowały ewaluację kształcenia prowadzonego w poszczególnych dziedzinach. Zarówno w przypadku lekarzy, prawników, pracowników społecznych, jak i nauczycieli, podstawowym sposobem ewaluacji była dokonywana przez profesjonalistów inspekcja szkół. Szerzej rozumiana inspekcja kwalifikacji polega na tym, że chirurdzy albo profesorowie, albo prawnicy oceniają kolegów pracujących w tej samej dziedzinie. Zazwyczaj członkowie grupy inspekcyjnej reprezentują ten sam poziom przygotowania zawodowego co ewaluowani.

Wiele z takich ewaluacji miało na celu potwierdzenie, że poziom nauczania w szkole zasługuje na uznanie. Na przykład od 1885 r. grupy szkół średnich zaczęły się organizować w celu standaryzacji wymagań dających wstęp do koledżów. Z działań tych w 1978 r. powstało stowarzyszenie Państwowe Badania Ewaluacji Szkoły (National Study of School Evaluation), które wydaje *Evaluative Criteria*. Publikuje ono kryteria ewaluacji szkół średnich.

Kryteria dotyczą poszczególnych przedmiotów — rolnictwa, sztuki, edukacji dla biznesu itd. Niektóre kryteria odnoszą się do takich dziedzin, jak kształcenie medialne, działania usługowe na rzecz uczniów, wyposażenie szkoły. Program edukacyjny jest podzielony na osiemnaście przedmiotów. Dodatkowo istnieje jeszcze dziewięć sfer dotyczących usług. W ramach każdego przedmiotu istnieje kilka list kontrolnych dotyczących tych aspektów, które mają być poddane ewaluacji.

Na przykład w odniesieniu do języka angielskiego lista taka zawiera ustalenia w rodzaju: „Lekcje angielskiego są obowiązkowe dla wszystkich uczniów" (s. 103); „Niektóre dzieła współczesne i klasyczne przerabia się szczegółowo, przy czym ich wybór jest dostosowany do zdolności, zainteresowań i dojrzałości uczniów" (s. 195); „Zwraca sie uwagę na poprawianie umiejętności uczenia się" (s. 111). Ewaluator ocenia realizację tych ustaleń według pięciopunktowej skali — od „niedostateczna" do „doskonała".

Celem prezentowanego podejścia jest zainspirowanie samoewaluacji pracowników szkoły. Przed dokonaniem inspekcji przez jednostkę zewnętrzną pracownicy szkoły zostają podzieleni na kilka grup. Grupy te dokonują ocen według listy kontrolnej. O wynikach informowani są wszyscy pracownicy szkoły. Agenda zewnętrzna (np. North Central Association) powołuje własny komitet, którego celem jest ocena trafności samoewaluacji. Szkoły mają prawo weta w kwestii składu członków tego komitetu. Liczba członków komitetu waha się od 10 do 30.

Komitet ten dzieli się na mniejsze grupy przedmiotowe, z których każda weryfikuje rezultaty samoewaluacji, zmieniając je w niektórych — wymagających tego — przypadkach. Zarówno każda grupa, jak i cały dokonujący inspekcji komitet przygotowują pisemne raporty. Przed opuszczeniem szkoły komitet przedstawia jej pracownikom krótki ustny raport, w którym wymienia mocne i słabe strony szkoły. Ponadto komitet przedstawia własne rekomendacje. Raport orzeka o poziomie profesjonalnym szkoły. Oczekuje się przy tym, że w rezultacie inspekcji podjęte zostaną w szkole kroki na rzecz wyeliminowania dostrzeżonych słabości. Wprawdzie kryteria i procedury różnią się do pewnego stopnia, ale podobne inspekcje odbywają się niemal w każdej szkole — od najbardziej elementarnych do *stricte* profesjonalnych. Organizują je różne agendy, zarówno stowarzyszenia profesjonalne, jak rządy stanowe.

Jednym z bardziej interesujących wariantów ewaluacji kwalifikacji jest gwałtownie upowszechniająca się ewaluacja katedr czy innych jednostek uniwersyteckich. Sponsorowana jest ona zwykle przez uniwersytecką administrację. Jedna z najwcześniejszych takich ewaluacji związana była z działaniami Council on Program Evaluation (COPE) na Uniwersytecie Stanu Illinois. W latach 1972–1979 ewaluowane były tam wszystkie jednostki akademickie.

W tym celu powołano radę składającą się z dziewięciu profesorów. Kierował nią przedstawiciel centralnej administracji uniwersyteckiej. Później dookoptowano do niej po dwóch studentów studiów magisterskich i studiów na poziomie bakalaratu. Rada ewaluowała na podstawie sześciu kryteriów:

1) jakości programu nauczania;

2) jakości badań, aktywności twórczej lub pracy naukowej;

3) jakości usług świadczonych na rzecz uniwersytetu i określonej profesji;

4) znaczenia danej jednostki dla innych jednostek uniwersyteckich i jej wkładu w ich rozwój;

5) wartości programu dla społeczeństwa lub jego wyróżniania się na poziomie stanu;

6) dalszych oczekiwań pod adresem danej jednostki.

Początkowo rada powoływała grupy zadaniowe składające się z 5–10 członków. Celem ich działania była analiza sytuacji w poszczególnych katedrach. Grupy zadaniowe przeprowadzały wywiady z pracownikami katedr. Na zakończenie formułowały na piśmie poufną ocenę katedry, którą przedstawiano radzie. Ta z kolei analizowała raport i na jego podstawie przedstawiała swoją własną skróconą ocenę — zwykle w formie listu, który przesyłano do wicekanclerza oraz do katedry. Katedra mogła ustosunkować się do listu i zawartych w nim rekomendacji.

Metoda ta, stosowana przez kilka lat, była tak bardzo czasochłonna, iż w konsekwencji zrezygnowano z tworzenia odrębnej grupy zadaniowej dla każdej katedry. Zdecydowano się na skonstruowanie zestawu formularzy dotyczących większości aspektów działania każdej katedry. Składały się one z kwestionariuszy wypełnianych przez katedry. Rada recenzowała otrzymane druki i dokonywała ogólnej oceny oraz przedstawiała — oparte na kryteriach formalnych — rekomendacje. Największy wpływ na ostateczną ocenę miała jakość badań naukowych. Jeśli natomiast ocena katedry wypadała szczególnie słabo lub wymagała dokładniejszych analiz, wówczas powoływano specjalną grupę, której zadaniem było przeprowadzenie dokładniejszej ewaluacji w kontaktach bezpośrednich (Petrie, 1979).

Oczywiście nie wszystkie inspekcje kwalifikacji działają w ten sam sposób. Na przykład komisjami ewaluacyjnymi w poszczególnych dyscyplinach medycznych kierują zespoły ludzi nie wyróżniających się kwalifikacjami zawodowymi, ale egzamin — zarówno pisemny, jak ustny — lekarz zdaje przed specjalistą. We wszystkich takich ewaluacjach standardy i kryteria ustanawiają sami profesjonaliści. Zakłada się, że najlepiej potrafią ocenić profesjonalizm w dziedzinie, którą reprezentują. Procedury ewaluacyjne są różne, jednak istotą jest holistyczne szacowanie programu w danej dziedzinie przez specjalistów.

Podejście quasi-sądowe (konfrontacyjne)

Quasi-sądowe procedury były od dawna używane zarówno w ewaluacji, jak i przy ustalaniu strategii decyzyjnych. W ostatnim okresie wiele ewaluacji prowadzono na wzór rozpraw sądowych lub naśladowano inne procedury konfrontacyjne, zwłaszcza agresywne postępowanie sądowe, przy rozstrzyganiu spornych kwestii w programach społecznych.

Tym niemniej nie wszystkie *quasi*-sądowe procedury mają charakter konfrontacyjny. Jeśli powoływanie grupy jurorów i różnego rodzaju komisji uzna się za formę procedur sądowych, wówczas należy uznać, że większość z nich nie ma takiego charakteru. Przepisy prawa cywilnego stanowią, że rewizje sądowe w sprawach administracyjnych muszą być przeprowadzane zgodnie z zasadami właściwymi dla procesu sądowego, ale nie mają one charakteru konfrontacyjnego.

Na przykład Komisja Warrena prowadząca dochodzenie w sprawie zabójstwa prezydenta Kennedy'ego była grupą powołaną przez prezydenta, która przesłuchiwała świadków, prowadziła własne dochodzenie i doszła do wniosków na temat prawdopodobnego przebiegu wydarzeń. Podobnie Komisja Kernera do Sprawy Niepokojów Społecznych (Kerner Commission on Civil Disorders) przedstawiła wnioski na temat prawdopodobnych przyczyn analizowanych wydarzeń. Często rodzaj takiego „sądu" powoływany jest dla rozwiania wątpliwości w przypadku zagadnień kontrowersyjnych. Jednym z ostatnich przykładów są działania komisji Wirtza (1977), która zajęła się problemem obniżenia (w skali całego kraju) wyników standardyzowanych testów.

W Anglii wykorzystywano często tego typu komisje i rady w celu ustalania kierunku polityki społecznej (Kogan, Packwood, 1974). Dobrymi przykładami mogą być Plowden Committee w 1967 r. do spraw edukacji na poziomie podstawowym, ostatnio Raport Taylora dotyczący lokalnego zarządzania szkołami. Takimi zespołami oceniającymi kierowali powszechnie szanowani obywatele. Uzyskiwali przy tym pomoc ze strony rządu.

Tego typu zespoły działały na wszystkich niemal płaszczyznach życia społecznego. Niekiedy powoływano je po to, aby rozwiązać jakiś problem w ramach określonej profesji. Na przykład finansowany przez National Education Association zespół badający system sprawozdawczości edukacyjnej w stanie Michigan (House, Rivers, Stufflebeam, 1974). To samo stowarzyszenie sponsorowało podobne badania na Florydzie (Tyler i współaut., 1978). W ramach działalności Biura d/s Edukacji (Office of Education) Joint Review Dissemination dokonywał analiz, których celem było ustalenie, czy uzasadnione jest upowszechnienie na terenie całego kraju programów tworzonych przez instytucje federalne.

Mimo iż takie zespoły i komisje oceniające nie są częścią systemu prawnego, to jednak w praktyce często wywodzą swój autorytet z procedur *quasi*-sądowych. Sprawa staje przed trybunałem. „Świadkowie" składają zeznania i przedstawiają „dowody". Istnieją jasne zasady określające zarówno, kto może dostarczać dowodów, jak i warunki, w jakich

powinny być składane zeznania. Przesłuchaniom nadaje się sądową symbolikę i powagę, a wystrój wnętrz imituje często salę sądową. Wielkie znaczenie przypisuje się też wysłuchaniu wszystkich „świadectw" za i przeciw. Mówiąc krótko, podstawowe problemy ewaluatora rozwiązuje się sięgając po *quasi*-sądowe procedury. Ewaluacja zapożycza swój autorytet od prawa.

W ostatnim okresie wykorzystywano w celach ewaluacyjnych procedurę konfrontacji sądowej. W punkcie wyjścia takich przesłuchań zakłada się, iż fakty dotyczące analizowanego przypadku mogą być stwierdzone wówczas, gdy każda ze stron przedstawi sądowi własne, przemawiające na jej korzyść dowody (Owens, 1973). Stronniczość oponentów pomaga w ujawnieniu niezbędnych dowodów.

Pierwsze tego typu konfrontacyjne przesłuchania przeprowadzono w 1970 r. na Hawajach. Chodziło o podjęcie decyzji, czy w szkołach publicznych na Hawajach należy realizować program *Man: A course of Study*. Rozprawie przewodniczyło trzech sędziów (wszyscy byli pedagogami). Uczestniczące w niej strony przygotowywały się przez wiele tygodni. Inna pozorowana rozprawa sądowa odbyła się na Uniwersytecie Stanu Indiana i dotyczyła ewaluacji programów kształcenia nauczycieli (Wolf, 1975). W toku rozprawy przedstawiono argumenty za i przeciw programowi. W ciągu dwóch dni wysłuchano zeznań trzydziestu dwóch osób i przesłuchano je w „krzyżowym ogniu pytań". Sąd składał się z trzynastu ekspertów w dziedzinie edukacji.

Organizator procesu pisze: „Za stosowaniem metodologii parasądowej przemawia być może głównie to, że oferuje ona system jasnych zasad i procedur pozwalających przed ogłoszeniem wyroku rozpatrzeć alternatywne konkluzje wynikające z przedstawionych danych. [...] Podejście prawnicze umożliwia systematyczne rozważenie alternatywnych argumentów i wniosków, dzięki czemu ewaluacja jest intelektualnie uczciwa i rzetelna" (Wolf, 1975, s. 185).

Proces według koncepcji Wolfa powinien składać się z czterech etapów: wytypowanie kwestii spornych; selekcja kwestii spornych; przedłożenie argumentów; przesłuchanie. Na etapie pierwszym przeprowadzano serię wywiadów, których celem było ustalenie trzydziestu kwestii spornych. Etap drugi polegał na zlustrowaniu tych kwestii w bezpośrednim zetknięciu z wykładowcami, uczniami i administratorami. Celem było zredukowanie kwestii spornych. Na etapie trzecim przygotowywano argumenty. Etap czwarty stanowiły sesje poprzedzające przesłuchania (podczas nich analizowano dowody) oraz przesłuchania właściwe. Cały proces trwał sześć miesięcy.

Polemizując z krytykami propagowanej przez siebie metody, Wolf starał się zminimalizować znaczenie jej konfrontacyjnego charakteru. „Najbardziej znaczącym aspektem tej całej metafory [...] jest być może wychowawczy charakter sposobu, w jaki sąd zbiera informacje" (Wolf, 1979, s. 21). Zeznania to „kamień węgielny dowodów". „Podstawę rozważań oraz wyroku stanowią nie tylko «fakty», ale także analiza różnych punktów widzenia, opinii, uprzedzeń i spekulacji, i to w kontekście wartości i przekonań" (tamże).

Stenzel (1979) przeanalizował w praktyce, czy modelem dla ewaluacji może być procedura przesłuchań prowadzonych w Kongresie. Biuro d/s Edukacji w Illinois utworzyło komitet, którego celem była analiza mocnych i słabych stron takiej ewaluacji w przypadku jednego z programów pedagogiki specjalnej. Do obowiązków komitetu należało określenie stopnia trafności ewaluacji, przedstawienie opinii dotyczącej rekomendacji, które zaproponowano w raporcie ewaluacyjnym, oraz pisemne ustosunkowanie się do treści tego raportu. Komitet składał się z trzech pracowników Biura d/s Edukacji Stanu Illinois. Na spotkaniu ewaluatorzy odpowiadali na pytania stawiane przez członków komitetu. Jak się okazało — procedury kongresowe, przedstawione przez Stenzela (1979) istotnie różniły się od procedur sądowych.

Istnieje wiele *quasi*-sądowych procedur, które mogłyby znaleźć zastosowanie w ewaluacji. Najlepiej znane z nich to grupa jurorów oraz przewód sądowy przeprowadzony przez skład sędziowski. Celem procedur parasądowych jest dojście — w ten czy inny sposób — do konkluzji. Większość takich procedur opiera się na rodzaju przesłuchań, w toku których przedstawia się argumenty „za" i „przeciw". Prawomocność tego podejścia zależy od stopnia akceptacji wykorzystanych w nim procedur szczegółowych, osób dokonujących przesłuchań oraz składu sędziowskiego podejmującego decyzję. Trzeba przy tym dodać, że w prawie zwyczajowym o postępowaniu sądu decyduje zazwyczaj sam zespół sędziowski.

Studium przypadku (albo transakcja)

W podejściu tym uwaga jest skoncentrowana na procesie realizacji programu oraz na tym, jak odbierają go ludzie. Podstawowe pytanie brzmi następująco: „Jak odbierają program różni znający go ludzie?". Metoda polega zwykle na wywiadach z wieloma ludźmi, obserwacjach uczestniczących oraz przedstawianiu wyników w formie studium przypadku. Metodologia opiera się głównie na podejściu jakościowym, stąd

pojęcia studium przypadku (wraz z transakcją) użyłem jako nazwy całego tego podejścia. Co prawda możliwe jest także ilościowe studium przypadku, ale byłoby ono sprzeczne z istotą prezentowanego podejścia. Podejście typu transakcja – studium przypadku jest niemal wyłącznie jakościowe, zarówno w warstwie metodologicznej, jak i przy prezentacji danych. Bardziej adekwatnym określeniem byłaby nawet „historia przypadku".

Celem tego podejścia jest przyjście z pomocą adresatom ewaluacji w lepszym zrozumieniu ewaluowanego programu. W tym celu przedstawia się im, jak odbierają go inni. Korzysta się tu z narzędzi wypracowanych przez antropologię, a nawet dziennikarstwo i literaturę. Podejście to jest zbliżone do etnometologii, a jego metody określa się często jako „naturalistyczne". Najbardziej reprezentatywnym jego przedstawicielem jest Stake (1975; 1980).

Stake dowodził, że „studium przypadku będzie preferowaną metodą badawczą często dlatego, ponieważ może epistemologicznie harmonizować z doświadczeniem adresata i tym samym stanowić naturalną podstawę do generalizacji. [...] Jeśli czytelnikami naszych raportów są osoby, które mieszkają w naszych domach, pracują w naszych szkołach, w naszym rządzie i przemyśle, i jeśli mamy pomóc im zrozumieć problemy i programy społeczne, to nasz sposób patrzenia na nie i komunikowania o nich musimy dostosować do ich sposobu myślenia. Ludzie ci postrzegają świat zarówno w oparciu o własne bezpośrednie doświadczenie, jak i to, które uzyskali z drugiej ręki" (Stake, 1980, s. 64).

Kiedy celem jest raczej zrozumienie niż wyjaśnianie i wnioskowanie na podstawie uzyskanej wiedzy, wtedy studium przypadku ma często przewagę nad innymi typami badań. Odkrywanie lub uzasadnianie praw za pomocą wielu badań ma swój głęboki sens, ale praktycznym celem ewaluacji jest przyczynianie się do realizacji programów. Dla Stake'a cechą wyróżniającą studium przypadku są kompleksowe, holistyczne opisy uwzględniające mnóstwo pozostających ze sobą w interakcji zmiennych. Dane są gromadzone w toku osobistych obserwacji, a pisemne raporty mają charakter nieformalny i narracyjny. Wykorzystuje się przy tym w nich często cytaty, ilustracje, aluzje i metafory. Porównania mają w większym stopniu charakter pośredni niż bezpośredni.

Podejście Stake'a koncentruje się głównie na gromadzeniu obserwacji i informacji w trybie „transakcyjnym" (stąd model ten nazywany jest często modelem transakcji). W „dialogicznej" ewaluacji Stake'a, ewalu-

ator negocjuje zwykle z klientem charakter swoich działań. W konsekwencji wychodzi on naprzeciw oczekiwaniom różnych odbiorców.

„Ewaluacja edukacyjna ma charakter dialogiczny wówczas, gdy zorientowana jest w większym stopniu na działania przewidziane w programie niż na jego zamierzenia docelowe; gdy odpowiada ona na zapotrzebowanie informacyjne odbiorców; gdy sukcesy i niepowodzenia programu rozpatrywane są w pespektywie różnych aktualnie uwzględnianych wartości. Ewaluacja dialogiczna wymaga od ewaluatora sporządzenia planu obserwacji i negocjacji. Zleca on różnym osobom obserwowanie przebiegu programu, następnie z ich pomocą przygotowuje krótkie narracje i opisy, przedstawia wyniki, grafy *etc.* Zyskuje rozeznanie, co ma wartość dla odbiorców, i zapoznaje się z różnymi punktami widzenia w tej sprawie. Oczywiście, weryfikuje także jakość zgromadzonych przez siebie danych, analizuje reakcje personelu programu na przedstawione opisy oraz reakcje odbiorców na dokonane przez siebie odkrycia. Przeprowadza to w sposób nieformalny — poprzez serie powtórzeń oraz notowanie akcji i reakcji. Wybiera media dające dostęp do odbiorców, co zwiększa możliwości w sferze wzajemnej komunikacji i dokładność przekazu. Przy czym przygotowując ostateczny raport pisemny nie musi brać pod uwagę uzgodnień poczynionych z klientami" (Stake, 1975, s. 14).

Stosując bardziej tradycyjną metodologię antropologiczną, Smith i Pohland (1974) ewaluowali komputerowo wspomagane programy nauczania w Appalachach, a Wolcott (1977) program planowania oraz system budżetowy w szkolnictwie lokalnym. W obu przypadkach ewaluatorzy wykorzystali obserwację uczestniczącą, nieformalne wywiady i analizę dokumentów.

W Anglii zwolennikiem studium przypadku był MacDonald (1974). Wykorzystywał je w podejściu, które nazwał ewaluacją „demokratyczną". „Demokratyczne studia ewaluacyjne to służba na rzecz społeczności, którą informuje się o właściwościach programu edukacyjnego. Uwzględnia się tu pluralizm wartości podzielanych przez członków społeczności, a w toku formułowania podstawowych pytań badawczych bierze się pod uwagę możliwie szeroki zakres ich interesów. Podstawową wartością dla ewaluatora jest poinformowana opinia publiczna, a jego rola polega przede wszystkim na pośredniczeniu w wymianie informacji pomiędzy różnymi grupami. Jego techniki gromadzenia materiałów empirycznych i prezentowania wniosków, jakie z tych materiałów wynikają, powinny być zrozumiałe dla nie mających specjalistycznego przygotowania odbiorców. Ewaluator powinien przede wszystkim

dowiedzieć się, jak ludzie reagują na określony program zmiany społecznej i jak go definiują. Uzyskiwane informacje opatrzone są klauzulą poufności, a informatorzy powinni mieć pełną kontrolę nad tym, jak ewaluator je wykorzystuje. Raport konkludujący studium ewaluacyjne nie powinien koncentrować się na rekomendacjach, a informacje nie mogą być przedmiotem manipulacji. Ewaluator systematycznie negocjuje wzajemne relacje ze zleceniodawcami i uczestnikami programu. Podstawowym kryterium jego sukcesu jest to, jak liczne jest grono odbiorców zainteresowanych wynikami jego pracy" (MacDonald, 1974, s. 18).

Inni brytyjscy ewaluatorzy, Parlett i Hamilton, opowiadali się za podejściem, które nazywali ewaluacją „wyjaśniającą". Jej celem są studia nad programem pozwalające ustalić, jak on działa; jaki wpływ ma na niego sytuacja; co osoby bezpośrednio zainteresowane uważają za jego zalety i wady; jaki ma on wpływ na rozwój intelektualny i doświadczenia uczniów w toku nauczania. „Jej celem jest odkrycie i udokumentowanie, co daje uczestniczenie w tym projekcie w roli nauczyciela oraz ucznia, ponadto zaś wyróżnienie i omówienie specyfiki wprowadzonych innowacji, czynników współtowarzyszących oraz kluczowych procesów" (Parlett, Hamilton, 1977, s. 19).

Niedawno akceptujący to podejście Guba (1978b) podjął próbę zdefiniowania trudności, przed którymi stoi ewaluacja „naturalistyczna". Główny problem dotyczy, jego zdaniem, autentyczności ewaluacji, podstawy zaufania pokładanego w jej wynikach. Inne problemy metodologiczne wiążą się z wyznaczeniem granic prowadzonych badań oraz doborem kategorii określających rodzaj danych, które mogą być porównywane i analizowane. Guba kładzie nacisk na deskryptywny charakter metodologii naturalistycznej — nie tylko w podejściu typu studium przypadku, lecz także w modelach krytyki artystycznej i *quasi*-sądowym.

Widać wyraźnie różnice między podejściem typu studium przypadku i podejściem typu „krytyka" (mimo iż są one podobne). Krytyk wykorzystuje swoje własne doświadczenie i stosuje standardy według własnego wyboru. W przypadku studium przypadku chodzi o przedstawienie sposobów postrzegania i wartościowania programu przez innych ludzi, a nie tylko przez ewaluatora. Niekiedy studium przypadku i podejście typu „krytyka" mogą mieć podobny charakter.

Studium przypadku (lub transakcja) zyskuje na popularności. Kiedyś ewaluacje tego typu traktowano jako zbyt subiektywne. Obecnie jest ich coraz więcej. Publikuje się też wiele tekstów na ich temat. Podej-

ście typu studium przypadku rozwiązuje problemy ewaluatora w ten sposób, że próbuje się dopuścić do głosu wszystkie liczące się stanowiska i reprezentowane wartości, przyjmuje sie je za podstawę kryteriów i standardów ewaluacji i umożliwia czytającemu raport samodzielne „ważenie" i „równoważenie" argumentów.

Inne podejścia

Przedstawiłem kilka głównych modeli ewaluacji, zarówno dyskutowanych w literaturze przedmiotu, jak realizowanych w praktyce. Bez wątpienia niektóre z nich nieświadomie pominąłem. Inne wykluczyłem, ponieważ wzorują się na głównych podejściach bądź stanowią ich kombinację. Przedstawione w formie typów idealnych podejścia nie muszą się wykluczać. Przedstawicielami nowych, wyłaniających się podejść są Apple (1979), Lundgren (1977), Kallos (1978), Paulston (w druku) i inni. Mimo dzielących ich różnic wszyscy ci autorzy kierują się ściśle określonymi założeniami w odniesieniu do społeczeństwa. Postrzegają oni społeczeństwo jako organizację generującą nierówności klasowe, edukację zaś jako gwaranta ideologicznych podstaw reprodukcji takiego społeczeństwa.

Jak dotąd, wymienieni autorzy koncentrują się głównie na krytyce istniejących podejść do ewaluacji, a nie na konstruowaniu podejścia opartego na nowych podstawach teoretycznych. Dotychczas niewiele ewaluacji przeprowadzono w oparciu o nowe propozycje. Jedna z nich wiązała się z długookresowymi badaniami przekonań, jakie kształtowały się wśród studentów wyższych studiów zawodowych w Szwecji (Franke-Wikberg, 1979). Takie podejście do ewaluacji mogłoby się przypuszczalnie skoncentrować na nierównościach społecznych występujących w ewaluowanym programie. Jeśli podejście takie rozwinie się w przyszłości, wówczas w ewaluacji powstanie trend zasadniczo różny od dominujących obecnie w ewaluacji podejść liberalnych.

Podsumowując kwestię klasyfikacji ewaluacji według jej typów idealnych, trzeba stwierdzić, że każda ewaluacja jest dyskusyjna. Wielu ewaluatorów będzie przekonanych, że ich podejście jest jedyne w swoim rodzaju i nie podpada pod żadną z wyżej zarysowanych kategorii. W pewnym sensie mają rację. Przedstawione podejścia stanowią typy idealne, konstrukty, które mogą być wykorzystywane do precyzowania założeń myślowych. Oczywiście, każde podejście jest do pewnego stopnia niepowtarzalne.

Z drugiej strony, jestem przekonany, iż wyróżnione przeze mnie podstawowe typy uwzględniają zdecydowaną większość przeprowadzanych obecnie ewaluacji. Istnieją przy tym silne argumenty na rzecz klasyfikowania konkretnych podejść według kategorii ogólnych. Nie znaczy to, że każda ewaluacja będzie kopiować któryś z wyróżnionych przeze mnie typów. W praktyce każda ewaluacja zależy od konkretnych okoliczności i warunków. Na przykład ewaluacje zorientowane na cel mogą znacznie różnić się w zależności od układu okoliczności, mimo to jednak na ogół będą miały najprawdopodobniej zupełnie inny charakter niż ewaluacja typu studium przypadku. Ostatecznym sprawdzianem przedstawionego schematu klasyfikacyjnego będzie zarówno jego trafność, jak przydatność dla zrozumienia ewaluacji. Warto dodać, że podobny pogląd na rolę kategoryzacji ewaluacji znajdziemy u Alkina i Elletta (1979) oraz Alkina, Daillaka i White'a (1979).

tłumaczył *Zbyszko Melosik*

Robert E. Stake

Studium przypadku

Robert E. Stake, *Case Studies*, w: *Handbook of Qualitative Research*, Norman K. Denzin, Yvonna S. Lincoln (eds.), Sage, London 1994.

Nie każde studium przypadku ma charakter jakościowy. W tym rozdziale skoncentruję się na studiach przypadku, w których dominuje podejście jakościowe o wyraźnym nachyleniu naturalistycznym, holistycznym, kulturowym bądź fenomenologicznym. Decydując się na studium przypadku, kierujemy się nie przesłankami metodologicznymi, lecz specyfiką obiektu naszych badań. Moglibyśmy go badać na wiele różnych sposobów. Lekarz bada dziecko, bo jest chore. Choroba ma objawy zarówno ilościowe, jak i jakościowe. W karcie choroby dane ilościowe przeważają nad jakościowymi. Pracownik socjalny bada dziecko, ponieważ jest ono zaniedbane. Zaniedbanie objawia się zarówno ilościowo, jak i jakościowo. W dokumentacji pracownika socjalnego przeważa opis jakościowy[1]. Studium przypadku stosuje się w wielu dziedzinach zawodowych i praktycznych. Kryterium definicyjnym studium przypadku jako sposobu prowadzenia badań są nie zastosowane metody, ale zainteresowanie indywidualnymi cechami przypadku.

[1]Studium przypadku może być jakościowe, ilościowe lub kombinowane. W poszukiwaniu najważniejszych cech wspólnych podejść jakościowych i ilościowych Robert Yin (1992) przeprowadził analizę trzech bardzo starannie przygotowanych studiów przypadku: ilościowego, którego celem było rozstrzygnięcie sporu co do autorstwa *Federalist Papers*; jakościowego, zmierzającego do rozszyfrowania zamiarów Związku Sowieckiego w czasie kryzysu kubańskiego; własnych badań nad rozpoznawaniem ludzkich twarzy. Yin wyodrębnił cztery komponenty wspólne analizowanym studiom przypadku: wykorzystanie w badaniu danego zjawiska wiedzy ekspertów, zgromadzenie wszystkich danych istotnych dla badanego zagadnienia, przetestowanie interpretacji konkurencyjnych, rozważenie, jak uzyskane wyniki mogą rzutować na inne badania. Są to komponenty ważne dla każdego badacza, bez względu na to, jaką stosuje metodę.

Niewykluczone, że większość badaczy zajmujących się studium przypadku posługuje się innymi określeniami. Howard Becker, na przykład, zapytany podczas drugiej Konferencji Cambridge (Simons, 1980), jak określa prowadzone przez siebie badania, odpowiedział z pewnym ociąganiem: „badaniami terenowymi" (*fieldwork*), dodając przy tym, że posługiwanie się tego typu etykietkami w niewielkim stopniu ułatwia zrozumienie, czym badacz naprawdę się zajmuje. Niektórzy z nas celowo kładą nacisk na nazwę *studium przypadku*, ponieważ zwraca ona uwagę na specyficzny rodzaj wiedzy, jakiej dostarcza pojedynczy przypadek. Czynnikiem organizującym ten rozdział będzie właśnie następujące pytanie epistemologiczne: Czego można się dowiedzieć, badając pojedynczy przypadek? Główny nacisk chcę położyć na schemat badawczy po to, by jak najlepiej zrozumieć przypadek — niezależnie od możliwości wynikających z niego uogólnień.

Identyfikacja przypadku

Przypadki bywają proste i złożone. Przypadkiem może być dziecko, klasa szkolna, działanie specjalistów podjęte w celu zbadania określonego problemu wieku dziecięcego. Wybrany do badań przypadek jest jednym z wielu. W każdym konkretnym badaniu koncentrujemy się na tym jednym jedynym przypadku. Możemy mu poświęcić wiele bądź niewiele czasu, ale dopóki się nim zajmujemy, prowadzimy studium przypadku.

Zwyczajowo utarło się, że nie wszystko jest przypadkiem. Przypadkiem może być dziecko. Może nim być lekarz, natomiast *zabiegi lekarskie* są już czymś zbyt nieokreślonym i rozmytym, by można je było nazwać przypadkiem[2]. Przypadkiem może być ośrodek opieki społecznej. Rzadko natomiast nazwalibyśmy przypadkiem powody, dla których rodzice zaniedbują swoje dzieci czy strategie radzenia sobie z takimi rodzicami. Są to bowiem problemy raczej ogólne niż szczególne. Przypadek zawsze jest czymś szczególnym. Co więcej, jest specyficzną jednostką funkcjonalną. Louis Smith (1978) nazywa przypadek „systemem ograniczonym". Przypadki, którymi zajmują się nauki społeczne i placówki usługowe, składają się z szeregu elementów funkcjonalnych, zwykle działają w sposób celowy, miewają nawet własne „ja". Przypa-

[2]Redaktorzy przypomnieli mi o etnometodologicznym podejściu do kwestii tematu i metody. Etnometodologowie traktują metodę jako temat badań, analizują sposoby działania itp. (Garfinkel, 1967). Zrozumienie przypadku zwykle poprzedza drobiazgowa analiza, jak to, o co chodzi, działa, jednak głównym przedmiotem studium przypadku są nie metody działania przypadku, lecz sam przypadek.

dek jest systemem zintegrowanym. Jego części składowe nie zawsze działają sprawnie, cele, ku którym zmierza, bywają irracjonalne, ale mimo wszystko jest to system.

Funkcjonowanie tego systemu podporządkowane jest pewnym wzorom. Na plan pierwszy wysuwa się spójność i sekwencyjność zachowania. Powszechnie uważa się, że pewne zachowania są częścią składową systemu, mieszczą się w obrębie danego przypadku, inne natomiast poza niego wykraczają, choć mogą stanowić jego istotny kontekst. Jak zauważyli William Goode i Paul Hatt (1952), osoba badająca dany przypadek nie zawsze potrafi określić, czy ma do czynienia jeszcze z dzieckiem, czy już z jego środowiskiem. Mimo to istnienie granic systemu i wzory jego zachowań mają kluczowe znaczenie dla zrozumienia danego przypadku (Stake, 1988).

Czasami interesuje nas szersze zjawisko albo cała populacja przypadków, bez poznania których nie jesteśmy w stanie zrozumieć tego, który badamy. Ale pracując nad danym przypadkiem wszystkie swoje skromne środki koncentrujemy właśnie na nim, starając się go zrozumieć w całej jego złożoności. Powstrzymujemy się na razie przed porównywaniem go z innymi. Możemy badać kilka przypadków na raz, niemniej jednak każde studium przypadku polega na skupieniu uwagi na jednym jedynym.

Samo pojęcie *przypadku* jest przedmiotem sporów[3], a termin *studium* nie jest terminem jednoznacznym (Kemmis, 1980). Określenie „studium przypadku" odnosi się zarówno do procesu poznawania przypadku, jak i do produktu tego procesu. Lawrence Stenhouse (1984) proponuje, by produkt końcowy nazwać „protokołem przypadku" (*case record*) i rzeczywiście czasem tak go nazywamy, ale przyjęło się również, aby dokument końcowy nazywać „studium przypadku" (*case study*). W sumie zasady nie zostały jeszcze na tyle ustalone, by ktokolwiek z badaczy (chyba że jest magistrantem) popadł w tarapaty z tego powodu, że posługuje się pojęciem studium przypadku według własnego upodobania[4]. Im szczególniejszym, bardziej niepowtarzalnym i lepiej wyodrębnio-

[3]Definicja przypadku nie jest niezależna od przyjętego paradygmatu interpretacyjnego i stosowanych metod badawczych. Ten sam przypadek z perspektywy różnych światopoglądów i w różnych sytuacjach nie jest taki sam. I bez względu na to, jak go zdefiniujemy wstępnie, w trakcie pracy nad przypadkiem jego definicja robocza się zmienia, również w zależności od stosowanych metod. Biograf Teodora Roosevelta, Edmund Morris (1979), nie tylko różnie przedstawia swego bohatera, ale zgoła różnie go definiuje, nazywa „gogusiem z Nowego Jorku", „drogim, starym, kochanym bratem", „wężem ukrytym w trawie", „brutalem", „najsławniejszym człowiekiem w Ameryce" itd.

[4]Historia studium przypadku, podobnie jak historia ciekawości czy zdrowego rozsądku, rozsiana jest po całej literaturze. Czytelnika poszukującego krótkiego przeglądu tej

nym systemem jest przedmiot naszych badań, tym przydatniejsze będą opisane w tym rozdziale założenia epistemologiczne.

Zainteresowanie przypadkami dla nich samych a zainteresowanie instrumentalne

Zainteresowanie pojedynczymi przypadkami może wynikać z różnych pobudek. Dla lepszego uświadomienia tych różnic warto wyodrębnić trzy typy studium przypadku. Typ pierwszy — nazwijmy go *autotelicznym* (*intrinsic case study*) — podejmowany jest w celu lepszego poznania konkretnego przypadku. Badacz nie dlatego podejmuje badania nad danym przypadkiem, że reprezentuje on inne, i nie dlatego, że dobrze ilustruje określoną cechę lub problem, ale dlatego, że choć pojedynczy i powszedni jest interesujący sam w sobie. Badacz zawiesza na pewien czas ciekawość innych zjawisk, by pozwolić danemu przypadkowi ujawnić swoją historię. Celem całego przedsięwzięcia nie jest zrozumienie jakiegoś abstrakcyjnego konstruktu czy zjawiska ogólniejszego, np. biegłości w mowie i piśmie, narkotyzowania się młodzieży dorastającej czy zakresu obowiązków dyrektora szkoły. Celem nie jest konstruowanie teorii (choć skądinąd i tym się badacz zajmuje). Przypadek — konkretne dziecko, konkretna poradnia czy konferencja, ten a nie inny program nauczania — badany jest dlatego, że sam w sobie jest ciekawy. Oto kilka przykładów książek prezentujących autoteliczne studia przypadku:

- *Akenfield* (Blythe, 1955/1969),
- *Argonauci Zachodniego Pacyfiku* (Malinowski, 1987),
- *Bread and Dreams: A Case Study of Bilingual Schooling in the U.S.A.* (Chleb i marzenia. Studium przypadku szkolnictwa dwujęzycznego w Stanach Zjednoczonych) (MacDonald, Adelman, Kushner, Walker, 1982),
- *God's Choice* (Boski wybór) (Peshkin, 1986).

O *instrumentalnym studium przypadku* (*instrumental case study*) mówimy wtedy, gdy badania określonego przypadku służyć mają lepszemu zrozumieniu jakiegoś innego problemu, udoskonaleniu teorii itp. Sam przypadek ma znaczenie drugorzędne; pełni wyłącznie rolę pomocniczą, ułatwia zrozumienie czegoś poza nim samym. Nie wyklucza to starannego zgłębiania danego przypadku, starannej analizy rozmaitych jego kontekstów, szczegółowego opisu charakterystycznych dla

problematyki odsyłam przede wszystkim do prac Bogdana i Biklena (1982), Delamonta (1992), Feagina, Oruma i Sjoberga (1991) oraz Stake'a (1980).

niego form aktywności, ale tylko po to, by lepiej zrozumieć coś, co jest względem niego zewnętrzne. Dany przypadek może, ale nie musi być typowym egzemplarzem innych przypadków (o tym, jak niewielkie znaczenie ma typowość powiem nieco później). Przypadek ten wybrany został dlatego, że daje nadzieję głębszego zrozumienia czegoś poza nim samym. Zwykle interesujemy się jednak kilkoma rzeczami naraz, często przelotnie, więc nie da się ostro odgraniczyć studium przypadku autotelicznego i instrumentalnego. Łączy je strefa zbieżnych celów. Przykłady instrumentalnych studiów przypadku znaleźć można w następujących książkach:

- *A Bright and Shining Lie: John Vann and America in Vietnam* (Wierutne kłamstwo. John Vann i Ameryka w Wietnamie) (Sheehan, 1988),
- *Boys in White: Student Culture in Medical School* (Chłopcy w bieli. Kultura studencka w szkole medycznej) (Becker, Geer, Hughes, Strauss, 1961),
- *Middletown: A Study of American Culture* (Middletown. Studium kultury amerykańskiej) (Lynd, Lynd, 1929),
- *La Vida* (Lewis, 1966).

Jeszcze mniejsze zainteresowanie pojedynczym przypadkiem wykazują badacze analizujący łącznie całą ich serię, by lepiej poznać określone zjawisko, daną populację czy stan ogólny. Podejście to można by nazwać *zbiorowym studium przypadku* (*collective case study*)[5]. Nie jest to badanie zbioru przypadków, lecz instrumentalne studium przypadku rozciągnięte na kilka przypadków. Badacz może, ale nie musi, wiedzieć z góry, że pojedyncze przypadki wchodzące w skład zbioru mają jakąś cechę wspólną. Poszczególne przypadki mogą być do siebie podobne, ale mogą też się różnić. Pewne cechy mogą się powtarzać, inne nie. Wybierając do badań te, a nie inne przypadki, badacz kieruje się przekonaniem, że lepsze zrozumienie tych właśnie przypadków pozwoli mu lepiej zrozumieć, a być może i lepiej ująć teoretycznie, zbiór jeszcze liczniejszy. Oto kilka przykładów książek obrazujących zbiorowe studium przypadku:

- *Children of Crisis* (Dzieci kryzysu) (Coles, 1967),
- *Habits of the Heart: Individualism and Commitment in American Life* (Z nawyku serca. Indywidualizm i zaangażowanie w życiu amerykańskim) (Bellah, Madsen, Sullivan, Swindler, Tipton, 1985),

[5] W zasadzie zbiorowe studium przypadku pokrywa się z tym, co Robert Herriott i William Firestone (1983) nazywają „badaniem jakościowym wielomiejscowym" (*multisite qualitative research*). Wielu socjologów niemieckich, np. Martin Kohli i Fritz Schütze, prowadziło zbiorowe studia przypadku w oparciu o teorię trajektorii życiowej Anselma Straussa.

- *Innovation up Close: How School Improvement Works* (Innowacja z bliska. Jak działa reforma szkolna) (Huberman, Miles, 1984),
- *Savage Inequalities* (Dzikie nierówności) (Kozol, 1991).

Mało który autor i mało który raport z badań da się precyzyjnie wpasować w taką kategoryzację, toteż trzy wyróżnione przeze mnie kategorie mają znaczenie raczej heurystyczne niż funkcjonalne. Peshkin wiadomość, jak zaszeregowałem jego pracę *God's Choice*, skomentował słowami: „Chcę tak przedstawić mój przypadek, by sam w sobie wzbudził zainteresowanie, ale w tle zawsze kryje się inny zamiar — przez pryzmat przypadku chcę dowiedzieć się czegoś o całej klasie jakichś zjawisk. Jakiej? Trzeba poczekać, zanim to sie do końca wykluje". Trzy lata poświęcił Peshkin na studiowanie jednej szkoły, Bethany Baptist Academy. O problemach, jakie się wyłoniły w toku tych badań, problemach niezwykle ważnych dla niego samego, dla środowiska szkolnego i dla ogółu obywateli, informuje czytelnika dopiero w ostatnim rozdziale. W pierwszej kolejności chodziło o zrozumienie przypadku. Bezpośredni powód zainteresowania się badanym obiektem tkwił w obiekcie jako takim. Metody, jakimi posługiwał się Peshkin, podyktowane były nie względami na społeczność, wolność czy możliwość przetrwania, ale na charakter samego przypadku.

Można by wymienić inne jeszcze typy studium przypadku. Jeden z nich, powszechnie stosowany w procesie nauczania, można by nazwać instruktażowym studium przypadku (*teaching case study*). Korzysta się z niego wtedy, gdy chce się zilustrować określony punkt widzenia, stan rzeczy, kategorię zjawisk, coś istotnego ze względu na proces nauczania (Kennedy, 1979). Wykładowcy Wydziału Prawa i Wydziału Zarządzania Uniwersytetu im. Harvarda od kilkudziesięciu lat afiszują się z tego typu przypadkami. Artykuły publikowane w „Journal of Case Research", najważniejszym czasopiśmie North American Case Research Association (Północnoamerykańskiego Stowarzyszenia Badaczy Przypadków), adresowanego do osób zajmujących się doskonaleniem zawodowym i szkoleniem kadr kierowniczych, to studia przypadku. Mają one charakter instruktażowy i konsultacyjny, ale biorą początek ze studium przypadku, który określiłem jako instrumentalne[6].

Do oddzielnej kategorii można by zaliczyć biografie. Kiedy Louis Smith zwraca uwagę, że autorzy biografii powinni nadawać swym pra-

[6]Badanie pojedynczego epizodu czy pojedynczej zależności jest powszechną praktyką historyków oraz politologów (*nide* oblężenie Moskwy przez Napoleona, kryzys kubański itp.). Ja sam nie nazwałbym tego typu badań studiami przypadku, ponieważ — niezależnie od stopnia złożoności, oddziaływań i powiązań — ani epizod, ani zależność, nie mają tożsamości i nie realizują własnych celów.

com ścisłą strukturę chronologiczną i bezwzględnie przestrzegać procedur chroniących osoby badane, również koncentruje się na przypadku. Specyficznymi dla siebie metodami posługuje się również telewizyjny film dokumentalny, też spełniający często kryteria studium przypadku. Inaczej definiują przypadek prawnicy. Operowanie prawem w praktyce samo w sobie jest niczym innym jak studium przypadku. Etnografowie, krytycy, demografowie i przedstawiciele wielu innych dyscyplin posługują się w swojej pracy schematami pojęciowymi i formalnymi, które nie tylko rozwijają systematykę rzeczywistości, ale także poszerzają bazę dla studiów przypadku w naukach społecznych i służbach socjalnych. Wprowadzona przeze mnie kategoryzacja ma wszakże bardziej ograniczony cel: wyodrębniając trzy typy studium przypadku, autoteliczny, instrumentalny i zbiorowy, pragnę zwrócić przede wszystkim uwagę na różne rozłożenie akcentów, jeżeli chodzi o zainteresowanie przypadkiem jako takim i metodologiczne podejście do niego.

Koncentracja na tym, co szczególne

Badacze zajmujący się studium przypadku szukają w danym przypadku zarówno tego, co go łączy z innymi przypadkami, jak i tego, co w nim szczególne, specyficzne, przy czym raport końcowy uwypukla zwykle to, co wyjątkowe, niepowtarzalne (Stouffer, 1941). To nastawienie na niepowtarzalność znamionuje wszystkie komponenty postępowania badawczego, rozciągając się na:

1 — rodzaj przypadku,
2 — jego tło historyczne,
3 — otoczenie fizyczne,
4 — inne konteksty (ekonomiczny, polityczny, prawny, estetyczny itp.),
5 — inne przypadki, na tle których wyróżnia się przypadek badany,
6 — osoby dostarczające informacji o przypadku.

Wielu badaczy, studiując dany przypadek, zbiera wszystkie wymienione wyżej rodzaje danych.

Nie każdy jest wielbicielem tego, co niepowtarzalne, szczególne, odmienne. Metodologia studium przypadku traci niekiedy na tym, że prezentuje ją ktoś, kto ma niewielki respekt dla badań nad czymś, co jedyne w swoim rodzaju (Denzin, 1978; Glaser, Strauss, 1967; Herriott, Firestone, 1983; Yin, 1989). Wielu socjologów pisząc o studium przypadku daje do zrozumienia, że badanie pojedynczego przypadku ze względu na jego indywidualne walory poznawcze ma mniejsze

znaczenie niż badanie populacji przypadków prowadzące do uogólnień. Autorzy ci zwracają uwagę, że przypadek należy traktować jako typowy przykład wielu podobnych przypadków, jako badanie eksploracyjne, za którym powinny iść badania zmierzające do formułowania uogólnień, a niekiedy jako wstępny krok w konstruowaniu teorii. Innymi słowy, studium przypadku jako metody poznawania czegoś, co ma autonomiczną wartość jako egzemplarz — np. biografii, autoanalizy instytucjonalnej, ewaluacji programu, praktyki terapeutycznej, ale i wielu innych dziedzin praktycznych — te szacowne autorytety nie darzą zbytnim szacunkiem. Natomiast w naukach społecznych posługujących się metodami jakościowymi daje sie zauważyć coraz mniejszy nacisk na konstrukcję teorii jako ostateczny cel badań.

Traktowanie studium przypadku jako drobnego kroku na drodze ku wielkim uogólnieniom ma swoje zalety (Campbell, 1975), ale nie w każdym badaniu trzeba zmierzać do generalizacji za wszelką cenę (Feagin, Orum, Sjoberg, 1991; Simons, 1980). Gdy przemożne dążenie do formułowania uogólnień lub budowania teorii odwraca uwagę badacza od elementów istotnych dla zrozumienia badanego przypadku, nic dobrego z tego wyniknąć nie może[7]. Każdy badacz biorący się za studium przypadku musi podjąć strategiczną decyzję: musi odpowiedzieć sobie na pytanie, jak szczegółowo i jak długo zamierza zajmować się danym przypadkiem. Nie wszystko uda mu się wyjaśnić, więc co wyjaśnić bezwzględnie powinien, a czego nie? Każdy badacz musi sobie sam odpowiedzieć na to pytanie.

Wyjątkowość sytuacji

Każdy przypadek ma własną niepowtarzalną historię i tworzy złożoną całość funkcjonującą w wielu różnych kontekstach (fizycznym, ekonomicznym, etycznym, estetycznym itd.). Przypadek jest jeden, ale można w nim wyodrębnić szereg jednostek składowych (np. produkcja, marketing, działy sprzedaży), grup (np. uczniowie, nauczyciele, rodzice), okoliczności (np. dni robocze, dni wolne od pracy, dni sąsiadujące z dniami wolnymi od pracy). Innymi słowy, każdy jednostkowy przypadek składa się z wielu wzajemnie powiązanych obszarów, tak niekiedy złożonych, że do badań można wybrać co najwyżej kilka próbek.

[7] „Jednym z podstawowych warunków wszelkiej racjonalnej pracy etnograficznej jest to, by praca ta podejmowała całokształt zagadnienia ukazanego w społecznych, kulturowych i psychologicznych aspektach życia danej społeczności" (Malinowski, 1987, s. 21). Sporo w tej wypowiedzi optymizmu, lecz całokształty umykają badawczemu spojrzeniu i za nic mają czas odmierzany zegarkiem.

Gdy chce się dany przypadek potraktować holistycznie, należy go badać w całej jego złożoności. Wiele badań jakościowych wychodzi z holistycznego założenia, że zjawiska społeczne, dylematy ludzkie, istota przypadków itp. mają charakter sytuacyjny i podlegają wpływom wielu różnych zdarzeń. Zwolennicy podejścia jakościowego zakładają niekiedy, że każde zdarzenie ma swoją przyczynę, częściej jednak podzielają pogląd Tołstoja wyrażony w *Wojnie i pokoju*, że przyczyna nigdy nie jest ani jedna, ani prosta. Wielu sądzi, że poszukiwanie przyczyn to strata czasu, i woli koncentrować uwagę na zbieżności zdarzeń, jedne postrzegając jako zamierzone, inne jako przypadkowe, liczne jako wzajemnie powiązane. Preferują oni poszukiwanie danych opisujących zachowania danego przypadku w różnych warunkach. Studium przypadku nie wymaga stosowania procedur sprawdzających wobec różnych problemów i kontekstów, niemniej jednak większość stosujących metody jakościowe tak właśnie postępuje.

Problemy badawcze

Bez względu na to, czy badacz tak go nazywa — temat, „abstrakcyjny wymiar", na jakim się koncentruje, często zahacza o jakiś *problem*, o jakieś problematyczne okoliczności przykuwające uwagę przedstawicieli różnych dyscyplin, np. socjologii, ekonomii, etyki czy krytyki literackiej. Badacze przypadków stosujący metody jakościowe szerzej widzą zagadnienie aniżeli zwolennicy metod eksperymentalnych i testowania hipotez, nastawieni są na kompleksowe łączenie codziennej praktyki w środowisku naturalnym z myśleniem abstrakcyjnym i zainteresowaniami typowymi dla różnych dyscyplin akademickich. To szersze spojrzenie stosują w odniesieniu do pojedynczego przypadku. Formułowanie uogólnień i ich udowadnianie (Becker, 1958) kryje w sobie jednak ryzyko błędu[8].

Nawet jeśli problemy badawcze podnosi się do rangi uogólnień, bada się je w jednym, konkretnym przypadku. Badacz znając temat, formułuje problemy przewidywane[9], po czym przystępuje do związanych z badanym problemem obserwacji, by przejść następnie do

[8]Formułowanie uogólnień na podstawie zbiorowych studiów przypadku omawiają Herriott i Firestone (1983), Lofland i Lofland (1984), Miles i Huberman (1984) i raz jeszcze Firestone (w druku).

[9]Malinowski twierdzi, że można wyróżnić dwie postawy wyjściowe: zamkniętą i wiedzącą czego szukać. „Dobra orientacja w teorii oraz znajomość jej najnowszych osiągnięć nie jest równoznaczna z obciążeniem «z góry przyjętymi koncepcjami». Gdy ktoś wyrusza na ekspedycję z zamiarem udowodnienia z góry przyjętych hipotez, a nie zdoła swych poglądów stale modyfikować i odrzucać bez większych oporów, od presją faktów, to zbyteczne jest podkreślać, że jego praca będzie bezwartościowa. Im

Przykład ewolucji problemu w toku badań

1. *Temat:* Cele programowe edukacji muzycznej.
2. *Problem przewidywany:* Większość społeczności lokalnej popiera dotychczasowy nacisk na granie w orkiestrze, śpiew chóralny i występy publiczne, ale niektórzy nauczyciele i działacze społeczni woleliby więcej uwagi poświęcać zagadnieniom bardziej rozwijającym intelektualnie, np. historii muzyki, literaturze muzycznej, krytyce muzycznej.
3. *Konkretyzacja problemu badawczego:* W jakim stopniu badana kadra nauczycielska zainteresowana jest prowadzeniem obowiązkowych lekcji muzyki?
4. *Teza końcowa:* Badana społeczność lokalna nie zgodziłaby się na wygospodarowanie dodatkowych funduszy na lekcje teoretycznych przedmiotów muzycznych dla wszystkich uczniów szkół średnich.

interpretacji tych danych, które układają się w jakąś logiczną całość i pozwalają nadać postawionym problemom badawczym postać twierdzeń. Ten proces ewolucji przedstawia tabela powyżej.

Wybór problemów organizujących studium przypadku podporządkowany jest określonemu celowi. Gdy przypadek ma egzemplifikować podobne przypadki, podstawowe pytanie badacza brzmi: Jakie problemy najpełniej uwydatniają nasze zainteresowania wyjściowe, podstawowy temat? Chcąc jak najlepiej zrozumieć dany przypadek, badacz zadaje inne pytanie: Jakie problemy należy drążyć, by ujawnić to, co w tym przypadku najbardziej niepowtarzalne? Z kolei podstawowe pytanie badacza, którego głównym celem są studia ewaluacyjne, brzmi: Jakie problemy należy wybrać, by jak najlepiej uwydatnić zalety i wady tego przypadku? Wszyscy zaś bez wyjątku zadają jedno pytanie ogólne: Jakie problemy ułatwią zaplanowanie tego badania i jego przebieg, inspirując badacza i dając nadzieję na sukces? Jednym z kryteriów wyboru problemów badawczych jest ich wartość poznawcza w dostępnych warunkach. Wybór ten podyktowany jest zarówno ważnością celu, do jakiego badacz zmierza, jak i indywidualnymi preferencjami autora. Można powiedzieć, że każdy badacz podpisuje z badanym zjawiskiem osobisty kontakt. Przy całym szacunku dla nauki i klienta, można by zadać pytanie następujące: Co takiego wnosi dany przypadek, że konkretny badacz nie może się oprzeć pokusie jego studiowania?

Nie zawsze problemy *organizujące* studium przypadku znajdują odzwierciedlenie w raporcie końcowym. Czym innym jest prowadzenie

więcej problemów i hipotez niesie ze sobą w teren, tym bardziej będzie skłonny do podporządkowania swych teorii faktom i widzenia faktów w oddziaływaniu na teorię, tym lepiej jest przygotowany do swej pracy badawczej. Z góry powzięte idee są zgubne w każdej pracy naukowej, ale umiejętność widzenia problemów, wyobraźnia naukowa stanowi podstawowe wyposażenie badacza naukowego, a rola studiów teoretycznych polega właśnie na tym, że naprowadzają obserwatora na te problemy" (1987, s. 37).

obserwacji, czym innym zaś raport prezentujący przypadek. Decydując się, co umieścić w końcowym sprawozdaniu, a co pominąć, badacz kieruje się niekiedy zainteresowaniami odbiorcy. Nawet wtedy, gdy nie do końca wiadomo, kim jest potencjalny odbiorca i co go interesuje, część badaczy pisze przede wszystkim z myślą o nim.

Sposób prezentacji przypadku

Badacze stosujący metody jakościowe typu studium przypadku apelują niekiedy, by pozwolić, aby przypadek „sam opowiedział swoją historię" (Carter, 1993; Coles, 1989). Nie ma żadnej gwarancji, że przypadek, który sam opowiada swoją historię, opowie ją dobrze i wyczerpująco. Mimo to etnograficzny etos badań *interpretacyjnych* (*interpretative study*), nastawiony na docieranie do indywidualnych znaczeń, jakie przypisuje rzeczywistości każda z badanych w ramach danego przypadku jednostek, trzyma się mocno. Istnieje wiele sposobów prezentacji przypadku. John Van Maanen (1988) wyróżnia siedem stylów prezentacji: realistyczny, impresyjny, osobisty, krytyczny, formalny, literacki i opowieść zbiorową. Nie sposób przewidzieć w punkcie startu, jakie zostaną poruszone problemy i ujawnione spostrzeżenia, jaka wyłoni się z tego teoria. Badacz przystępujący do studium przypadku spodziewa się, czasem jest wręcz przekonany, że pewne zdarzenia, zagadnienia, relacje okażą się ważne, a przecież często okazuje się później, że jest inaczej (Parlett, Hamilton, 1977). Warstwa treściowa przypadku ewoluuje w trakcie pisania.

Nie kto inny przecież niż sam badacz, choć zobowiązany do wczuwania się i uwzględniania wielu wymiarów rzeczywistości, decyduje o tym, co stanowi „własną historię" danego przypadku i jak tę historię ująć. Jego poszukiwania zatoczą szerszy krąg, aniżeli początkowo zamierzał. Nie wszystko, czego się dowiedział, znajdzie się w końcowym raporcie. I choć kompetentny badacz kieruje się tym, co sam przypadek uzna za ważne, choć mecenasi i koledzy po fachu podsuwają własne pomysły — nie kto inny jak tylko on sam podejmuje ostateczną decyzję, co jest dla zrozumienia danego przypadku konieczne. Jeśli nawet podstawowe składniki historii opowiada sam przypadek, sos przyrządza badacz. Nie znaczy to bynajmniej, że należy dążyć do znalezienia takiej formuły, która zaprezentuje dany przypadek najlepiej. Chodzi tylko o przypomnienie, że o kryteriach przedstawienia decyduje w ostatecznym rozrachunku badacz.

Niejeden badacz chciałby opowiedzieć wszystko, ale nie jest to możliwe: nikt nie zna całej historii, a gdyby ją znał, nie potrafiłby jej

opowiedzieć. Nawet badacze o skrajnie maksymalistycznych zapędach czują się silnie zobligowani do oddzielenia ziarna od plew i przedstawienia spójnego opisu. Można powiedzieć bardzo wiele albo zgoła nic. Między tymi dwoma biegunami rozciąga się wiele możliwości. Zarówno badacz nastawiony holistycznie, jak i badacz nastawiony na rozwikłanie jednego tylko problemu musi wybierać. Kryteriów wyboru jest wiele (Van Maanen, 1988). Niektóre ustalają fundatorzy, inne narzucają potencjalni odbiorcy, konwencja wypowiedzi, model kariery badacza, możliwość publikacji. Niektóre kryteria określa zasada, że przypadek należy przedstawić w sposób pełny, maksymalnie wszechstronny, zadość czyniący okazanej gościnności. Jak każdy wybór obiektu badań, tak i ten wybór jest subiektywny. Pewne decyzje podejmuje się już w fazie planowania studium przypadku, inne — w trakcie jego trwania, aż po ostatnie godziny pracy nad przypadkiem.

Czego uczy konkretny przypadek

Praca badacza przypomina pracę nauczyciela. Nauczyciel może stosować dwojakiego rodzaju metody (Eisner, 1985). Metoda pierwsza, tzw. *dydaktyczna*, polega na uczeniu tego, co się samemu wie. Metoda druga, zwana przez pedagogów *uczeniem się przez odkrywanie*, polega na dostarczaniu materiału, na podstawie którego uczący się sam zdobywa wiedzę — zarówno tę, którą posiada już nauczyciel, jak i tę, której nauczyciel nie posiada. Czego zatem można się nauczyć badając pojedynczy przypadek? Epistemologią tego, co szczególne, indywidualne zajmują się Donald Campbell (1975), David Hamilton (1980c), Stephen Kemmis (1980), Robert Yin (1989) i William Firestone (w druku). O tym, czy badając pojedynczy przypadek możemy się czegoś nowego dowiedzieć, decyduje stopień, w jakim badany przypadek przypomina inne przypadki, a w jakim się od innych różni. Ale — jak wykażę później — bezpośrednie porównywanie jest niezbyt pouczające.

Raporty relacjonujące studia przypadków dostarczają dwojakiego rodzaju wiedzy: postulatywnej (*propositional knowledge*) i odwołującej się do doświadczenia (*experiential knowledge*) (Geertz, 1983; Polanyi, 1962; Rumelhart, Ortony, 1977; von Wright, 1971). Czytelnik asymiluje w pamięci pewne opisy i twierdzenia. A zatem, gdy badacz swoją narracją stwarza okazję do doświadczeń zastępczych, czytelnik uzupełnia swoje pamięciowe zapisy zdarzeń o nowe treści. Opis przypadku zawierający materiał przyrodniczy czy etnograficzny uzupełnia w pewnej mierze doświadczenia bezpośrednie, zasilając najbardziej podstawowe procesy

świadomości i rozumienia. Procesy te nazwaliśmy z Deborah Trumbull (Stake, Trumbull, 1982) *uogólnianiem naturalistycznym* (*naturalistic generalization*). Czytelnik dowiaduje się o tym, o czym opowiada autor, tak jak gdyby sam tego doświadczał. Staje się to źródłem trwałych znaczeń, modyfikowanych i wzmacnianych w czasie następnych tego typu „spotkań".

W realnym życiu jednostka rzadko dokonuje uogólnień w samotności, częściej w obecności (nie zawsze bezpośredniej) innych ludzi. Jest to zatem proces społeczny. Ludzie wspólnie naginają, komponują, utrwalają i wzbogacają swój sposób rozumienia pewnych zjawisk. O tym, co się wydarzyło, dowiadujemy się między innymi słuchając, co inni mają do powiedzenia na temat swoich doświadczeń. Badacz przygotowujący studium przypadku wychodzi od jednego doświadczenia społecznego, od obserwacji, i aranżuje inne, jakim jest raport. My, konstruktywiści, uważamy, że każda wiedza jest konstrukcją społeczną. Innymi słowy, autor studium przypadku pomaga swoim odbiorcom wiedzę tę konstruować.

W jaki sposób badacz przekazuje wiedzę odbiorcom

Zarówno badacz, jak i odbiorca potrzebują struktur pojęciowych, nadrzędnych struktur porządkujących (*advanced organizers*) (Ausubel, Fitzgerald, 1961), schematów (Anderson, 1977), rusztowania (*scaffolding*) (Cazden, 1988), wykładni realizacyjnej (*unfolding of realization*) (Bohm, 1985). Potrzeba ta nie musi być uświadamiana. Myśląc, rozmawiając z kimś, a zwłaszcza pisząc, człowiek łączy poszczególne zwroty w większe całości, formuje z określeń konstrukty. Funkcję ogniskującą pełni uwaga. Generalizacja nie musi być procesem świadomym.

Każda idea podlega prywatnej, osobistej strukturalizacji, uwydatnieniu, podporządkowaniu, łączeniu, osadzeniu w kontekście, osadzeniu *wraz* z odpowiednią ilustracją, przyjmowana jest chętnie lub z wątpliwościami. Bez względu na to, jak bardzo pragnie się podzielić swoimi pomysłami z innymi, jak mądre i rozbudowane jest to, co pisze, autor studium przypadku, jak każdy piszący, przekazuje swym czytelnikom tylko niektóre osobiste interpretacje zdarzeń i zależności między nimi. Wie, że czytelnik coś od siebie doda, coś opuści, sam coś wymyśli, coś doprecyzuje; innymi słowy, zrekonstruuje wiedzę, inaczej ją powiąże, uczyni bardziej dla siebie użyteczną.

Droga, jaką przebywa wiedza zanim trafi od piszącego do odbiorcy, najeżona jest niebezpieczeństwami. Piszący musi zatem przesięwziąć odpowiednie środki ostrożności. Już w momencie, gdy czytający przy-

stępuje do czytania, a często nawet o wiele wcześniej, opisywany przypadek ma swoje miejsce wśród znanych mu wcześniej przypadków. Z punktu widzenia czytelnika nowy przypadek może być tylko kombinacją przypadków już znanych. Przypadek nie mający żadnych cech wspólnych z dotychczas poznanymi byłby niezrozumiały. Ale z drugiej strony, gdyby nowy nie różnił się niczym od dotychczasowych, nie zasłużyłby na uwagę. Badacz nie wie, jakie przypadki czytelnik już poznał ani jakimi ścieżkami wędrują jego myśli, i dlatego zawsze zastanawia się, jak zabezpieczyć i uprawomocnić przekazywaną wiedzę.

Badacz zdaje sobie sprawę, że musi się dostosować do zasobu wiedzy czytelnika. Każdy z nas zmuszony jest to czynić codziennie i czerpie w tym celu z doświadczeń całego życia, a mimo to wciąż tak niewiele wiadomo o mechanizmach asymilacji nowych doświadczeń. Spiro, Vispoel, Schmitz, Samarapungavan i Boerger (1987) twierdzą, że większość doświadczeń osobistych jest „źle ustrukturowana" — nie ma w nich ładu ani pedagogicznego, ani epistemologicznego. Płynie z tego wniosek następujący: najlepszym sposobem „przekazywania" wiedzy odwołującej się do doświadczenia wcale nie musi być prezentacja dobrze uporządkowana, postulatywna. Czytelnik charakteryzuje się pewną „giętkością poznawczą", gotów jest sam skonstruować odpowiedni do danej sytuacji schemat, korzystając z fragmentarycznej wiedzy, jaka płynie z nowego doświadczenia.

„Jeżeli chcemy, by prezentowana wiedza była na tyle elastyczna poznawczo, by można było z niej korzystać w przyszłości, najlepiej jest w nauce własnej i nauczaniu innych stosować metody prezentacji oparte na przypadkach, dany zakres treści traktując jak krajobraz, który poznaje się «przemierzając» go wielokrotnie wzdłuż i wszerz i zaglądając ponownie w każdy «kąt» (każdy przypadek), za każdym razem w kontekście innych przypadków sąsiednich, oraz porównując poszczególne przypadki przy użyciu wielu różnych abstrakcyjnych wymiarów" (Spiro i współaut., 1987, s. 178).

Nadal jednak trudno zrozumieć, na czym polega przekazywanie wiedzy. Jeszcze trudniej pojąć, w jaki sposób drobny aspekt danego przypadku, nawet nietypowego, może, jak twierdzi wielu czytelników, zmodyfikować dotychczasowy sposób rozumienia przypadków w ogóle. Przygotowując studium przypadku pewnej szkoły położonej na terenie getta, obserwowałem pracę jednego z nauczycieli. Nauczyciel ten dla utrzymania dyscypliny wprowadził w klasie pewien zestaw bezwzględnie obowiązujących zasad. Równocześnie zaś musiał wynajdywać coraz to inne, bardziej liberalne zasady względem jednego z uczniów,

chłopaka nieposkromionego, któremu groziło wydalenie ze szkoły. Po przeczytaniu mojego opisu, nauczyciele z bardzo różnych szkół poparli oba podejścia. „Tak, zasad należy ściśle przestrzegać". „Tak, czasami trzeba zasady naginać." Każdy z tych nauczycieli dostrzegł w moim opisie wyjątek i uznał, że można ten wyjątek uogólnić. Przypadek, choć nietypowy, pozwolił im przecież wejrzeć w kondycję ludzką. Niekiedy ten wgląd akceptowany jest zbyt pochopnie. Autor studium przypadku musi zarówno uwiarygodnić swoje obserwacje, jak i uzasadnić trafność formułowanych na ich podstawie generalizacji.

Triangulacja

Przy „źle ustrukturowanej" prezentacji przypadków i „źle ustrukturowanej" recepcji, w klimacie, jaki stwarza konstruktywizm, nie należy się dziwić tolerowaniu niejednoznaczności i orędowaniu za pluralizmem. Mimo to większość badaczy zabiega o trafność swoich komunikatów. Przekazywane znaczenia nigdy nie docierają do odbiorcy w nienaruszonym stanie, zawsze do pewnego stopnia odciska się na nich indywidualny sposób rozumowania odbiorcy, mimo to oczekuje się pewnej korespondencji między znaczeniami przypisywanymi sytuacji, obserwacją, komunikatem i recepcją. Joseph Maxwell (1992) apeluje o odrębne traktowanie trafności opisu, interpretacji, teorii, uogólnień i ocen.

Istnieje szereg procedur zmniejszających prawdopodobieństwo błędnej interpretacji, np. wielokrotne gromadzenie danych, stosowanie procedur sprawdzających proponowane wcześniej interpretacje itp. (Denzin, 1978; Goetz, LeCompte, 1984). W jakościowych badaniach studium przypadku procedury te określa się ogólnie mianem *triangulacji*. Triangulacja, mówiąc ogólnie, polega na rozpatrywaniu badanego obiektu z wielu punktów widzenia, aby móc go lepiej zrozumieć, i weryfikowaniu powtarzalności danej obserwacji czy interpretacji. Zważywszy jednak, że replika obserwacji czy interpretacji nigdy nie jest doskonała, triangulacja służy również precyzowywaniu rozumienia przez wskazanie różnych sposobów patrzenia na dane zjawisko (Flick, 1992).

Porównywanie

Badacze wiedzą, że opisywane przez nich przypadki będą porównywane z innymi. W różnym jednak stopniu czują się odpowiedzialni za dokonywane porównania, za dostarczenie odbiorcy właściwych odniesień, zdając sobie sprawę, że dla każdego czytelnika punktem od-

niesienia może być inny przypadek. Większość badaczy stosujących podejście sytuacyjne, etnograficzne lub fenomenologiczne stara się badany przez siebie przypadek przedstawić na tyle szczegółowo, by czytelnik mógł dokonać stosownych porównań. Czasem badacz sam podsuwa ewentualne pomysły na ten temat. Badacze ujmujący swoje przypadki w kategoriach ilościowych zwykle sami dokonują porównań, omawiając jeden lub więcej przypadków stanowiących punkty odniesienia lub podając statystyczne wyznaczniki grup odniesienia, które mógłby reprezentować hipotetyczny przypadek porównawczy. Ale mimo tradycji komparatystycznych w antropologii porównawczej i dziedzinach pokrewnych — ani podejście ilościowe, ani jakościowe nie zapewnia porównań dokładnych i wyczerpujących (Firestone, w druku; Ragin, 1987; Tobin, 1989).

Ja sam przypisuję porównywaniu odrębną funkcję epistemologiczną, równie ważną jak gromadzenie wiedzy o przypadku i zdobywanie wiedzy na jego podstawie. Porównywanie jest mechanizmem skutecznym w dochodzeniu do sedna sprawy, pozwala bowiem skupić uwagę na kilku porównywanych właściwościach przypadku z pominięciem pozostałych. Opis porównawczy jest przeciwieństwem czegoś, co Clifford Geertz (1973) nazywa opisem rozdętym (*thick description*). Rozdęty opis programu kształcenia muzycznego uwzględniałby np. skład personalny prowadzących, najnowsze zmiany programowe, charyzmę dyrygenta chóru, kontakty robocze z organistą kościelnym, wynik głosowania rady szkoły i ubolewania z powodu braku zainteresowania uczniów nauką gry na klarnecie. Oczywiście każdy z tych punktów z osobna może być obiektem porównań, ale opis porównawczy zwykle dotyczy zmiennych bardziej ogólnych, tradycyjnie uwzględnianych w programach kształcenia muzycznego — np. spraw kadrowych, kosztorysu, koncepcji i organizacji występów gościnnych. Badacz koncentruje się przede wszystkim na tych elementach, które stanowią podstawę porównań, a elementów tych jest niewiele. Interpretuje elementy niepowtarzalne i kompleksowe. W porównawczym schemacie badawczym przedmiotem zainteresowania badacza nie jest *sam przypadek*, lecz *porównywanie* przypadków.

Bez względu na to, z jakim typem przypadku mamy do czynienia — autotelicznym, instrumentalnym czy zbiorowym — przypadki wskazane przez badacza jako materiał porównawczy niezbyt wzbogacają wiedzę odbiorcy. Przy większej liczbie przypadków interesujących samych w sobie porównywanie ich może być, rzecz jasna, pożyteczne, zazwyczaj jednak z punktu widzenia badacza interesujący sam w sobie

jest co najwyżej jeden przypadek. Czytelnik zainteresowany danym przypadkiem jako takim więcej się o nim dowie studiując bezpośrednio jego opis, a materiał porównawczy traktując tylko jako dodatkowe uzupełnienie. Czytelnik zapoznający się z instrumentalnym studium przypadku poznaje konkretny przejaw szerszego zjawiska. Badacza rzadko interesują te różnice między porównywalnymi przypadkami, które wykraczają poza analizowane zjawisko, ponieważ potencjalnych różnic jest tu zbyt wiele[10] . Uogólnienia formułowane na podstawie analizy różnic między dwoma dowolnymi przypadkami są bardziej wątpliwe niż uogólnienia oparte na analizie jednego przypadku. Pokazanie, jak dane zjawisko konkretyzuje się w ściśle określonym przypadku, może być źródłem cennych i wiarygodnych informacji.

Jest wiele sposobów konceptualizacji przypadku pomocnych w uzyskaniu maksimum informacji. Przyjmuje się zwykle, że przypadek, to coś, co funkcjonuje, co działa; studiując przypadek, obserwujemy jak działa. Jest w nim coś, co należy opisać i zinterpretować. Większość naturalistycznych, holistycznych, etnograficznych i fenomenologicznych studiów przypadku kładzie nacisk na oddzielenie obiektywnego opisu od subiektywnej interpretacji, konieczność szanowania kulturowo uwarunkowanych różnic w percepcji różnych zjawisk i otwartości na te różnice oraz konieczność przedstawiania warunków lokalnych w sposób empatyczny. Wszystko to razem wzięte (a może raczej wrzucone do jednego worka) składa się na epistemologię konstruktywistyczną.

Metody

Najprostszą chyba receptą na metodę jakościowego studium przypadku jest umieszczenie najtęższych głów w centrum zdarzeń. Na pierwszy rzut oka podstawowym zadaniem badacza jest obserwacja, ale tak naprawdę jest nim refleksja. (Ja sam wolałbym tu mówić o *interpretacji*, by podkreślić, że chodzi o animowanie rozumienia, ale etnografowie, mówiąc o interpretacji, mają na myśli „poznawanie specyficznych zapatrywań aktorów*, znaczeń lokalnych"; por. Erickson, 1986). Badacz

[10]Studia ewaluacyjne polegające na porównywaniu programu innowacyjnego z kontrolnym rzadko bywają wiarygodne. Bez względu na to, jak rzetelnie prowadzimy badanie, przypadek kontrolny niedostatecznie reprezentuje znane już czytelnikowi przypadki. Poprzez wszechstronny opis ocenianego programu badacz powinien pomóc czytelnikowi w formułowaniu wyważonych uogólnień.

*Termin, który zyskał już prawo obywatelstwa w polskiej socjologii, oznacza osobę działającą, uczestnika zdarzeń, kogoś czynnego w danych warunkach. Aktorem w tym rozumieniu jest człowiek w sytuacji społecznej — *przyp. red.*

zachowujący we wszelkich okolicznościach postawę refleksyjną stara się przede wszystkim przemyśleć własne impresje, to, co zapamiętał i zanotował. Nie musi przy tym zgadzać się ani z teoretykami, ani z aktorami, ani z odbiorcami swoich przemyśleń (Carr, Kemmis, 1986). Ważne są znaczenia lokalne; ważne są znaczenia projektowane (*foreshadowed meanings*); ważne są znaczenia, jakich doszuka się następnie czytelnik. Autor studium przypadku boryka się ze wszystkimi trzema rodzajami znaczeń i znajduje racje, dla których zajmuje sie jednymi a nie innymi. Zawsze wchodzi tu w grę refleksja[11].

W typowym jakościowym studium przypadku dane (czasem wcześniej zakodowane) bezustannie podlegają interpretacji — już przy pierwszym zetknięciu, a potem jeszcze raz i jeszcze raz. Zapisy i tabele czyta się starannie nie tylko po to, by móc je skategoryzować i dostrzec w nich prawidłowości, ale także po to, by poddać je „krzyżowej": refleksji (Spiro i współaut., 1987). Charakterystyczną cechą jakościowego studium przypadku jest to, że osoba kierująca badaniem spędza osobiście wiele czasu w terenie, stykając się bezpośrednio z przypadkiem, sposobem i zakresem jego działania, zastanawiając się nad nim, odczytując wciąż na nowo sens tego, co się dzieje.

Badacz przypadku — socjolog, etnograf — stara sie wyszukiwać to, co istotne w obserwowanych zdarzeniach, okolicznościach, przejawach wartości. O tym, czego sam dostrzec nie może, dowiaduje się z rozmów ze świadkami wydarzeń lub z zapisów tych wydarzeń.

Przeglądając literaturę zauważyłem, że o metodach studium przypadku piszą przede wszystkim ludzie przekonani, że celem tych badań są naukowe uogólnienia. Autorami większości studiów przypadku są natomiast ludzie, których przypadki te interesują *same w sobie*. Każdy z tych badaczy stara się zaplanować swoje studium przypadku tak, by zrozumieć istotę tego konkretnego przypadku z jego własnej perspektywy, a nie z perspektywy badacza i teoretyka, by wydobyć z niego jego własne problemy, konteksty, interpretacje. Inne jest zadanie metod instrumentalnych studiów przypadku, mają one pomóc badaczowi pokazać, jak w danym przypadku wygląda to, co interesuje badacza i teoretyka. Ponieważ tu jest bardziej prawdopodobne, że badacz już w punkcie wyjścia wie, na co powinien zwrócić uwagę (określa to mię-

[11]Mówiąc o *refleksyjności*, etnografowie czasem pragną zwrócić uwagę jedynie na potrzebę podawania w wątpliwość trafności swoich uniwersalnych (*etic*) problemów badawczych, swojego układu odniesienia, swoich nastawień (*bias*) kulturowych. To podawanie w wątpliwość jest ważne, ale mówiąc tutaj o refleksyjności analizy jakościowej przypadku mam na myśli, podobnie jak Donald Schön (1983), ogólne nastawienie mentalne badacza.

dzy innymi reprezentowana przez niego dyscyplina badawcza), schemat badawczy w większym stopniu opiera się na gotowych narzędziach i wcześniej ustalonych zasadach kodowania.

Autorzy autotelicznych studiów przypadku nie unikają uogólnień — byłoby to niemożliwe. Z całą pewnością formułują generalizacje na temat tego, co się może zdarzyć badanemu przypadkowi w przyszłości lub w innych sytuacjach. Od czytelnika oczekują zarówno zrozumienia proponowanych im interpretacji, jak i przemyśleń własnych. Stosują więc takie metody, które pozwolą na tyle poznać dany przypadek, by można było zawrzeć w końcowym raporcie całą złożoność jego znaczeń, a zarazem charakterystykę przypadku przedstawić w formie na tyle opisowej, by czytelnik mógł zrelacjonowane zdarzenia przeżyć zastępczo i wyciągnąć z nich własne wnioski.

Dobór przypadku

Tym, co najbardziej chyba wyróżnia studium przypadku spośród innych metod stosowanych w naukach i służbach społecznych, jest sposób doboru przypadków do badań. Przypadki badane dla ich walorów autotelicznych z reguły narzucone są z góry. Lekarz, pracownik socjalny, ewaluator programu szkolnego nie wybiera swoich przypadków — zostają mu one narzucone. Przypadek jest przedmiotem zainteresowania przed rozpoczęciem formalnego badania. Przypadki analizowane z racji instrumentalnych i zbiorowe studia przypadku z reguły muszą być specjalnie dobierane. Właściwy dobór jest często warunkiem zrozumienia zjawiska, o które chodzi (Patton, 1990; Yin, 1989). Załóżmy, że chcemy zrozumieć zachowanie porywaczy i decydujemy się wysondować to zjawisko posługując się studium przypadku. Porwania nie są zjawiskiem częstym — jest ich kilka w skali całego świata. Przypuśćmy, że w danym momencie mamy do wyboru: sprawcę napadu na bank, porywacza samolotu, odsuniętego od rodziny ojca porywacza własnego dziecka oraz grupę islamskich szyitów. Chcielibyśmy dojść do jakichś uogólnień na temat populacji porywaczy, zdajemy sobie jednak sprawę, że każdy z tych przypadków (prób) jest mało *reprezentatywny* dla całej interesującej nas grupy.

Jeżeli planujemy badanie zgodnie z zaleceniami Milesa i Hubermana (1984), właściwa selekcja przypadków jest sprawą absolutnie najważniejszą. Jest to problem właściwego doboru próby. Wyselekcjonowane przypadki mają być reprezentatywne dla określonej populacji przypadków. Przejawy interesującego nas zjawiska w konkretnym, poddanym obserwacji przypadku powinny być reprezentatywne dla tego zjawiska

w ogóle. Zdaniem Milesa i Hubermana, Yina, Malinowskiego, studium przypadku ma służyć przede wszystkim nauce. Chodzi przede wszystkim o to, by jak najlepiej wyjaśnić interesujące nas zjawisko (von Wright, 1971). W punkcie startu mamy po prostu zjawisko, dopiero analiza konkretnych przypadków stwarza szansę jego zbadania.

Mamy się zająć porywaniem zakładników. Chcemy to zjawisko lepiej zrozumieć, ogarnąć z perspektywy dotychczasowej wiedzy na temat przestępczości, rozwiązywania konfliktów, stosunków międzyludzkich — różnych „abstrakcyjnych wymiarów"[12]. Zdajemy sobie sprawę z istnienia dużej populacji przypadków hipotetycznych i mniejszych subpopulacji przypadków dostępnych. Interesuje nas przede wszystkim uogólniona wiedza na temat zjawiska, jakim jest porywanie zakładników, a nie któryś z dostępnych przypadków z osobna wziętych. Z punktu widzenia reprezentatywności wartość poznawcza tych przypadków jest niewielka, ale jesteśmy optymistami: wierzymy, że każdy niemal przypadek może nam coś ważnego ujawnić. Wybieramy jeden lub niewielką ich liczbę. Zazwyczaj zakładnikami są przypadkowi ludzie, którzy akurat byli pod ręką. Możemy zatem wykluczyć ojca, który porwał własne dziecko. Być może ten akurat typ porwań zdarza się częściej niż inne, z ojca jednak rezygnujemy. Bardziej nas interesują te przypadki porwań, które noszą znamiona przestępstwa, porwanie w celu wymuszenia azylu. Badacz bierze pod uwagę różne interesujące aspekty danego zjawiska, wybierając przypadek z jakichś względów typowy, ale przede wszystkim skłania się ku tym przypadkom, które stwarzają szansę na *dowiedzenie się czegoś nowego*. Ja sam dokonałbym takiego właśnie wyboru[13]. Być może po rozważeniu wszystkich za i przeciw wybrałbym w końcu przypadek, z którym mógłbym spędzić jak najwięcej czasu. Potencjał poznawczy to nie to samo, co reprezentatywność, ale czasem on właśnie jako kryterium jest ważniejszy. Często lepiej zbadać przypadek nietypowy, z którego można się wiele dowiedzieć, niż doskonale typowy, który niewiele wzbogaca naszą wiedzę.

Jeszcze jeden przykład: Przypuśćmy, że interesuje nas atrakcyjność wystaw uczestniczących organizowanych w muzeach dla dzieci. (Wy-

[12]Jak już wspomniałem wcześniej, ja sam nazywam je problemami (*issues*) lub obszarami problemowymi (*issue areas*). Mary Kennedy (1979) nazywa je „istotnymi właściwościami" (*relevant attributes*). Spiro i współaut. (1987) używają terminu „abstrakcyjne wymiary" (*abstract dimensions*). Malinowski (1987) nazywa je „teoriami". W wypadku naszych badań nie chodzi tu o wielkie teorie funkcjonujące w ramach takiej czy innej dyscypliny naukowej, lecz o teorie robocze.

[13]Nacisk położony tu został na możliwość dowiedzenia się jak najwięcej zarówno o danym przypadku, jak i o szerszym zjawisku, a zwłaszcza o tym ostatnim, jeżeli wyjątkowość okoliczności stwarza możliwość niecodziennego wglądu w badany problem.

stawa uczestnicząca to taka wystawa, gdzie zwiedzający może brać eksponaty do ręki, prosić o dodatkowe wyjaśnienia itp.). Środki pozwolą nam uwzględnić cztery muzea, przeprowadzić zbiorowe studium czterech przypadków. Zaczęlibyśmy prawdopodobnie od typologii muzeów uwzględniającej: (a) charakter muzeum — sztuk pięknych, nauki i techniki, historyczne; (b) charakter miasta — duże, bardzo duże; (c) charakter wystawiennictwa — wystawa tradycyjna, ekspozycja uczestnicząca. W wyniku otrzymujemy dwunastopolową macierz. Zapewne nie uda się każdego pola wypełnić odpowiednim przykładem, zresztą i tak nie starczyłoby środków na 12 muzeów. Wiedząc, że możemy uwzględnić 4 muzea, decydujemy się początkowo na jedno muzeum sztuk pięknych, jedno historyczne i dwa muzea nauki i techniki (ponieważ w tym ostatnim typie placówek łatwiej trafić na wystawę uczestniczącą); dwie placówki w dużym mieście i dwie w mieście bardzo dużym; dwa muzea organizujące wystawy tradycyjne i dwa — ekspozycje uczestniczące. Ale po bliższym przyjrzeniu się istniejącym przypadkom, rozważeniu problemów organizacyjnych, posiadanych środków i kilku istotnych czynników dodatkowych wybieramy ostatecznie cztery muzea, w sumie co prawda nie w pełni reprezentatywne, ale za to zróżnicowane pod względem charakterystyk, te cztery, które stwarzają najlepszą szansę dowiedzenia się czegoś o wystawach uczestniczących[14]. Żadne cztery muzea, najlepiej nawet dobrane ze względu na zamierzenie badawcze, nie będą reprezentatywną próbą wszystkich muzeów, a już z całą pewnością nie dadzą podstaw statystycznych do formułowania uogólnień w kwestii zależności między charakterem ekspozycji a formalnymi cechami muzeów. Zazwyczaj z jakichś cech musimy zrezygnować, choć nie są bez znaczenia. Nawet wtedy, gdy zamierzamy przeprowadzić zbiorowe studium przypadku, poszczególnych własności analizowanych obiektów nie należy traktować jako najważniejszego kryterium doboru próby. Równoważność zmiennych i ich różnorodność to na pewno sprawy ważne; najważniejsza jednak jest możliwość dowiedzenia się czegoś nowego.

Dobór próby w ramach przypadku

Praca nad autotelicznym studium przypadku również zakłada pewną selekcję. Mimo że przypadek jest nam (zwykle) z góry narzucony, musimy podjąć szereg dalszych decyzji na temat osób, miejsc

[14] Firestone (w druku) zaleca maksymalizowanie zróżnicowania i dbałość o „możliwie jak największe podobieństwo do interesującej nas populacji".

i zdarzeń, które będziemy obserwować. I tym razem doświadczenie zawodowe i intuicja podpowiadają, że powinniśmy szukać odpowiedniej próby. Przypuśćmy, że przedmiotem naszej analizy jest program zaopatrywania czwartoklasistów w komputery domowe służące jako pomoc w nauce[15]. Przypadki (szkoły) zostały wybrane z góry. O niektórych czynnościach badawczych decyduje koordynator badań, ale każdy podwykonawca odpowiedzialny jest za przeprowadzenie jednego studium przypadku. Jedną z najważniejszych spraw wymagających rozważenia jest wpływ programu na rodzinę: umieszczenie komputera w domu wiąże się z pewnymi oczekiwaniami co do możliwości jego użytkowania. (Rodzina powinna mieć dostęp do komputera i móc używać go do redagowania tekstów, prowadzenia dokumentacji i do zabawy, ale czwartoklasista musi mieć zarezerwowany czas na odrabianie lekcji). W jednym okręgu badawczym 50 rodzin ma już własne komputery. Każda z tych rodzin może udzielić badaczowi pewnych informacji, ale bezpośrednią obserwację można prowadzić tylko w nielicznych domach. Które domy należy wybrać? Podobnie jak przy zbiorowym studium przypadku i tym razem badacz sporządza najpierw listę zmiennych, które go interesują: płeć czwartoklasisty, obecność rodzeństwa, struktura rodziny, poziom dyscyplinowania życia domowego, wcześniejszy kontakt rodziny z komputerami i innymi urządzeniami elektronicznymi, itp. Omawia następnie te punkty ze swoimi informatorami, zbiera rekomendacje, odwiedza kilka domów, zbiera potrzebne dane. Wreszcie wybiera swoją próbę[16] (maksymalnie zróżnicowaną, choć niekoniecznie reprezentatywną), nie zważając zbytnio na typowość, za to kierując się w dużej mierze takimi kryteriami, jak dostępność, a nawet gościnność, ponieważ czasu jest niewiele, a wrogo nastawieni rodzice nie udzieliliby zapewne zbyt wielu informacji. I tym razem więc podstawowym kryterium jest szansa dowiedzenia się czegoś o przypadku.

Kwestie etyczne

Autorzy studiów przypadku często zajmują się sprawami, które żywo interesują opinię publiczną, ale o których ani opinia publiczna, ani uczony „nie mają prawa wiedzieć". Ani przydział funduszy badawczych, ani motywacja poznawcza, ani dyplom socjologa nie uprawnia

[15]Tak było z programem Buddy (Buddy Project), jednym ze składników podjętej w latach 1990–1993 próby reformy oświaty publicznej w stanie Indiana (por. Quinn, Quinn, 1992).

[16]Problem sukcesywnego doboru prób w dłuższym odcinku czasu omawiają Patton (1990), Strauss i Corbin (1990) i Firestone (w druku).

automatycznie do wkraczania w czyjś prywatny świat. Korzyści płynące ze znakomicie przeprowadzonego badania rzadko kiedy równoważą krzywdę wyrządzoną człowiekowi, którego tajemnice wywleczone zostały na światło dzienne. Każdy badacz stosujący jakościowe metody naukowe jest gościem w prywatnych przestrzeniach tego świata. Powinien mieć dobre maniery i bezwzględne zasady moralne.

Studium przypadku, podobnie jak większość metod jakościowych, charakteryzuje się żywym zainteresowaniem dla poglądów i okoliczności życia osób badanych. Osoby te narażone są więc na zdemaskowanie i wstyd: na utratę szacunku, pracy, dobrej samooceny. Należy zawczasu przedyskutować problemy, które mają być obiektem obserwacji i opisu. Należy z osobami badanymi uzgodnić, kto może mieć dostęp do informacji, a kto nie. Ustaleń tych należy ściśle przestrzegać. Osoby badane powinny mieć możliwość zapoznania się ze wstępnymi wersjami studium, by móc sprawdzić, jak je tam opisano, które wypowiedzi zostały zacytowane, jak materiał zinterpretowano, badacz z kolei powinien uważnie wysłuchać wszystkich zgłoszonych zastrzeżeń. Jest to zalecenie ważne choć niewystarczające. Należy przedsięwziąć wszelkie możliwe kroki zmniejszające ryzyko i ściśle przestrzegać zasad ochrony osób badanych. Co więcej, badacz powinien sam te zasady zaostrzyć, unikając drążenia kwestii drażliwych, które niewiele wnoszą do tematu, i prosząc osoby trzecie o nadzór nad systemem zabezpieczeń.

Podsumowanie

Podstawowe zadania autora jakościowego studium przypadku są zatem następujące:

1. Określenie granic przypadku, konceptualizacja przedmiotu badań.
2. Wybór zjawisk, tematów, problemów (pytań badawczych).
3. Szukanie wyłaniających się ze zgromadzonych danych prawidłowości, by móc doprecyzować problemy badawcze.
4. Triangulacja (rozpatrywanie z wielu różnych punktów widzenia) kluczowych obserwacji i przesłanek interpretacji.
5. Wybór i testowanie alternatywnych interpretacji.
6. Formułowanie twierdzeń lub uogólnień na temat badanego przypadku.

Z wyjątkiem punktu pierwszego, są to zadania wspólne dla wszystkich metod jakościowych. Im bardziej badacz interesuje się danym przypadkiem samym w sobie, tym bardziej będzie się koncentrował na tym, co w danym przypadku wyjątkowe, na jego specyficznym kontek-

ście, specyficznych problemach, historii. Badacz pracujący nad studium przypadku musi podjąć szereg decyzji formalnych. Oto niektóre z nich:

1. Jak dalece raport końcowy powinien mieć formę narracji?

2. Jak dalece należy badany przypadek porównywać z innymi przypadkami?

3. Jak dalece uogólniać wiedzę o przypadku na inne przypadki i czy nie pozostawić procesu dochodzenia do wniosków ogólnych czytelnikowi?

4. W jakim stopniu uwzględniać w raporcie opis badacza jako uczestnika?

5. Czy utrudniać identyfikację przypadku i w jakim stopniu?

Studium przypadku jest metodą naukową (jedną z wielu), ale służy nie tylko nauce. Jeden lub kilka przypadków to liczba zbyt mała, by mogła być reprezentatywna dla całej populacji przypadków, nie dająca podstaw do wielkich uogólnień, natomiast przypadek będący wyjątkiem od reguły może ukazać granice takich uogólnień. Z mniejszym przekonaniem będziemy twierdzić, że zawsze lepiej jest, by dziecko rozwiedzionych rodziców zostało przy matce, jeśli poznamy choć jeden przypadek wskazujący na szkodliwe konsekwencje takiego rozwiązania. Studia przypadku pomagają precyzować teorie, są źródłem pomysłów do dalszych badań i pomagają określić granice dopuszczalnej generalizacji.

Studia przypadku pomagają też kształtować opinię publiczną i zmuszają do refleksji nad ludzkim życiem. Zapoznawanie się z cudzymi doświadczeniami pełni ważną rolę w kształtowaniu preferencji i oczekiwań. Epistemologia formalna wciąż nie jest jeszcze nauką doskonałą, mimo to ludziom udaje się jakoś z opisów pojedynczych przypadków wyciągać wnioski na użytek innych przypadków. Nie zawsze są to wnioski trafne, ale bywają przekonujące nawet przy rozbieżności poglądów.

Celem studium przypadku jest nie poznanie świata, lecz poznanie konkretnego przypadku. Badania, których celem jest trafna generalizacja, muszą spełniać określone kryteria. Kryteria te nie pasują do badań, których podstawowym celem jest trafny opis tego, co szczególne, trzeba je zatem modyfikować. Studia przypadku dlatego są tak użyteczne dla praktyków i polityków, że uzupełniają doświadczenia osobiste. Metody jakościowego studium przypadku to w gruncie rzeczy nic innego jak metody dyscyplinowania doświadczeń osobistych i jednostkowych.

tłumaczyła *Helena Grzegołowska*

Rozdział IV

Projektowanie i metody badawcze

Colin Robson

Projektowanie ewaluacji

Colin Robson, „Designing Evaluations", rozdział z książki tegoż autora *Real World Research. A Resource for Social Scientists and Practitioner-Researchers*, Blackwell Publishers, Oxford — Cambridge Mass. 1995.

Ewaluacja — cel, a nie strategia

Ewaluacja nie jest jakąś nową czy też inną strategią badawczą, ewaluacja to badanie, które służy realizacji pewnego specjalnego celu. Celem tym jest ocena rezultatów i efektywności — zwykle innowacji lub interwencji — jakiegoś działania o charakterze politycznym, praktycznym lub usługowym. Ewaluację przeprowadza się przy użyciu różnych strategii badawczych: eksperymentu, sondażu, studium przypadku lub metodami kombinowanymi. [...]

Ewaluacja jest faktycznie rodzajem działania bardzo wrażliwym na reakcje społeczne, wiąże się z nim ryzyko (obowiązek?) ujawniania nieadekwatności i błędów, twoje intencje jako ewaluatora mogą być źle rozumiane, a twoje ustalenia zignorowane lub wykorzystane w niewłaściwy sposób. Projektowanie ewaluacji wymaga od ciebie starannego przemyślenia, co robisz i dlaczego.

Znaczenie ewaluacji

Rozliczanie (*accountability*) jest obecnie kluczowym pojęciem całego spektrum oferowanych usług publicznych takich, jak edukacja, zdrowie i opieka społeczna. Bierze się to częściowo z przyczyn politycznych i ideologicznych, nasila się bowiem (w Wielkiej Brytanii, w innych częściach Europy, a zwłaszcza w Stanach Zjednoczonych) tendencja do

sytuowania usług publicznych w strukturach działających według tych samych zasad, według których działa sektor prywatnej gospodarki nastawionej na zysk. Istnieją także, czy może powinny istnieć, bardziej „społeczne" bodźce nakłaniające do przeprowadzania ewaluacji. Na przykład w kwestii informowania o AIDS Freudenberg (1990) twierdzi, że „od początku epidemii było jasne, iż naszym najpotężniejszym narzędziem zapobiegania rozprzestrzenianiu się zakażenia wirusem HIV jest upowszechnianie wiedzy na temat AIDS. Niestety, po dziesięciu latach więcej wiemy na temat biochemii HIV i udziału komórek T w organizmie w różnych stadiach choroby niż o tym, od czego zależy efektywność programów edukacji na temat AIDS i jak takie programy wprowadzać w życie, by osiągnąć zamierzony skutek" (s. 295).

Podobne uwagi sformułowano odnośnie do programów oszczędzania energii (Kushler, 1989).

Niezależnie od tego, w czym tkwią przyczyny — poszukiwana przez nas strategia zrozumienia i analitycznej oceny funkcjonowania usług oraz programów ulepszeń w tej dziedzinie stoi przed wielkim wyzwaniem. Sporne kwestie wiążą się z wykonawcą, ze sposobem działania i jego celem. Sytuuje nas to właśnie w szybko rozwijającej się dziedzinie badań ewaluacyjnych.

De facto większość badań nad otaczającą nas rzeczywistością prowadzonych przez psychologów i innych przedstawicieli nauk społecznych może być traktowana jako pewien rodzaj ewaluacji: *jako próba oceny wartości innowacji lub interwencji, usługi lub podejścia.* Jest to oczywiste w dziedzinie edukacji, medycyny lub badań rynkowych, ale wymiar ewaluacyjny ma również spora część badań stosowanych, nielaboratoryjnych.

Ewaluacja jest rodzajem badań stosowanych. Ogólnie rzecz biorąc, za badania stosowane uważa się takie, które definiują problemy konkretnej rzeczywistości bądź poszukują takich strategii, sposobów działania lub programów, które pomogłyby te problemy rozwiązać. Ewaluacja jest przede wszystkim postrzegana jako opisywanie i określanie efektów danego podejścia, polityki lub programu. Począwszy od lat sześćdziesiątych obserwujemy szybki rozwój badań ewaluacyjnych. Rozwój ten wspomógł rząd Stanów Zjednoczonych, który pewną część budżetu wielu rozpoczynanych wtedy programów społecznych przeznaczył na badania ewaluacyjne. Okazało się jednak, że w odniesieniu do programów o szerokim zasięgu ich konkluzje są dość wątpliwe, że badania te zwykle nie osiągają założonych celów (por. Mullen i współaut., 1972). Zaczęto szeroko krytykować jakość wielu ewaluacji (por. Weiss, 1977).

Jednakże dyskusja nad problemami, jakie nastręczały ewaluacje na dużą skalę (por. Cronbach, 1982), okazała się bardzo cenna dla organizacji badań w mniejszej skali.

Charakterystyczne cechy badań stosowanych ponad wszelką wątpliwość odnoszą się również do ewaluacji. Na przykład badania ewaluacyjne zamawiane są zazwyczaj przez klienta bądź sponsora, którego interesy bezpośrednio wiążą się z przedmiotem ewaluacji. W tej sytuacji o kwestiach, którymi zajmować się będzie ewaluator, decyduje nie on sam, lecz ktoś inny — zwykle jednak sprawa ta staje się przedmiotem negocjacji między wykonawcą ewaluacji a klientem. Ma to nawet swą pozytywną stronę: badania ewaluacyjne mają szansę bardziej oddziaływać na rzeczywistość (lub przynajmniej na jej wycinek, jakim jest oceniany program czy innowacja) niż badania tradycyjne. Pojawia się tu jednak wiele kwestii etycznych. Czyim interesom służy interwencja? Kto jest rzeczywistym klientem (czy jest nim zamawiający badanie czy też ci, którzy mają zyskać w efekcie danej interwencji)? Jak splecione są ze sobą różne wchodzące w grę interesy? Ewaluacja, jej rezultaty i sposób, w jaki są prezentowane, mogą mieć wymierne skutki dla pracy, wykształcenia i stanu zdrowia wielu osób. Równie trudno uchronić się przed kwestiami politycznymi. Zarówno wybór rodzaju i stylu ewaluacji, jak przyjęte kryteria mogą być równoznaczne z wyborem perspektywy, wartości i celów którejś z zainteresowanych stron.

Ewaluacje dotykają także kwestii związanych ze zmianą. Usługa, program bądź inna oceniana praktyka mogą chcieć spowodować zmianę lub do niej zachęcać. Ewaluacja może wykazać, że program, jeśli ma być efektywny, wymaga zmian. Ustalenia poczynione w trakcie ewaluacji są jednakże tylko jedną składową złożonego układu wpływów, którym poddany jest przebieg programu. Ewaluatorzy powinni informować o rezultatach swej pracy i płynących stąd wnioskach nie tyle swoich kolegów badaczy, którzy zapewne są wystarczająco dobrze poinformowani i przychylnie nastawieni do badań empirycznych, ile klientów i osoby podejmujące decyzje. Oznacza to, iż trzeba zastanowić się nad sposobem komunikowania się oraz stylem i charakterem raportów z ewaluacji.

Badania każdej konkretnej rzeczywistości przysparzają także wielu problemów praktycznych. Zwykle ewaluacja ma ściśle wyznaczony, niezbyt odległy, termin jej zakończenia. Dostęp do jej uczestników może być utrudniony zarówno z uwagi na ich zajęcia i obowiązki, jak i z uwagi na ich niechęć do uzewnętrzniania swoich spraw. Trudności mogą robić „strażnicy firmy", urzędnicy średniego szczebla zarządzania

albo zastępcy dyrektorów szkół. Zmienić się też może w trakcie badania w sposób istotny system lub otoczenie. Źródłem zakłóceń stać się mogą czynniki zewnętrzne, od strajku począwszy, a na burzy śnieżnej skończywszy.

Uwaga na temat możliwych przedmiotów ewaluacji

Lista takich przedmiotów nie ma końca. My podajemy przykład z dziedziny edukacji (ramka 1). Ewaluować można, by przywołać kilka losowo wybranych przykładów, nową restaurację, eksperymentalny schemat komunikacji, kampanię reklamową, pakiet oprogramowania lub seminaria uniwersyteckie. Każdy, kto raz zajmie się ewaluacją, przekona się jak działa Kaplana „prawo młota" (dając komuś do ręki młot, zmieniasz jego otoczenie w zbiór rzeczy, które trzeba uderzyć młotem). Dalej mówić będziemy głównie o ewaluacji „programów". Przez określenie „program" należy rozumieć wszystko, co można poddać ewaluacji. Skupimy się zaś na ludziach, na „ludzkim" aspekcie ewaluacji. W niektórych przypadkach — na przykład, gdy mówimy o ewaluacji oprogramowania — trzeba wyjść poza ludzi i odnieść się do podstawowych kwestii technicznych.

Definiowanie ewaluacji

Twierdzimy tutaj, że *ewaluacja pod względem projektowania, technik gromadzenia danych oraz metod analizy nie różni się zasadniczo od innych przedsięwzięć badawczych* (co ma podkreślać używane niekiedy określenie „badania ewaluacyjne"). Nie wszyscy podzielają ten pogląd (por. Worthen, Sanders, 1973; MacDonald, 1974; Norris, 1990).

Ewaluację można realizować i faktycznie realizuje się przy użyciu różnych, czasem kombinowanych strategii badawczych. Twierdzi się niekiedy, że najlepsze efekty daje eksperyment, w toku którego porównuje się co najmniej dwie grupy (jedną objętą nowym programem lub usługą i drugą nie objętą). Niestety, zazwyczaj bardzo trudno albo znaleźć odpowiednią grupę kontrolną, albo uzyskać losowe rozmieszczenie różnych grup, albo wreszcie odizolować je od siebie na tyle skutecznie, aby uniknąć zniekształcenia wyników.

Istnieją wreszcie powody przemawiające za ewaluacją metodą studium przypadku. Po pierwsze, zapewnia ona większą elastyczność przy projektowaniu i wykonywaniu badania. Po drugie, celem wielu ewaluacji jest stwierdzenie, na ile efektywne i adekwatne są wprowadzone

Ramka 1. Przykład ewaluacji na małą skalę

Saunders przeprowadził badania ewaluacyjne zajęć organizowanych w czasie weekendu dla studentów mieszkających w akademikach — słuchaczy kursu A dwóch kolejnych lat. Zajęcia prowadził sam Saunders z pomocą kolegi. Zainspirowała ich napotkana w literaturze opinia, iż na zajęciach wstępnych ważniejsze są dla studentów sprawy ich tożsamości i motywacji niż przyswajanie sobie umiejętności studiowania.

Na zajęciach nie poruszano wobec tego zagadnień związanych z programem studiów, lecz zajęto się ustaleniem pewnych społecznych i osobowościowych uwarunkowań osiągania dobrych wyników w nauce.

Badanie miało charakter studium przypadku i gromadziło dane z wielu źródeł. Zwracano przy tym uwagę na mocne i słabe strony użytych metod, które obejmowały:

a) *Inwentarz umiejętności*: Studenci bezpośrednio przed i bezpośrednio po weekendzie wypełniali prosty, 18-punktowy test dotyczący ich predyspozycji osobowych i społecznych. Wyniki tych testów w obu grupach mieszkających w akademiku wykazały znaczący postęp, natomiast w grupie kontrolnej (studentów mieszkających w domach) — dużo mniejszy.

Komentarz: Posłużono się tu *quasi*-eksperymentem, w którym uczestniczyły dwie celowo dobrane grupy eksperymentalne oraz jedna grupa kontrolna. Można mieć wiele zastrzeżeń do wartości tego badania. Zwiększają ją bardzo zbliżone wyniki dla dwóch kolejnych lat.

b) *Profile studentów*: Studenci zostali poproszeni o zrelacjonowanie własnych doświadczeń na specjalnych kwestionariuszach zawierających pewną liczbę odpowiednio zatytułowanych rubryk odnoszących się do tego, co robili w czasie minionego weekendu. Materiał ten posłużył następnie do rozmów ze studentami jako punkt wyjścia ich refleksji nad sobą i samooceny.

Dla celów ewaluacji kursu Saunders poddał kwestionariusze formalnej analizie treści. Wszystkie stwierdzenia w poszczególnych rubrykach przyporządkowane zostały do odpowiednich kategorii, np. „przeżycia", „zachowanie", „relacje w grupie". Każde stwierdzenie oznaczone zostało następnie jako „pozytywne" lub „negatywne" względem pewnego zestawu kategorii nadrzędnych (np. „zmiana osobowa", „realizacja celów").

Większość wypowiedzi miała charakter pozytywny, a zatem profile pokazują, zdaniem Saundersa, że w czasie weekendu miał miejsce proces rzeczywistego uczenia się.

c) *Obserwację uczestniczącą*: Do głębszej analizy efektów kursu wykorzystane zostały obserwacje i wrażenia osoby go prowadzącej.

d) *Obserwację nieuczestniczącą*: Obserwację uczestniczącą uzupełniły obserwacje osoby z dużym doświadczeniem w dziedzinie edukacji psychologicznej i społecznej dokonane bezpośrednio przed i zaraz po zajęciach weekendowych.

Ocena ogólna wypadła pozytywnie. W podsumowaniu Saunders pisał: „Praktyczne podejście do mierzenia rezultatów może dostarczyć danych, które, niezależnie od innych walorów, mogą pomóc podjąć decyzję, czy warto przeznaczać zawsze ograniczone środki na podobne przedsięwzięcia".

Źródło: Saunders, 1990.

zmiany lub programy w danym specyficznym otoczeniu. Bardziej więc chodzi o konkretny przypadek niż o próbę.

Często celem ewaluacji jest nie tylko ocena wagi lub wartości zmian, ale także wsparcie procesu ulepszania tego, co jest poddawane ewaluacji. Michael Quinn Patton, wybitny amerykański ewaluator, piszący

zreszta teksty bardziej przystępne niż zazwyczaj spotykane w tej dziedzinie, stwierdził, że „ewaluacja jest procesem systematycznego gromadzenia informacji na temat działania, właściwości i rezultatów programów, personelu oraz produktów; informacje te są wykorzystywane przez specjalistów do redukowania niepewności i poprawienia efektywności, a także służą podejmowaniu decyzji z uwagi na to, co owe programy, personel lub produkty robią oraz czego dotyczą" (Patton, 1982, s. 15).

Definicja ta nie musi być powszechnie akceptowana, ale zwraca uwagę na:

— potrzebę *systematycznego* gromadzenia informacji;

— *szeroki wachlarz spraw* będących przedmiotem ewaluacji;

— fakt, że ewaluacja, jeśli ma być efektywna, musi być przez kogoś *spożytkowana*;

— *ogromną różnorodność celów* ewaluacji.

Ponadto pozwala ona dostrzec, że ewaluacja może być czymś więcej niż techniką umożliwiającą stwierdzenie, w jakim stopniu zrealizowane zostały założone cele. Definicja Suchmana (1967), który określa ewaluację jako „metodę oceniania stopnia, w jakim planowany program osiągnie pożądane cele", jest typowa dla tego podejścia i odzwierciedla wciąż powszechnie spotykany punkt widzenia. Będzie to zapewne i w przyszłości ważny aspekt ewaluacji programów mających wyraźnie określone cele, ale bez wątpienia jest on tylko wycinkiem tego, czym może zajmować się ewaluacja. Bardzo ważne choć niespodziewane mogą okazać się rezultaty i procesy nie zaplanowane. Przykładem jest szereg ewaluacji wykonanych przez Rossa (1973).

Różnorodność podejść

Istnieje wiele *modeli ewaluacji* — różnych koncepcji jej przeznaczenia oraz różnych zasad jej przeprowadzania. Ramka 2 zawiera kilka takich modeli. Każdy z modeli ma właściwe sobie wady i zalety. Podejście systemowe, na przykład, dostarcza łatwych do wykorzystania informacji o działaniu programu, zarazem jednak koncentruje się na samym jego przebiegu, zamiast pytać o podstawowe wartości. Podejście obliczone na „podjęcie decyzji" utożsamia ewaluację z wyborem jednego z alternatywnych rozwiązań, nie biorąc pod uwagę, że podejmujący decyzje nie muszą respektować ustaleń ewaluacji oraz że ewaluacja może służyć czemuś innemu niż tylko szukanie przesłanek konkretnych decyzji. Model „konfrontacyjny" może uwzględniać fakty pomijane przez inne

Ramka 2. Modele ewaluacji

1. *Analiza systemowa*: Pomiar ilościowy danych na wejściu i wyjściu, badanie efektywności i wydajności.
2. *Zoperacjonalizowane zadania*: Ustalenie, w jakim stopniu osiągane są mierzalne, z góry określone cele.
3. *Ewaluacja przyjmująca za podstawę potrzeby*: Bada się, na ile zaspokojone zostały potrzeby klienta.
4. *Ekspertyza*: Celem jest stwierdzenie, na ile program odpowiada standardom, które ewaluator-ekspert uznaje za doskonałe.
5. *Inspekcja kwalifikacji*: Eksperci z zewnątrz oceniają, w jakim stopniu program spełnia obowiązujące standardy profesjonalne.
6. *Model konfrontacyjny*: Dwa zespoły ewaluatorów prowadzą spór na temat zalet i wad programu oraz zastanawiają się, czy program powinien być kontynuowany.
7. *Model transakcyjny*: Koncentracja na przebiegu programu.
8. *Podejmowanie decyzji*: Ewaluacja jest zdeterminowana rodzajem decyzji, które należy podjąć.
9. *Analiza rozbieżności*: Porównuje się zakładane rezultaty programu z jego efektami faktycznymi.
10. *Ewaluacja wyjaśniająca*: Koncentruje się na metodach jakościowych, analizie indukcyjnej oraz badaniu naturalistycznym".
11. *Ewaluacja dialogiczna*: Kładzie nacisk na uwzględnienie wszystkich stron zaangażowanych w proces i zainteresowanych wynikami ewaluacji.

modele, jest jednak nieodporny na słabości spierających się stron oraz nie liczy się z tym, że kwestie związane z wartościami nie powinny być rozstrzygane metodą albo–albo. Taką listę plusów i minusów można by sporządzić dla każdego z modeli.

Każdemu, kto poważnie traktuje swoje studia nad ewaluacją, należałoby zalecić głębsze zaznajomienie się z różnymi modelami. Szczegółowo omawia je House (1978a). Określenia poszczególnych modeli nie zawsze są pomocne. Na przykład ewaluacja dialogiczna (*responsive*) zakłada całą serię posunięć, z których nie wszystkie dadzą się określić jako wrażliwe (*responsive*). Posunięcia te to (Stake, 1976):

a) identyfikacja problemów na podstawie rozmów z osobami bezpośrednio zaangażowanymi w program;

b) posłużenie się dokumentami programu w celu określenia pozostałych problemów;

c) bezpośrednia obserwacja działania programu (jeszcze przed fazą projektowania ewaluacji);

d) zaprojektowanie ewaluacji opartej na informacjach (a), (b) i (c);

e) przedstawienie ustaleń bezpośrednio wszystkim zainteresowanym, dobierając odpowiednie podejście do różnych adresatów.

Podobnie Tylerowskie podejście skoncentrowane na zoperacjonalizowanych celach behawioralnych ewaluowanych programów było istotnie

zawężone przez późniejsze interpretacje behaviorystyczne nastawione na specyfikowanie instrumentów testowania zachowań. Jak podkreślał Cronbach (1982, s. 222), Tayler był znacznie bardziej zainteresowany wyraźnym sprecyzowaniem niejasnych często celów programu i gdyby jego twórca twierdził, że celem tym jest „kształtowanie charakteru", Tyler wysondowałby, jacy uczniowie czego muszą dokonać, aby wskazywało to na jego realizację.

Następne komplikacje wynikają z rozbieżności stanowisk zarówno protagonistów poszczególnych modeli, jak i tych, którzy próbują je modyfikować. O ile więc wiedza na temat różnic dzielących konkretne modele wydaje się użyteczna dla przypomnienia choćby o szerokim zakresie istniejących możliwości, o tyle przesadą byłoby twierdzenie, że prowadzi ona do eklektyzmu. Potrzebna jest natomiast elastyczność i zdolność łączenia potrzeb i celów konkretnej ewaluacji.

Modele ewaluacji usiłują mieć charakter zadaniowy, co może być sposobem patrzenia na ewaluację w ogóle. Dlatego przydatna może być analiza od strony jej celów.

Cele ewaluacji

Różne rodzaje ewaluacji koncentrują się na różnych pytaniach i różnych aspektach ewaluacji. Amerykańskie Stowarzyszenie Badań Ewaluacyjnych (American Evaluation Research Society) sporządziło w 1980 r. kategoryzację typów ewaluacji, biorąc pod uwagę zarówno cel, jak i rodzaj podejmowanych działań:

a) *Analiza typu „front–end"* (przystępuje się do niej przed wprowadzeniem programu, obejmuje ona kontekst i możliwości zrealizowania programu). Celem jest poradnictwo w fazie planowania i wprowadzania programu w życie.

b) *Ocena możliwości przeprowadzenia ewaluacji.* Przedmiotem oceny są możliwości zastosowania różnych metod i sposobów ewaluacji.

c) *Ewaluacja formatywna* (rozwoju, procesu). Dostarcza informacji potrzebnych do korygowania programu, jego modyfikowania i sterowania nim.

d) *Ewaluacja skutków* (konkluzje, rezultat, efektywność). Określa rezultaty i efektywność programu, zwłaszcza pod kątem podjęcia decyzji o jego kontynuowaniu, rozwijaniu, zredukowaniu i finansowaniu.

e) *Monitorowanie programu.* Sprawdza zgodność z zakładanymi celami, analizuje wykonane usługi, liczy klientów.

f) *Ewaluacja ewaluacji* (ewaluacja wtórna, metaewaluacja, kontrola ewaluacji). Krytyka raportów z ewaluacji, ponowna analiza danych, ocena z zewnątrz ewaluacji wewnętrznej.

Celem tej listy nie było zasugerowanie, że każda ewaluacja powinna reprezentować jakiś czysty typ, raczej wskazanie różnych możliwości. W praktyce ewaluacja zwykle realizuje wiele, jeśli nie wszystkie z wymienionych celów i działań. W ramach każdej z sześciu generalnych kategorii wyróżnić można więcej szczegółowych rodzajów działań ewaluacyjnych:

a) *Ewaluacja świadomości.* Kto wie o działaniu programu? Jaka jest ta wiedza?

b) *Ewaluacja kosztów i korzyści.* Relacja między kosztami programu i zyskami (w kategoriach finansowych).

c) *Ewaluacja kosztów i efektów.* Relacja między kosztami programu i zyskami (w kategoriach innych niż finansowe).

d) *Ewaluacja na podstawie ustalonego kryterium.* Na ile udało się osiągnąć jakiś konkretny cel, przyjmując za kryterium pewien pożądany poziom.

e) *Weryfikacja jakości.* Czy wystarczająco systematycznie osiąga się minimum akceptowanych standardów? Jak jakość, o którą chodzi, można monitorować i prezentować.

Patton (1981) wymienił ponad sto typów ewaluacji. Zauważył także, że ewaluatorzy pozwalają sobie na działania nie cieszące się uznaniem: na *ewaluację szybką i pobieżną* (robioną możliwie szybko i jak najmniejszym kosztem), *ewaluację mierzoną w kilogramach* (opasły raport), *ewaluację oparta na przeświadczeniach i szacunkach* (formułowanie wniosków nie popartych danymi), *ewaluację „po uważaniu"* (ponieważ realizacją programu kierują ludzie mili, przyjacielscy, chętni do pomocy itd.).

Podobną listę „pseudoewaluacji" przedstawił Suchman (1967). Uwzględniała ona kilka potencjalnych motywów sponsorów ewaluacji: *mydlenie oczu* (nacisk na to, co widać powierzchownie), *wybielanie* (próby ukrycia ograniczeń lub niepowodzeń programu), *krecia robota* (zrobienie politycznego użytku z ewaluacji celem zniszczenia programu), *pozorowanie* (celebrowanie rytuału ewaluacji bez rzeczywistego zainteresowania albo zamiaru wykorzystania jej ustaleń; dzieje się tak, kiedy ewaluacja jest warunkiem sfinansowania programu), *odwlekanie* (wykorzystanie potrzeby przeprowadzenia ewaluacji jako pretekstu do opóźniania lub unikania działań). Jakkolwiek możliwości te zostały potraktowane nieco żartobliwie, to jednak sygnalizują ostrożność, z jaką powinno

się przystępować do ewaluacji — działalności zawsze mającej podtekst polityczny.

Ewaluacja formatywna i konkluzywna

Przedstawiona wyżej kategoryzacja Stowarzyszenia Badań Ewaluacyjnych wymienia podejścia do ewaluacji *formatywne* i *konkluzywne*. Rozróżnienie to przywoływane jest w wielu tekstach. Zajmiemy się nim nieco bardziej szczegółowo, ponieważ klienci znający trochę ewaluacyjny żargon mogą wyrażać swoje preferencje posługując się tą właśnie terminologią. Powtórzmy, rozróżnienie to odnosi się głównie do celu badania. *Ewaluacja formatywna ma pomóc w realizacji programu*, innowacji czy też innego jeszcze przedsięwzięcia, które poddajemy badaniom oceniającym. *Ewaluacja konkluzywna koncentruje się na ocenie rezultatów i skutków programu*. Odnosi się ona do całościowego jego oddziaływania; zajmuje się nie tylko odpowiedzią na pytanie, w jakim stopniu osiągnięte zostały założone cele, ale w ogóle wszystkimi możliwymi do wykrycia konsekwencjami. Zadania te nie muszą się wykluczać. Zwłaszcza ewaluacja konkluzywna może z powodzeniem oddziaływać na przyszłe projekty w fazie ich tworzenia. Oceny formułowane w wyniku ewaluacji w większości przypadków nie są ani wyłącznie negatywne, ani bez reszty pozytywne i zwykle dostarczają ważnych wniosków dla zmian w kolejnych programach (Birney, 1988).

Ewaluacja formatywna, rzecz jasna, powinna być prowadzona w takim momencie, by w razie potrzeby znalazł się jeszcze czas na dokonanie zmian. Występuje tu pewna kolizja między presją czasu a wymogami jakości: badanie wykonane szybko może okazać się wykonane byle jak i nie spełniać należycie warunków rzetelności i trafności, badanie dobre jakościowo może zająć zbyt wiele czasu, aby jego ustalenia mogły wpłynąć na istotne decyzje podejmowane w fazie tworzenia projektu. Takie są realia świata, w którym pracujesz. Do takich realiów należy też czas — tempo pracy bowiem zazwyczaj zależy nie od badacza, ale od kogoś innego. Podejmując badanie ewaluacyjne trzeba się należycie zastanowić nad warunkami jego realizacji. Projekt badania musi uwzględniać ograniczenia czasu i możliwości; sposób gromadzenia danych; możliwości uzyskania koniecznych do przeprowadzenia badań zezwoleń i współpracy; dostępność źródeł informacji itd.

Ewaluacja rezultatów i ewaluacja przebiegu programu

Podział na ewaluację *rezultatów* i ewaluację *przebiegu programu* odnosi się do podobnych kwestii, jak rozróżnienie ewaluacji formatywnej i konkluzywnej. Tradycyjnie pytania ewaluacyjne dotyczą kwestii związanych z rezultatami programu. Celem jest tu zbadanie, na ile program, praktyka, innowacja, interwencja czy polityka osiągają sformułowane na początku zadania lub zamierzenia. Podejście to, przez wielu wciąż uważane za właściwą ewaluację, coraz częściej traktowane jest jako niewystarczające.

Ewaluacja przebiegu programu przynieść ma odpowiedź na pytanie typu „jak" lub „co się dzieje". Polega na systematycznej obserwacji i analizie tego, co faktycznie wydarza się w toku programu, interwencji lub czegokolwiek, co jest przedmiotem ewaluacji. Bez tego rodzaju analiz badana kwestia może okazać się źle rozumiana lub w ogóle niejasna. Niezgodność między „oficjalnym" wyobrażeniem tego, co powinno być zrobione, i tym, co faktycznie ma miejsce, może być zupełnie zasadnicza. Nowy program, dajmy na to, nauki czytania może zakładać codzienne zajęcia indywidualne. Jednakże warunki pracy szkoły oraz nikle zainteresowanie programem dyrektora lub jego zastępcy sprawia, że większość takich zajęć nigdy się nie odbędzie. Ogólnie rzecz biorąc, poleganie na oficjalnych założeniach nie jest bezpieczne.

Ewaluacja przebiegu programu jest użytecznym uzupełnieniem ewaluacji rezultatów metodą *analizy systemowej* lub modelu *zoperacjonalizowanych zadań*. Ten ostatni działa w istocie na zasadzie „czarnej skrzynki" — koncentruje się na analizie tego, co wchodzi do „skrzynki" (czyli do programu), oraz tego, co z niej wychodzi. Badanie zachodzących procesów może rzucić na nie światło i ujawnić związki przyczynowe. Taka analiza procesów może być cenna sama w sobie, a także stworzyć lepsze podstawy do ewaluacji rezultatów. W pewnych okolicznościach wskazane być może ześrodkowanie uwagi na doświadczeniach i interakcjach będących konsekwencją programu, czyli przyjęcie ich za główne kryterium oceny wartości programu.

Przeprowadzanie ewaluacji

Jeżeli nie ma szans na przeprowadzenie badań ewaluacyjnych we właściwy sposób, nie należy ich podejmować. Ramka 3 zawiera kryteria, które powinny być spełnione, zanim zaangażujesz się w badania ewaluacyjne. Kryterium *przydatności* podkreśla wagę, jaką dla ewaluacji ma jej faktyczna użyteczność. W przeciwnym razie ewaluacja jest bezcelowa.

Ramka 3. Wyróżniki ewaluacji

1. *Przydatność*. Ewaluacja nie ma sensu, jeśli brak widoków na to, że ktoś ją wykorzysta.
2. *Wykonalność*. Ewaluację należy przeprowadzać tylko wtedy, gdy jest to możliwe ze względów politycznych i praktycznych, a efekty równoważą koszty.
3. *Przyzwoitość*. Ewaluacji dokonywać się powinno wtedy, gdy można wykazać, że zostanie przeprowadzona *uczciwe* i *zgodnie z zasadami etycznymi*.
4. *Adekwatność techniczna*. Mając pewność co do jej przydatności, wykonalności i przyzwoitości, ewaluację trzeba wykonać fachowo i precyzyjnie.

Pamiętajmy: „Ewaluacja ma służyć nie uzasadnieniu [programu], lecz poprawie [istniejącego stanu rzeczy]".

Lepiej też nie podejmować badań, których wyniki będą utrzymane w tajemnicy ze względów politycznych. Nie ma sensu realizowanie zlecenia, gdy wiadomo, że brak czasu i możliwości wyklucza poważne i odpowiedzialne badanie. Stronić także należy od „kreciej roboty", czyli badania mającego uzasadnić likwidację programu bądź przeciwnie — dostarczyć argumentów dla działań, które i tak już zostały postanowione. Unikać należy, oczywiście, wszystkiego, co w jakiś sposób może ograniczyć twoją swobodę działania.

Wreszcie, co wiąże się także z etyką, nie należy angażować się w badania, dopóki nie zostaną spełnione wszystkie potrzebne uwarunkowania techniczne. Ramka 4 przedstawia listę kwestii, o których należy pamiętać planując ewaluację. Ramka 5 zawiera zestawienie podstawowych umiejętności niezbędnych przy realizacji różnego typu ewaluacji.

Polityka ewaluacji

Właściwie nie zdarza się, by ewaluacja nie miała swego wymiaru politycznego. Inicjatywy i przedsięwzięcia społeczne mają swoich sponsorów i rzeczników. Większości towarzyszy krytyka i sceptycyzm. Personel odpowiadający za realizację programu może wiele zyskać lub stracić, w zależności od wyników ewaluacji. W grę wchodzić mogą nawet miejsca pracy. Ocena pozytywna może spowodować rozszerzenie programu, zwiększenie jego budżetu i innych zasobów, które mogą mieć znaczący wpływ na życie klientów programu.

Ewaluacje dotyczą zwykle projektów lub inicjatyw, które są jakoś widoczne na scenie politycznej, ogólnokrajowej lub lokalnej, albo znajdują się w polu zainteresowania biznesu, szkoły czy innej organizacji. Berk i Rossi (1990, s. 2) zauważyli, że „ewaluacje niemal bez reszty mieszczą się w obszarze bieżącej polityki". Innymi słowy, wiążą się z ocenianiem

Ramka 4. Kwestie ważne w fazie planowania ewaluacji.

1. *Powody, cele i motywy*
   - Czy ewaluacja ma służyć tobie czy komuś innemu?
   - Dlaczego przeprowadzasz ewaluację?
   - Kto powinien otrzymać informacje?
2. *Wartości*
   - Czy w wyniku ewaluacji można przedsięwziąć działania lub podjąć decyzje?
   - Czy jest ktoś lub coś, co blokuje te decyzje?
3. *Interpretacja*
   - Czy strony zaangażowane w ewaluację uzgodniły jej rodzaj?
4. *Przedmiot*
   - Jakiego rodzaju informacji potrzebujesz?
5. *Ewaluator (ewaluatorzy)*
   - Kto zbiera informacje?
   - Kto pisze raporty?
6. *Metody*
   - Jakie metody należy zastosować dla uzyskania wymaganych informacji?
   - Czy mogą być one opracowane i zastosowane w czasie, jakim dysponujesz?
   - Czy metody zostaną zaakceptowane przez osoby, których dotyczy ewaluacja?
7. *Czas*
   - Ile czasu można zarezerwować na ewaluację?
   - Czy wystarcza czasu na zebranie i analizę danych?
8. *Pozwolenie i kontrola*
   - Czy uzyskane zostały wszelkie pozwolenia konieczne do przeprowadzenia ewaluacji?
   - Czy udział w badaniach jest dobrowolny?
   - Kto decyduje o tym, co ma się znaleźć w raporcie?
9. *Wykorzystanie*
   - Kto decyduje o sposobie wykorzystania ewaluacji?
   - Czy uczestnikom ewaluacji przedstawiona zostanie robocza wersja raportu?
   - Czy forma raportu dostosowana jest do jego odbiorców (styl, objętność)?

Pamiętaj także:
- *Zachowaj jak największą prostotę* — unikaj analiz skomplikowanych.
- *Myśl defensywnie* — jeśli coś może się nie udać, to na pewno się nie uda, tak więc próbuj uprzedzać potencjalne problemy.

Opracowano według Harlen, Elliott, 1982; patrz także Robson i współaut., 1988, s. 85.

polityki i projektów, w które angażują się ludzie parający się polityką. Kilka lat temu amerykański Departament Edukacji i Nauki pod presją żądań rzeczników wprowadzenia specjalnego programu edukacji dla dzieci z porażeniem mózgowym, na wzór działań Peto Institute w Budapeszcie, sfinansował duże badanie ewaluacyjne mające ocenić jego efektywność. Obecnie większość funduszy przeznaczonych na ewaluację pochłaniają badania nad projektami znajdującymi się w centrum uwagi opinii publicznej w związku z wprowadzaniem Narodowego Programu Edukacji, m.in. nad Ujednoliconymi Testami Wyników Nauczania (Standard Attainment Testing).

Ramka 5. Umiejętności potrzebne w toku ewaluacji

Wymieniamy najbardziej podstawowe:
- pisanie projektu,
- precyzyjne określenie celów ewaluacji,
- dobór, zorganizowanie i współpraca w zespole badawczym,
- wybór projektu i technik zbierania danych,
- przeprowadzenie wywiadów,
- konstruowanie kwestionariusza i posługiwanie się nim,
- obserwacja,
- posługiwanie się złożonymi systemami informacji,
- analiza danych,
- pisanie raportów, formułowanie rekomendacji,
- zapewnienie spożytkowania wyników,
- wrażliwość na kontekst polityczny.

Ze względu na implikacje polityczne ewaluacją i jej wynikami interesują się wszyscy: władze federalne, stanowe i lokalne, politycy i urzędnicy, agencje i ich przedstawiciele odpowiedzialni za realizację projektów, osoby odpowiedzialne za organizację usługi, jej klienci, związki zawodowe, które same muszą troszczyć się o własne interesy, może także podatnicy i obywatele w ogóle. Mało prawdopodobne jest, by interesy wszystkich grup były zbieżne. Wyniki ewaluacji dla jednych będą korzystne dla innych nie.

Oznacza to między innymi, że prowadzenie ewaluacji nie jest zajęciem odpowiednim dla osób nieodpornych na sprzeczne naciski i niepewnych swoich racji w obliczu kontrowersji. Krytyka może mieć charakter metodologiczny (dotyczyć sposobu prowadzenia analiz) albo polityczny (dotyczyć ustaleń); niekiedy wreszcie kwestie metodologiczne mogą służyć za parawan, za którym kryje się polityka (Berk, Rossi, 1990). Wynika stąd przede wszystkim, że zarówno w fazie planowania badań, jak w trakcie ich realizacji nie ma takich spraw, choćby najbardziej szczegółowych, które można by uznać za mało ważne. Warto też upewnić się, czy uwzględnione zostały uzasadnione interesy osób, z którymi przyjdzie mieć do czynienia w toku badań.

Szacowanie potrzeb

Program innowacyjny lub nowy rodzaj usług z reguły są wprowadzane ze względu na jakieś potrzeby, o których sądzi się, że nie są zaspokajane. Logicznie rzecz biorąc, ocena tych potrzeb powinna być dokonana przed opracowaniem programu i odpowiadać za nią powinny osoby planujące program. Na ogół jednak to właśnie ewaluatorzy są

proszeni o pomoc przy szacowaniu potrzeb lub wręcz zleca się im to zadanie. Bywają także sytuacje, gdy te same osoby realizują program, a potem przeprowadzają jego ewaluację. Może więc przydać się kilka uwag na ten temat.

Oszacowanie potrzeb wymaga najpierw ustalenia, jakie potrzeby występują, a potem określenia hierarchii tych potrzeb. Wszyscy wiedzą, co oznacza „potrzeba" w rozumieniu potocznym, warto może jednak przypomnieć, że potrzeby pojawiają się wtedy, gdy pogłębia się rozziew między faktycznie obserwowanym a pożądanym stanem rzeczy. Proponowano kilka sposobów szacowania potrzeb (McKillip, 1987). Ramka 6 przedstawia jeden z nich. Satcher ze swoimi współpracownikami (1980) opisał podobną strategię zastosowaną przy opracowywaniu programu usług medycznych dla mieszkańców jednego z ośrodków miejskich (patrz też Kosccoff, Fink, 1982).

Ramka 6. Etapy szacowania potrzeb

1. *Określenie potencjalnych celów.* Pierwszym etapem jest skompletowanie możliwie wyczerpującego zestawu celów, które mogą być brane pod uwagę. Podpowiedzieć je może literatura (jeśli takowa istnieje), eksperci w danej dziedzinie, ci, którzy mogą być adresatami przewidywanego programu.

2. *Decyzja, które z tych celów należy uznać za ważne.* Wybierz grupy reprezentujące różne strony: prawdopodobnych odbiorców usług lub mających korzystać z programu, osoby dostarczające usług oraz menedżerów programu. Niech ocenią, na ile ważny jest każdy z potencjalnych celów.

3. *Określenie środków, jakimi można w danym momencie dysponować dla zrealizowania tych uważnych celów.* W niektórych przypadkach etap ten można połączyć z etapem poprzednim i za jednym zamachem uzyskać opinie i oceny respondentów, a jeśli to możliwe — także informacje na temat obecnego poziomu zaspokojenia potrzeb. Innym sposobem może być zamówienie wypowiedzi nagranych na taśmach lub przeprowadzenie małego sondażu.

4. *Ostateczny wybór celów.* Należy wybrać te cele, które są uznawane za ważne, a równocześnie są w niewielkim stopniu osiągane. Można brać także pod uwagę inne kryteria, np. możliwość realizacji. Programy bądź usługi należy tak dostosować do potrzeb i celów, by jak najlepiej służyły ich realizacji (np. projektodawcy nastawiali się na rodzaj samokształcenia, ale ocena potrzeb wykazuje, że ich zaspokojeniu lepiej służyć będą indywidualne konsultacje).

Należy zdawać sobie sprawę z niebezpieczeństw, jakie niesie ze sobą koncentrowanie się na potrzebach „możliwych do zaspokojenia". Judd (i współaut., 1991, s. 408) pokazał, że takie czysto techniczne decyzje pociągają za sobą konsekwencje w postaci pewnej ideologii. Jako przykład podaje program zapobiegania przestępczości. Jeśli zdarzenia prowadzące do przestępstwa uznać za łańcuch przyczynowy, to interwencja w jakimkolwiek jego punkcie powinna zmniejszyć prze-

stępczość. Interweniować można różnie: można zająć się warunkami, w jakich dzieci wyrastają na przestępców; można organizować szkolenie zawodowe dla bezrobotnych, żeby mogli znaleźć pracę; można wreszcie propagować odpowiednie zabezpieczanie domów i mieszkań. Programy, które koncentrują się na zakładaniu zamków w drzwiach, dodatkowego oświetlenia posesji i alarmów antywłamaniowych są łatwe do zrealizowania, ale nie są obojętne ideologicznie: sugerują mianowicie, że chroniąc potencjalne ofiary ograniczamy przestępczość.

Uwagi końcowe

Jako potencjalny ewaluator musisz orientować się w zagadnieniach związanych z rozwojem wstępnej fazy projektowania, wyborem ogólnej strategii badania oraz specyficznych metod i technik zbierania danych. Istniała tendencja w badaniach realizowanych na mniejszą skalę do utożsamiania ewaluacji z użyciem kwestionariuszy (jako jednej z podstawowych metod), ale ewaluacja jest znacznie bardziej złożoną dziedziną, w której najlepiej stosować metody kombinowane. W wielu ewaluacjach stosuje się metody zarówno jakościowe, jak ilościowe, dlatego znać musisz szeroki zakres właściwych metod i technik analizy. Jedną z podstawowych kwestii jest także zrozumiały sposób prezentowania wyników ewaluacji. Jest to szczególnie istotne dla tych, którzy na ich podstawie podejmować będą konkretne działania.

Twoja praca polega na wyborze strategii badawczej, metod analizy i zbierania danych. Zalecana jest elastyczność w wyborze najbardziej odpowiednich strategii i metod, które pozwolą na realizację konkretnego zadania. Równie ważna jest dyscyplina i systematyczność zbierania danych. *Szczególnie istotna jest jednak praktyczna przydatność uzyskanych danych.* Jones (1985) pisze: „Jeżeli możesz dowiedzieć się czegoś potrzebnego ci o programie rozmawiając z kilkoma niezadowolonymi pracownikami, rozmawiaj z nimi. Jeżeli jedyną drogą uzyskania potrzebnych ci danych jest obserwacja uczestnicząca, uczestnicz i obserwuj (nie zapomnij tylko o właściwym dokumentowaniu). Jeżeli potrzebujesz danych archiwalnych, zlokalizuj właściwe źródła i wykorzystaj je. Wykorzystaj wszystkie metody, jakimi dysponujesz, aby wykonać swoją pracę" (s. 258).

tłumaczył *Marcin Ciechomski*

Michael Q. Patton

Obserwacja — metoda badań terenowych

Michael Q. Patton, „Fieldwork and Obserwation", rozdział z książki tegoż autora *How to Use Qualitative Methods in Evaluation*, Sage Publications, London 1987.

Ewaluacja jakościowa dostarcza szczegółowego opisu działań, procesów i osób uczestniczących w ocenianym programie. Ważnym źródłem danych jakościowych jest bezpośrednia obserwacja. Oznacza to, że badacz musi „udać się w teren", terenem zaś jest miejsce, w którym jest realizowany program interesujący ewaluatora. Zadaniem ewaluatora-obserwatora jest jak najbardziej szczegółowe i precyzyjne rejestrowanie tego wszystkiego, co zaobserwował i doświadczył tam „w terenie", notatki zaś, które sporządził w trakcie swojej pracy, będą mu służyć jako podstawa ewaluacji jakościowej.

Jakość danych zebranych w wyniku obserwacji w dużym stopniu zależy od kwalifikacji, przygotowania i kompetencji ewaluatora. „Praca w terenie" wymaga czegoś znacznie więcej niż tylko umiejętności pobieżnej obserwacji wystarczającej w życiu codziennym. Wyszkolony obserwator umie rozpoznać i precyzyjnie opisać pełne znaczeń procesy i interakcje zachodzące między ludźmi. Poza odpowiednim przygotowaniem i niezbędną praktyką badacz terenowy musi posiadać umiejętność koncentracji, musi być cierpliwy, czujny, wrażliwy i odporny psychicznie. Obserwacja jako metoda badawcza stawia wysokie wymagania i fizyczne, i intelektualne. Przy omawianiu niedokładności i błędów codziennej obserwacji nienaukowej zapomina się często, ile wysiłku, wprawy i przygotowania wymagają metody jakościowe.

Obserwacje osób nie przeszkolonych i nie przygotowanych

W swoim popularnym „Przewodniku dla osób zajmujących się badaniami społecznymi" Katzer, Cook i Crouch (1978) rozdział dotyczący obserwacji zatytułowali „Nie wierzyć własnym oczom" („Seeing Is Not Beliving"). W rozdziale tym omawiają często powtarzaną historię, która ilustruje istotę problemu związanego z obserwacją jako źródłem danych.

„Na środek sali, w której odbywało się to spotkanie naukowe, wyskoczył nagle człowiek. Gonił go inny mężczyzna z rewolwerem. Zaczęli się bić na oczach zgromadzonych naukowców; z lufy rewolweru błysnął ogień, po czym obaj szamoczący się mężczyźni wybiegli z sali. Całe zdarzenie trwało około dwudziestu sekund. Zaraz po tym zajściu, osoba prowadząca spotkanie poprosiła wszystkich jego uczestników o relację z tego, co widzieli. Obserwatorzy nie wiedzieli, że awantura została zaaranżowana i zainscenizowana z dużą dbałością o szczegóły. Całe to zajście sfilmowano. Spośród czterdziestu raportów tylko jeden zawierał mniej niż 20% błędnych informacji dotyczących podstawowych faktów, większość zaś raportów zawierała ponad 40% błędnych informacji. To zdarzenie bez wątpienia skupiło na sobie całą uwagę obserwatorów, było dla wszystkich wyraźnie widoczne i trwało tylko dwadzieścia sekund, a jednak nie byli w stanie zaobserwować wszystkiego, co tam się stało. Parę osób chichotało ukradkiem czytając później własne raporty. Podobne sytuacje opisywano wiele razy i można powiedzieć, że naukowcy nie różnią się w swoich reakcjach od innych ludzi" (s. 21–22).

Historia ta zdaje się wskazywać, że dane zebrane w wyniku obserwacji są niewiarygodne, a ludzie są obserwatorami, na których nie można polegać. Jednakże w rzeczywistości opowieść ta ilustruje fakt, że niewiarygodne są te obserwacje, które zostały poczynione przez osoby do tego nie przygotowane i nie przeszkolone. Analizując sytuacje takie, jak cytowana wyżej, często nie dostrzega się, że, po pierwsze, naukowcy ci nie byli *przeszkoleni* do roli wykwalifikowanych obserwatorów oraz, po drugie, nie byli przygotowani do przeprowadzenia obserwacji w tym konkretnym momencie.

Badanie naukowe metodą obserwacji wymaga odpowiedniego przeszkolenia i dokładnych przygotowań. Sam fakt, że człowiek wyposażony jest w pięć zmysłów — wzroku, smaku, słuchu, węchu i dotyku — nie kwalifikuje go jako obserwatora. To, że parę przypadkowych osób będących świadkiem jakiegoś zdarzenia opisało je w całkiem różny sposób, nie oznacza, że *przeszkoleni i przygotowani* obserwatorzy nie potrafią opisać go dokładnie, trafnie i rzetelnie.

Przeszkolenie i przygotowanie wykwalifikowanego obserwatora

Szkolenie obserwatora obejmuje opanowanie umiejętności opisywania, sporządzania notatek w trakcie obserwacji, wychwytywania ważnych szczegółów spośród nieistotnych detali oraz stosowania metod precyzyjnej kontroli trafności obserwacji. Przeszkolenie badacza tak, by stał się wnikliwym i wykwalifikowanym obserwatorem, jest zadaniem trudnym, wiele bowiem osób sądzi, że zdolność obserwacji jest cechą przyrodzoną człowieka, a więc niczego nie muszą się tu uczyć. Tymczasem wykwalifikowani obserwatorzy muszą być przeszkoleni nie mniej rygorystycznie niż ci, którzy, zajmując się naukami społecznymi, stosują w swej pracy metody ilościowe.

Staranne przygotowanie się do obserwacji jest tak samo ważne jak systematyczne szkolenie. Przygotowanie to ma wymiar intelektualny, fizyczny, psychologiczny i materialny. Ludwik Pasteur powiedział kiedyś: „Co się tyczy obserwacji, to szansę ma tylko umysł przygotowany". Częścią tego przygotowania umysłu do obserwacji jest nauczenie się koncentracji uwagi podczas obserwowania interesującego nas zdarzenia. Naukowa obserwacja nie jest czynnością, którą można podjąć bez przygotowania w dowolnym momencie — podobnie mistrz olimpijski nie pobije rekordu świata, gdy ktoś zaczepi go na ulicy i zechce sprawdzić, jak szybko potrafi biegać, a światowej klasy bokser nie obroni tytułu w przygodnej bójce.

Na dwie sprawy trzeba tu zwrócić uwagę. Po pierwsze, zwyczajna zdolność obserwacji, będąca niczym innym jak tylko wybiórczą percepcją otaczającej nas rzeczywistości, może wystarczać, gdy chodzi o udział w codziennych wydarzeniach. Po drugie, wykwalifikowany obserwator jest w stanie poprawiać dokładność, trafność i rzetelność swoich obserwacji dzięki ukierunkowanemu szkoleniu i dokładnemu przygotowaniu.

Rozróżnienie to jest ważne również dla określenia zalet i niedoskonałości różnych metod badawczych. Jakość *każdej* ewaluacji w dużym stopniu zależy od kompetencji i profesjonalizmu ewaluatora, w przypadku zaś ewaluacji jakościowej — przede wszystkim od przeszkolenia i przygotowania ewaluatora obserwującego realizowany program.

Wartość danych uzyskanych w wyniku obserwacji

Dane zgromadzone w toku obserwacji mają służyć przede wszystkim pełnemu i szczegółowemu *opisaniu* realizowanego programu dla potrzeb jego ewaluacji. Opis ten musi uwzględniać działania podjęte

w ramach programu, osoby w nim uczestniczące, a także znaczenia, jakie to, co zostało zaobserwowane, miało dla wszystkich uczestników. Raporty z obserwacji muszą zawierać tyle i takich szczegółów, aby osoby, do których raport ten jest adresowany, otrzymały pełny obraz sytuacji. Opisy muszą odpowiadać faktom, muszą być dokładne i kompletne. Wszelkie nieistotne szczegóły i uogólnienia, rozmywające obraz realizowanego programu, powinny zostać usunięte z końcowego opisu. Podstawowa wskazówka przy sporządzaniu raportu z obserwacji powinna brzmieć: pisz tak, aby czytelnik twojego sprawozdania miał wrażenie, że uczestniczy w opisywanej przez ciebie sytuacji.

Podstawową zaletą obserwacji sytuacyjnych jest to, że dane są gromadzone w terenie, czyli tam, gdzie realizowane są zadania określone przez dany program. Świadczą o tym terminy określające sposób gromadzenia danych metodą obserwacji: „obserwacja uczestnicząca, obserwacja terenowa, obserwacja jakościowa, obserwacja bezpośrednia lub badanie terenowe. Wszystkie te określenia wskazują na to, że badacz jest «w» lub «obok» dziejących się sytuacji społecznych po to, aby te sytuacje poddawać analizie jakościowej" (Lofland, 1971, s. 93).

Obserwacje terenowe mają wiele zalet dla realizacji zadań ewaluacyjnych. Po pierwsze, bezpośrednia obserwacja programu daje ewaluatorowi możliwość lepszego zrozumienia *kontekstu*, w jakim podjęto działania wynikające z realizacji danego programu, a zrozumienie kontekstu jest istotne dla holistycznej perspektywy badawczej.

Po drugie, własne doświadczenia ewaluatora w zetknięciu z realizacją programu pozwalają mu swobodniej poruszać się w zagadnieniu. Obserwując wszystko z bliska, ewaluator nie musi ograniczać się do założeń wyjściowych programu. Może osobiście przekonać się, jak program funkcjonuje. Dzięki temu jego ewaluacja nastawiona jest na wydobycie tych zjawisk, które świadczą o faktycznej, a nie tylko zakładanej dynamice programu.

Po trzecie, zaletą metod obserwacyjnych jest to, że specjalnie przeszkolony ewaluator może dostrzec te zjawiska, które umykają codziennej uwadze nie przygotowanych uczestników programu. Ktoś, kto ma przekazać jakieś informacje, musi mieć pewność, że posiada właśnie te informacje, których się od niego oczekuje. Ponieważ każdy system społeczny funkcjonuje do pewnego stopnia rutynowo, uczestnicy takich rutynowych działań przyjmują je jako coś oczywistego i nie są świadomi ważnych niuansów, które może dostrzec tylko osoba z zewnątrz, obserwator wcześniej nie uczestniczący w owych rutynowych działaniach.

Po czwarte, wartość podejścia badawczego związanego z bezpośrednią obserwacją polega w dużym stopniu na tym, że obserwator może dostrzec te fakty, o których uczestnicy programu w wywiadach mogą mówić niechętnie. Pewne problemy mogą być dla nich drażliwe i podczas wywiadu nie będą chcieć lub nie będą w stanie opowiadać o niektórych faktach ważnych dla opisu całości programu. Tak więc jedynie dzięki bezpośredniemu doświadczeniu i obserwacji konkretnych wydarzeń ewaluator ma możliwość dotarcia do tych informacji, które w inny sposób byłyby dla niego niedostępne.

Po piąte, obserwacje pozwalają ewaluatorowi wyjść poza selektywne spostrzeżenia innych osób. Wywiady odtwarzają sposób rozumienia respondentów. Rozumienie to może być źródłem ważnych i często kluczowych informacji, jednakże trzeba mieć na uwadze, że wywiady zawsze informują o spostrzeżeniach, a spostrzeżenia są bardzo wybiórcze. Oczywiście, spostrzeżenia ewaluatora mogą być równie selektywne, lecz właśnie dzięki uwzględnieniu własnych spostrzeżeń jako części danych ewaluator ma możliwość zaprezentowania bardziej wszechstronnego obrazu analizowanego programu.

Na koniec trzeba zauważyć, że własne doświadczenia ewaluatora związane z faktem, iż ma bezpośredni kontakt z badanym przez siebie środowiskiem społecznym, powodują że ma on również bezpośredni dostęp do osobistej wiedzy i indywidualnych doświadczeń uczestników programu, co jest pomocne w rozumieniu i interpretacji tego wszystkiego, co jest przedmiotem obserwacji. Ważną częścią badań terenowych jest właśnie refleksja i introspekcja. Wrażenia i odczucia obserwatora stają się częścią zgromadzonych danych, które mogą posłużyć do lepszego zrozumienia całego programu i ludzi w nim uczestniczących. Obserwator zbiera informacje i różne odczucia uczestników programu, które są czymś więcej niż tylko szczegółowymi notatkami z pracy w terenie. Rozumienie jest ważne dla analizy zebranego materiału, a także dla rekomendacji na przyszłość, tj. wskazówek służących usprawnieniu programu w dłuższej perspektywie czasowej. Rekomendacje oparte na danych jakościowych mających źródło w bezpośrednich doświadczeniach i bezpośrednich obserwacjach programu — są szczególnie cenne, jako że czerpią z bezpośredniego rozumienia realiów realizowanego programu, a nie tylko abstrakcyjnych, wysmażonych przy biurku założeń leżących u jego podstaw. Ewaluacyjna praca w terenie powinna dostarczać jak najbardziej praktycznych i ważnych rekomendacji służących przyszłemu doskonaleniu programu.

Implikacje obserwacji jako metody

Wybór obserwacji jako metody badawczej nie zamyka sprawy. Decyzja, że informacje o programie będzie się zbierać metodami terenowymi, jest tylko pierwszym krokiem, który pociąga za sobą szereg konsekwencji. Wybierając metodę pracy w terenie, trzeba włączyć się w program jako pewien układ sytuacyjny, zdawać sprawozdanie ze wszystkiego, co się zaobserwowało, i poznać zapatrywania wszystkich uczestników programu. Podjąwszy decyzję o takim zaangażowaniu się, trzeba następnie zadecydować, które ze szczegółowych technik obserwacji są właściwe dla przeprowadzenia ewaluacji danego programu.

Widz czy uczestnik?
Różne zaangażowanie obserwatora

Strategie obserwacyjne różnią się między sobą przede wszystkim i zasadniczo tym, do jakiego stopnia ewaluator-obserwator jest także uczestnikiem tych wszystkich poczynań programowych, które są przedmiotem jego analizy. Naprawdę chodzi o coś więcej niż zwykły wybór między uczestniczeniem i nieuczestniczeniem. Między tymi skrajnościami rozciąga się kontinuum możliwości — od całkowitego wtopienia się w program jako pełny uczestnik do całkowitego odseparowania się od obserwowanych działań i przyjęcie roli „widza".

Decyzja obserwatora, do jakiego stopnia chce być także uczestnikiem, nie przesądza sprawy raz na zawsze. Stopień uczestnictwa może się zmieniać wraz z upływem czasu. W niektórych przypadkach ewaluator może zacząć od roli widza, by w toku dalszych badań stać się uczestnikiem podejmowanych działań. Kiedy indziej ewaluator — chcąc osobiście wniknąć w sytuację — może rozpoczynać swoją pracę od pełnego uczestnictwa, a wraz z upływem czasu stopniowo wycofywać się, aby pod koniec znaleźć się w roli niezależnego i nie uczestniczącego obserwatora widza.

Obserwacja uczestnicząca jest strategią badawczą wymagającą posługiwania się wieloma różnymi technikami: „[...] analizą dokumentów, wywiadem socjologicznym, obserwacją przy bezpośrednim współudziale i introspekcją" (Denzin, 1978, s. 183). W obserwacji uczestniczącej ewaluator nawiązuje bliski kontakt z ludźmi, do których program jest adresowany, bierze udział w ich życiu i działaniach. Celem takiego zbliżenia jest uzyskanie spojrzenia na to, co się dzieje, *od wewnątrz*. Oznacza to, że ewaluator nie tylko ogląda to, co dzieje się dookoła niego, lecz także *odczuwa* tak, jak gdyby był członkiem obserwowanej grupy.

Doświadczanie sytuacji z perspektywy ich uczestnika jest tym, co *konstytuuje* istotę obserwacji uczestniczącej. Jednakże równie ważna jest bezstronna *obserwacja „z boku"*. Prawdziwym wyzwaniem dla ewaluatora jest połączenie uczestnictwa i obserwacji w taki sposób, aby zrozumieć to, czego jest świadkiem, z perspektywy uczestnika, a opisać z perspektywy widza.

Do jakiego stopnia ewaluator może stać się uczestnikiem tego, w czym bierze udział, zależy częściowo od specyfiki obserwowanego programu. Na przykład zajmując się jakimś programem edukacyjnym dla dzieci ewaluator nie może stać się uczniem i podzielić doświadczeń dzieci, może natomiast wcielić się w rolę kogoś dorosłego związanego z tym programem, np. jednego z rodziców, wolontariusza lub kogoś ze szkolnego personelu, i w ten sposób wyrobić sobie pogląd na ten program z perspektywy jego uczestnika.

Programy, których realizacja dotyczy jakiejś szczególnej grupy osób, mogą stwarzać również pewne naturalne bariery utrudniające ewaluatorowi *pełne* uczestnictwo. Przykład: ewaluator, który nie jest uzależniony od jakichś używek, nie będzie w stanie zostać autentycznym uczestnikiem programu adresowanego do osób uzależnionych, aczkolwiek może włączyć się do takiego programu, jako uczestnik-obserwator, aby osobiście zaznajomić się ze sposobem leczenia. Taki udział w programie terapeutycznym w roli klienta może zaowocować pełniejszym rozeznaniem, jak taki program powinien wyglądać, tyle że ewaluator musi pozbyć się złudzeń, że jest to udział w pełnym tego słowa znaczeniu.

Taką sytuację ilustruje dialog między ewaluatorem a młodym więźniem, zarejestrowany podczas obserwacji uczestniczącej w więzieniu.

Więzień: Co tu robisz, człowieku?
Ewaluator: Wpadłem tu zobaczyć, jak to jest być w więzieniu.
Więzień: Co ty gadasz? Zobaczyć, jak to jest w więzieniu?
Ewaluator: Jestem tutaj, żeby przekonać się na własnej skórze, jak jest w więzieniu, a nie dowiadywać się z zewnątrz.
Więzień: Nie rozśmieszaj mnie, człowieku. Chyba nie mówisz poważnie. Na własnej skórze? Cholera! Człowieku, możesz pójść do domu, kiedy tylko poczujesz, że masz dość. No nie?
Ewaluator: W zasadzie tak.
Więzień: Więc nie dowiesz się, jak tu jest naprawdę.

Uczestnictwo mogą także ograniczać czynniki społeczne i polityczne. Jeśli członkowie jakiejś zbiorowości znają się bardzo dobrze, mogą nie życzyć sobie, aby dołączył do nich ktoś z zewnątrz. Tam gdzie występują wyraźne różnice społeczne między ewaluatorem a uczestnikami programu, może zachodzić podejrzenie, iż w uczestnictwie obserwatora

kryje się jakiś podstęp. Również organizatorzy programu mogą obawiać się, że konieczność włączenia ewaluatora obciąży skromny budżet, a dodatkowy uczestnik naruszy równowagę ilościową między zespołem obsługującym a klientami.

Trzeba w tym miejscu zaznaczyć, że ewaluatorzy którzy stosują jakościowe strategie badawcze, muszą unikać fałszywego przekonania, że *idealem* jest zawsze pełne uczestnictwo w programie, które często określa się jako „wejście w skórę" (*going native*). *Idealem jest ustalenie i zastosowanie takiego stopnia uczestnictwa, które dostarczy najbardziej znaczących danych pozwalających scharakteryzować uczestników programu, określić rodzaj pytań wymagających odpowiedzi oraz kontekst społeczno-polityczny, w którym program jest realizowany.*

I na koniec jeszcze jedna uwaga: plany i intencje ewaluatora związane z jego zaangażowaniem w program mogą się nieoczekiwanie zmienić. Lang i Lang (1960) podają przykład dwóch naukowców prowadzących obserwację uczestniczącą zachowania publiczności uczestniczącej w kazaniach Billy'ego Grahama. Ewaluatorzy ci „zwrócili się ku Bogu", porzucili swoje posady i przyłączyli się do krucjaty wielebnego Grahama. Jest to przykład ryzyka zawodowego (lub korzyści — zależnie od przyjętej perspektywy), jakie niesie za sobą badanie metodą obserwacji uczestniczącej.

Obserwacja jawna i niejawna

Kwestia trafności i rzetelności danych uzyskanych w wyniku obserwacji wiąże się głównie z wpływem obserwatora na to, co jest przedmiotem jego obserwacji. Ludzie mogą zachowywać się zupełnie inaczej, gdy widzą, że są obserwowani, inaczej zaś, gdy nie są tego świadomi. Stąd przekonanie, że obserwacja niejawna częściej pozwala odkryć, co się rzeczywiście dzieje w ramach programu, obserwacja jawna natomiast powoduje, że osoby uczestniczące w programie i świadome, iż to, co robią, jest przedmiotem czyjegoś zainteresowania, mogą w jakiś sposób swoje zachowania modyfikować.

Na temat etycznych i moralnych stron obserwacji niejawnej powiedziano już prawie wszystko. Na jednym krańcu kontinuum sytuuje się Edward Shils, który całkowicie neguje wszelkie formy obserwacji niejawnej. Sprzeciwia się „jakiejkolwiek technicznie wykonalnej obserwacji zachowań prywatnych bez wiedzy i wyraźnej zgody osób obserwowanych". Głosi pogląd, że cel badań nigdy nie powinien być tajemnicą, tak więc technika obserwacji uczestniczącej jest „moralnie naganną manipulacją", jeśli obserwator zaraz na wstępie nie wyłoży uczestnikom

badań swoich pytań badawczych (Shils, 1959, cyt. za Webb i współaut., 1966, s. VI).

Przeciwne stanowisko reprezentuje Jack Douglas (1976) proponując podejście nazwane „śledztwem społecznym" (*investigative social research*). Douglas dowodzi, że tradycyjne metody antropologicznych badań terenowych opierają się na jednomyślnym założeniu, że ludzie są z gruntu gotowi do współpracy i niesienia pomocy, że pragną, aby ich punkt widzenia był rozumiany i podzielany przez innych. W przeciwieństwie do tego modelu opartego na konsensie społecznym, Douglas proponuje paradygmat wychodzący od konfliktu społecznego. Paradygmat ten pozwala mu wierzyć, że wszystkie niejawne metody badawcze, które zbliżają badacza do prawdy, powinny być uznane i zaakceptowane jako dopuszczalne.

„Paradygmat śledczy opiera się na założeniu, że życie społeczne przenika głęboki konflikt interesów, wartości, uczuć i działań. Trzeba przyjąć jako coś oczywistego, że wiele osób — być może w mniejszym lub większym stopniu wszyscy — ma coś do ukrycia przed innymi, a nawet oszukuje się nawzajem. Zamiast wierzyć innym i oczekiwać, że oni nam uwierzą, mamy wiele powodów, aby być podejrzliwi wobec innych ludzi, tak jak inni mogą być podejrzliwi wobec nas. Konflikt jest rzeczywistością naszego życia, podejrzliwość zaś jego wiodącą zasadą. [...] Jest to wojna wszystkich przeciw wszystkim i nikt nie daje nikomu niczego za darmo, a przede wszystkim — prawdy" (Douglas, 1976, s. 55).

Wyjawienie celu badań

Tak jak w przypadku obserwacji uczestniczącej nie był to wybór albo-albo, tak i jawność obserwacji oraz celów ewaluacji nie jest kwestią albo-albo. Poinformowanie uczestników ewaluowanego programu, że są obserwowani, oraz powiadomienie ich o celach prowadzonej ewaluacji jest stopniowalne w granicach od pełnej jawności do całkowitego utajnienia.

Złożoność badań ewaluacyjnych implikuje wiele poziomów, których mogą dotyczyć decyzje o jawności lub niejawności prowadzonej obserwacji oraz o wyeksplikowaniu lub ukryciu celów ewaluacji. Czasem tylko osoba zamawiająca ewaluację zna jej pełny zakres i rzeczywisty cel prowadzonych obserwacji. Kiedy indziej o uczestnictwie obserwatora w realizowanym programie mogą być poinformowane jedynie osoby na kluczowych stanowiskach. Innym razem ewaluator może ujawnić cel i charakter własnego uczestnictwa jedynie uczestnikom najniższego

szczebla, prosząc ich jednocześnie o współpracę w sekrecie przed decydentami i innymi wpływowymi osobami. Czasami sytuacja, w której znalazł się ewaluator, jest tak złożona, że on sam już nie wie, kto z uczestników jest poinformowany o jego pracy. Może to prowadzić do komicznych sytuacji, kiedy wszyscy uczestnicy wiedzą, że są obserwowani przez ewaluatora, on sam zaś nie wie, że inni znają jego rolę w tym programie.

Na koniec wymienić należy problem poufności. Ci, którzy są zwolennikami badań niejawnych i zalecają utrzymanie w tajemnicy celów ewaluacji, zwykle domagają się zatajenia w raporcie końcowym wszelkich nazwisk, miejsc i innych łatwo dających się zidentyfikować informacji, które w przyszłości mogłyby przynieść szkodę uczestnikom realizowanego programu. Jednakże w badaniu ewaluacyjnym, w którym ważne jest wyraźne opisanie tego, co kto powiedział, zachowanie poufności zebranych informacji jest niezwykle trudne. Co więcej, taka poufność może rozmijać się z celami, dla których ewaluacja była przeprowadzona.

Ewaluator musi znaleźć właściwe rozwiązanie tych problemów w przypadku każdej ewaluacji oddzielnie. Musi kierować się własnym sumieniem, celami ewaluacji, realiami politycznymi oraz względami etycznymi. Przy tym wszystkim musi mieć na uwadze etyczne standardy ewaluacyjne wypracowane przez Połączony Komitet do Spraw Standardów Ewaluacji Edukacyjnej (Joint Committee on Standards for Educational Evaluation, 1981, s. 81, 86, 20):

- Ewaluacje powinny być planowane i realizowane tak, aby wszelkie prawa i dobra osobiste były respektowane i chronione.
- Ewaluatorzy powinni respektować ludzką godność we wszystkich swoich interakcjach z innymi osobami uczestniczącymi w ewaluacji.
- Wnioski wynikające z ewaluacji powinny być przedstawiane klientom lub innym uprawnionym do tego adresatom tak, aby mogli oni wnioski te oszacować i wykorzystać.

Czas trwania obserwacji

Czas przeznaczony na zebranie danych jest kolejnym ważnym wymiarem różnicującym badania metodą obserwacji. W badaniach terenowych wyrosłych z tradycji antropologicznej przyjmuje się, że uczestnik-obserwator spędzi minimum sześć miesięcy, a czasem nawet parę lat w świecie kultury będącej przedmiotem jego obserwacji. W socjologicznych badaniach subkultur poświęca się na te studia od paru miesięcy do roku. Holistyczny obraz całej kultury lub subkultury wymaga wiele czasu. W naukach społecznych celem podstawowych badań terenowych jest ukazanie zasadniczych związków i wzorów rzeczywistości

społecznej. Socjolog prowadzący takie badania ma nadzieję, iż uda mu się sformułować i zweryfikować stwierdzenia teoretyczne dotyczące funkcjonowania danej kultury i społeczeństwa.

Program ewaluacji lokuje się na przeciwnym krańcu kontinuum. Przeważnie jego celem jest zadanie znacznie skromniejsze: przygotowanie użytecznych informacji dla potrzeb przyszłych działań. Jeśli informacja będąca wynikiem ewaluacji ma być użyteczna, musi być przekazana w odpowiednim momencie. Osoby podejmujące decyzje nie mogą czekać latami, aż osoba prowadząca badania terenowe powróci z odległej wioski przywożąc notatki ze swoich obserwacji. Wielu ewaluatorów prowadzących badania znajduje się pod silną presją upływającego czasu i ograniczonych środków na te badania. Po wytypowaniu danych obserwacyjnych potrzebnych dla danej ewaluacji, czas trwania obserwacji będzie zależał w dużym stopniu od czasu i środków niezbędnych do spełnienie oczekiwań głównych organizatorów programu. O ile jedne projekty ewaluacji mogą wymagać zbierania szczegółowych informacji latami, przez cały czas trwania analizowanego programu, o tyle inne studia ewaluacyjne mogą zadowolić się jedno- lub dwugodzinną obserwacją jego wybranych komponentów. Dla ustalenia po prostu różnic w poziomie realizacji programu w różnych miejscowościach wystarczą krótkie wizje lokalne tam, gdzie program jest realizowany.

Czas trwania obserwacji powinien zależeć od charakteru pytań badawczych, na które ewaluator stara się odpowiedzieć, a nie od wyobrażenia, ile czasu potrzeba na przeprowadzenie typowej obserwacji uczestniczącej. Czasami dla potrzeb niektórych programów ewaluacyjnych niezbędny jest długi okres obserwacji. W innych przypadkach, kiedy oczekuje się od wykwalifikowanego ewaluatora sformułowania bardzo syntetycznych wniosków, może wystarczyć pobieżne, godzinne „przyjrzenie się" temu, jak program jest realizowany. Oczywiście będzie to miało wpływ na wnikliwość, szczegółowość i dokładność zgromadzonych danych.

Zasięg obserwacji

Starałem się pokazać, że obserwacje mogą się różnić między sobą na cztery sposoby: stopniem uczestnictwa obserwatora w ewaluowanym programie; stopniem utajnienia badań; stopniem publicznego wyeksplikowania celu ewaluacji; czasem prowadzenia obserwacji. Głównym czynnikiem wpływającym na te wymiary jest zasięg lub przedmiot prowadzonych badań. Ewaluacja może mieć zasięg szeroki lub bardzo wąski, wycinkowy.

Tradycja antropologicznych badań terenowych eksponuje ważność holistycznego ujęcia kultury. Różne podsystemy społeczne są tu rozumiane jako wzajemnie powiązane ze sobą części. Podsystemy takie, jak gospodarka, wierzenia, polityka, pokrewieństwo i inne specyficzne podsystemy społeczne, najlepiej opisywać i interpretować we wzajemnych relacjach i powiązaniach. Jednakże w praktyce obserwacje terenowe koncentrują się na poszczególnych segmentach danego społeczeństwa lub danej kultury, zależnie od zainteresowań badacza, a czas im poświęcany zależy od tego, które z tych segmentów są dla badacza najważniejsze. Tak więc konkretne studia badawcze mogą bardzo ogólnie zaprezentować daną kulturę, natomiast bardzo szczegółowo potraktować w raporcie końcowym system religijny właściwy tej kulturze.

Podobnie różny może być zasięg i przedmiot ewaluacji — bardzo szeroki lub bardzo wąski. Trzeba podjąć decyzję, co jest bardziej celowe i przydatne: pogłębiona analiza jednego lub paru pytań badawczych czy bardziej pobieżna analiza wielu problemów. Dalej, we fragmencie zatytułowanym „Co obserwować", znajduje się szczegółowe omówienie paru możliwych rozwiązań. Szerszy lub węższy zakres obserwacji zależy od środków i czasu, jakie można przeznaczyć na badania, oraz celu prowadzonej ewaluacji. To zatem, co dla danej ewaluacji jest kwestią zasadniczą, w znacznym stopniu przesądzi, które spośród rozwiązań badawczych jest i bardziej korzystne, i uzasadnione praktycznie.

Podsumowanie: wymiary różnicujące badania terenowe

Warianty obserwacji terenowych można opisać za pomocą pięciu podstawowych wymiarów. Przedstawia je w formie graficznej tabela na sąsiedniej stronie. Wykres ten może być pomocny przy wyborze parametrów konkretnej ewaluacji. Może także posłużyć do umiejscowienia danej ewaluacji na każdym z wymiarów — najpierw w fazie studiów przygotowawczych, a następnie w fazie prowadzenia obserwacji.

Co obserwować

Nie można zaobserwować wszystkiego. Człowiek obserwujący nie jest kamerą filmową, zresztą i kamera, mając ograniczone pole widzenia, musi być zwrócona w odpowiednią stronę, aby zarejestrować to, co nas interesuje. Tak więc i człowiek, i kamera filmowa muszą być skoncentrowani na przedmiocie obserwacji. W ewaluacji przedmiot obserwacji wyznaczają pytania stawiane przez ewaluatora. Przed wyruszeniem „w teren" ewaluator musi tak uporządkować sobie obszar rzeczywi-

Pięć wymiarów różnicujących badania terenowe

I. Rola obserwatora-ewaluatora

obserwacja w pełni uczestnicząca	obserwacja z udziałem do pewnego stopnia	obserwacja z zewnątrz, z perspektywy widza

II. Ewaluator w oczach innych

obserwacja jawna: organizatorzy i uczestnicy programu wiedzą, że są obserwowani i wiedzą, *kto* jest obserwatorem	rola obserwatora jest znana niektórym, lecz nie wszystkim	obserwacja niejawna: organizatorzy i uczestnicy programu nie wiedzą, że jest prowadzona obserwacja albo że ktoś ich obserwuje

III. Wyjawienie celu ewaluacji innym

wyjawianie rzeczywistego celu ewaluacji wszystkim	wyjawienia częściowe	zatajenie celu przed wszystkimi organizatorami i uczestnikami	wprowadzenie w błąd organizatorów i/lub uczestników co do celu ewaluacji

IV. Czas trwania obserwacji ewaluacyjnej

obserwacja jednorazowa o ograniczonym czasie trwania (np. jedna godzina)	obserwacja wielokrotna lub długotrwała (np. miesiące lub lata)

V. Zasięg obserwacji

ograniczony pojedynczy element lub komponent programu	szeroki holistyczny obraz całego programu i wszystkich jego elementów

stości określonej przez program, aby jego obserwacje były wykonalne i odpowiadały założonym celom.

Postaram się omówić parę takich obszarów, na których mogą ogniskować się obserwacje. Wybór ten może uwrażliwić obserwatora na te informacje, które mogą mu być potrzebne dla adekwatnego opisania programu i jego uczestników.

Doświadczeni obserwatorzy wiedzą, że pewne typy działań i zdarzeń są źródłem informacji szczególnie użytecznych. Strategia, która kładzie szczególny nacisk na pewnego typu obserwacje, często wymaga sięgnięcia po coś, co w metodologii jakościowej nazywa się pojęciami

uwrażliwiającymi (*sensitizing concepts*). Zamiast narzuconych z góry kategorii i zoperacjonalizowanych zmiennych, „uważliwiające pojęcia" wyznaczają jedynie ogólne ramy wskazujące na ważność pewnych typów zdarzeń, działań i zachowań. „Bycie w grupie" jest takim pojęciem uwrażliwiającym. Pokrewieństwo, przywództwo, doświadczenia socjalizacyjne, władza i inne podobne pojęcia mają analogiczny uwrażliwiający charakter, sygnalizują nam bowiem pewien typ doświadczeń i sterują decyzjami dotyczącymi tego, co należy obserwować. Dalej zajmiemy się pojęciami uwrażliwiającymi w terenowych badaniach ewaluacyjnych.

Opis środowiska, w którym program jest realizowany

Opis obserwowanego środowiska powinien być dostatecznie szczegółowy, by czytelnik mógł je sobie wyobrazić. W tym celu warto zwłaszcza cytować wypowiedzi uczestników programu dotyczące ich reakcji i spostrzeżeń odnośnie do otoczenia. Pomocne mogą być także fotografie, pod warunkiem, że zrobiono je w sposób dyskretny.

Dobrym ćwiczeniem dla początkujących obserwatorów jest sporządzenie opisu sytuacyjnego i udostępnienie go osobie postronnej z pytaniem, czy jest w stanie wyobrazić sobie to, co zostało zawarte w opisie. Równie pomocne może być sporządzenie opisu przez dwie osoby, które następnie wspólnie przeanalizują swoje opisy, zwracając uwagę na przymiotniki interpretujące lub abstrakcyjne (np. środowisko „sprzyjające") użyte zamiast opisowych (np. śródmiejskie biuro sprzedaży). Pisanie w sposób opisowy polega na dostarczaniu czytelnikom raportu wyczerpujących informacji. Wyczerpujących to znaczy takich, które nie każą im domyślać się, co obserwator miał na myśli, lub szukać dodatkowych informacji pozwalających zinterpretować otrzymany opis. Na przykład wyrażenie „zatłoczony pokój" jest interpretacją. Opis powinien brzmieć: „Pokój był wystarczająco duży, aby mieściła się w nim kanapa na trzy osoby pod jedną ścianą, sześć krzeseł pod ścianą obok i trzy krzesła naprzeciwko kanapy, obok drzwi. W pokoju było dość miejsca dla dwudziestu osób, ale kiedy wszystkie stały, to nieomal ocierały się o siebie. Niektórzy z obecnych narzekali, że ten pokój jest za ciasny".

Fizyczne otoczenie sytuacji społecznej może być niezwykle ważne dla tego, co się dzieje. Częstym błędem popełnianym przez obserwatorów zajmujących się ewaluacją jest przyjmowanie otoczenia za oczywiste. W raportach końcowych możemy natrafić na stwierdzenia, że program był realizowany „w szkole". Obserwator może mieć wyobrażenie „szkoły", które odpowiada temu, co obserwował, lecz trzeba pamiętać, że szkoły mogą się różnić wielkością, wyglądem, tym, co je

otacza, a przede wszystkim tym, jakie są wewnątrz. To samo można powiedzieć o otoczeniu odnośnie do przestępczości, zdrowotności, lokalnych programów ochrony zdrowia psychicznego i wszystkich innych sytuacji społecznych.

Społeczne środowisko człowieka

Tak jak różne jest otoczenie fizyczne, tak różne może być środowisko społeczne, w którym dany program jest realizowany. Sposób, w jaki ludzie kontaktują się ze sobą, wytwarza środowiskową konstelację społeczną, która oddziałuje na ich zachowania względem siebie. Opisując środowisko społeczne obserwator powinien zwracać uwagę na sposób, w jaki ludzie organizują się w grupy i podgrupy. Wzory interakcji, ich częstotliwość, ukierunkowanie wzorów komunikacji, a także wszelkie zmiany tych wzorów ujawniają stan rzeczy ich środowiska społecznego. Charakterystyki osób wchodzących w różne układy grupowe, interakcje między kobietami i mężczyznami, między osobami należącymi do różnych warstw lub ras oraz różnych grup wiekowych, sygnalizują obserwatorowi wzory funkcjonujące w środowisku społecznym danego programu.

Wzory podejmowania decyzji w danym środowisku społecznym mogą być przedmiotem szczególnego zainteresowania. Kto podejmuje decyzję co do przyszłych działań? Jak te decyzje są komunikowane?

Tak jak w przypadku otoczenia fizycznego, tak i tu ważne jest, aby obserwator nie mylił opisu tego, co się zdarzyło, z relacją o reakcjach i opiniach osób uczestniczących w tym, co się zdarzyło. Opis środowiska społecznego sporządzony przez obserwatora nie musi pokrywać się ze spostrzeżeniami osób uczestniczących w tym środowisku. Przeciwnie, byłoby czymś niezwykłym, gdyby wszyscy postrzegali dane środowisko społeczne w ten sam sposób. Spostrzeżenia osób uczestniczących zawsze powinny być podane w formie cytatów z powołaniem się na źródła informacji, tak aby oddzielić je od opisu obserwatora i jego własnej interpretacji sytuacji.

Działania programowe a zachowania osób uczestniczących w programie

Ośrodkiem zainteresowania większości obserwacji są działania podejmowane w ramach programu oraz zachowania osób w nim uczestniczących. Co uczestnicy programu naprawdę w nim robią? Jakie są ich oczekiwania? Co oznacza „bycie uczestnikiem" programu w obser-

wowanym otoczeniu? Jakie rodzaje aktywności mają w nim miejsce? Są to przykłady pytań, jakie stawia sobie ewaluator rozpoczynający obserwację realizacji programu.

Przy opisywaniu działań podejmowanych w ramach programu zazwyczaj niezbędne jest pogrupowanie działań w pewne sekwencje. Działania zwykle mają swoją chronologię, swój początek, rozwinięcie i zakończenie. Obserwując program, ewaluator zwraca uwagę właśnie na te następujące po sobie człony podejmowanych działań. Co więcej, sekwencje te są zazwyczaj określone i nazwane albo przez organizatorów programu, albo przez jego uczestników; czasem jednak ewaluator może uchwycić sekwencje zdarzeń dające się wyodrębnić w ramach większych całości. Mogą nimi być na przykład: lekcja, zebranie rady pedagogicznej, przerwa obiadowa, różnego rodzaju spotkania, wizyty domowe, wykraczające poza program, lub konsultacje. Podsumowując, są to wszelkie formalne działania programowe.

Ewaluator, obserwując interakcje i działania, dąży do wyrobienia sobie ogólnego poglądu na to, co dzieje się w ramach obserwowanego przez niego programu. Aby opis interakcji był wyczerpujący, informacja o pełnej sekwencji wydarzeń musi zawierać odpowiedzi na następujące pytania: Jak działania i interakcje się rozpoczynały? Kto był obecny na początku? Co dokładnie powiedziano, gdy rozpoczynano dany program? Jak uczestnicy programu zareagowali na to, co zostało im powiedziane, lub jak skomentowali to, co zostało im powiedziane? Są to przykłady pytań związanych ze wstępną fazą działań podejmowanych przy rozpoczęciu danego programu.

Tego samego rodzaju pytania wymagają odpowiedzi w kolejnej fazie, przy obserwacji pełnej sekwencji wydarzeń: Kto jest w nie zaangażowany? Kto co powiedział? Co się zdarzyło? Co robili uczestnicy programu? Jakie występują różnice w zaangażowaniu uczestników w realizację programu? (Zanotowane przez „uczestniczącego obserwatora" jego własne spostrzeżenia są częścią zebranych przez niego danych obserwacyjnych). Jak zmieniały się zachowania i interakcje między uczestnikami programu w trakcie realizacji podejmowanych działań?

Wreszcie obserwacja zakończenia programu. Jakie są sygnały tego, że działania i interakcje zmierzają do końca? Kto jest obecny przy jego zakończeniu? Co kto powiedział? Jak uczestnicy zareagowali na zakończenie programu? Jak się ma zakończenie całości działań danego programu do innych działań i interakcji między jego uczestnikami?

Dla celów obserwacji wskazane jest traktowanie poszczególnych sekwencji działań jako niezależnych wydarzeń. Analiza będzie polegać

na poszukiwaniu wzorów łączących te niezależne od siebie człony. We wstępnych fazach obserwacji ewaluator będzie więc zajęty próbami uchwycenia względnie niezależnych obiektów obserwacji, wzory zaś i wzajemne realacje między poszczególnymi sekwencjami działań wyłonią się w następstwie późniejszej analizy.

Fragment ten dotyczył obserwowania formalnych działań podejmowanych w ramach programu, wyznaczających jego początek i wszystkie inne znaczące fazy jego realizacji aż do zamknięcia programu. Formalne działania składają się na zasadniczą treść każdego programu, gdyż każda jego realizacja polega przede wszystkim na przeprowadzeniu serii działań wcześniej zaplanowanych w sposób formalny. Jednakże obserwator zajmujący się ewaluacją, dążący do zrozumienia programu i określenia, jak miał on wpływ na jego uczestników, nie może poprzestać na obserwacji tych rodzajów aktywności, które były zaplanowane w sposób formalny. Dalej zajmiemy się obserwacją tego, co znajduje się „poza" sformalizowanymi działaniami programowymi.

Nieformalne interakcje i nie planowane działania

Ewaluatorzy, którzy nie wyjdą z obserwacją poza sformalizowaną aktywność programową, mogą przegapić wiele danych. Niektóre programy są tak skonstruowane, że w „czasie wolnym", czy po prostu nie zagospodarowanym, pozwalają uczestnikom przemyśleć to, co miało miejsce w czasie sformalizowanych zajęć programowych, i zapewniają (uczestnikom i organizatorom) niezbędne do tego pomieszczenie. Inne programy ograniczają ilość czasu nie zagospodarowanego, ale rzadko który program lub instytucja jest w stanie zaplanować każdą chwilę swoim uczestnikom.

Zorganizowanie obserwacji takich nieformalnych interakcji i nie zaplanowanych działań może być szczególne trudne. Dzieje się tak dlatego, iż ludzie w takich sytuacjach są w ruchu, wchodzą i wychodzą, łączą się w grupki, które zaraz się rozpraszają, czasem siedzą samotnie, czasem coś piszą, starają się odpocząć, a wszyscy na różny sposób zaangażowani są w działania, które wydawać się mogą przypadkowe. Jak więc ewaluator-obserwator ma gromadzić dane w takich sytuacjach?

Na tym przykładzie widać, jak ważne jest, by być otwartym na dane pojawiające się niejako spontanicznie. Nie można przewidzieć, jakie nowe fakty ujawni obserwacja nie zaplanowanych działań. Nie spiesząc się z interpretacją czy też z przypisywaniem znaczenia zachowaniom, których jest właśnie świadkiem, obserwator powinien po prostu kon-

tynować zbieranie danych opisujących, co ludzie robią, a zwłaszcza co mówią.

Ostatnia uwaga jest szczególnie ważna. W czasie nie zorganizowanym przez program jego uczestnicy mają doskonałą okazję do wymiany opinii i porozmawiania z innymi o własnych doświadczeniach związanych z tym programem. Niekiedy wystarczy, by obserwator przysłuchiwał się tym rozmowom. Kiedy indziej odpowiednią strategią badawczą będą nieformalne wywiady — z pojedynczymi osobami w formie niezobowiązującej rozmowy albo z niewielką grupą osób. W takich przypadkach ewaluator zadaje normalne, konwersacyjne pytania: „Tak więc, co pan sądzi o tym, co zdarzyło się dzisiaj rano"? „Czy było dla pana jasne, o czym oni dzisiaj mówili"? „Jak pan myśli, czy to, co dzisiaj robiliśmy, pasuje do całości naszych zadań"?

Równocześnie nie wolno zapominać, że wszystko, co dzieje się w ramach i obok programu, może być źródłem informacji. Znaczenia, wzory, istotność danych może się zmienić, ale ludzie cały czas coś robią, coś przeżywają, a program trwa. Fakt, iż żaden z uczestników nie mówi o zajęciach po ich zakończeniu — to informacja. Fakt, że ludzie natychmiast po zakończeniu spotkania rozchodzą się do swoich zajęć — to informacja. Fakt, że uczestnicy plotkują i rozmawiają o sprawach osobistych, w żaden sposób nie związanych z programem, w którym biorą udział — to informacja.

Nie jest wcale czymś niezwykłym, że najważniejsi uczestnicy wielu programów uczą się czegoś nowego właśnie w czasie nie zagospodarowanym przez organizatorów. Osobisty kontakt z innym uczestnikiem programu może być zdarzeniem najważniejszym. Aby uzyskać holistyczną wizję programu, ewaluator zajmujący się obserwacją musi zachować czujność w czasie tych nieformalnych i nie zaplanowanych przez organizatorów działań.

Obserwacja uczestnicząca jest zawsze kombinacją obserwacji i nieformalnych wywiadów. Jest niezwykle ważne, aby ewaluator nie przyjmował wstępnych założeń co do znaczenia tego, co obserwuje, nie uwzględniających tej perspektywy, w jakiej osoby uczestniczące w danym programie postrzegają własne zachowanie i doświadczenia.

Język, jakim posługują się uczestnicy programu

Jednym z aksjomatów antropologii jest to, że nie zrozumie się innej kultury, nie poznawszy języka, jakim posługują się ludzie w tej kulturze żyjący. Język jest sposobem organizacji świata. Nazwy, które ludzie nadają rzeczom, mówią nam, co jest ważne dla ich kultury

lub dla ich grupy. Tak więc Eskimosi mają wiele słów na określenie śniegu, a Arabowie wiele słów na określenie wielbłąda. Podobnie artysta posługuje się wieloma określeniami czerwieni i różnego rodzaju pędzli, a ewaluator wieloma terminami opisującymi różne rodzaje ewaluacji.

To samo zjawisko można zaobserwować w programach socjalnych i edukacyjnych. Osoby zajmujące się realizacją programów z różnych dziedzin wypracowały swój własny język dla opisu problemów, z którymi mają do czynienia. Ci, którzy pracują z osobami opóźnionymi w rozwoju, posługują się rozbudowaną terminologią opisującą różne typy opóźnienia umysłowego. Ci, którzy mają do czynienia z przestępczością, wypracowali własny język różnicujący typy napastników i rodzaje zagrożenia. Do zadań obserwatora danego programu należy nauczenie się języka, jakim posługują się jego uczestnicy. Oznacza to poznanie nie tylko dosłownego znaczenia używanych słów, ale także konotacji i całej symboliki związanej z ich słownictwem. Nie jest niczym niezwykłym, że uczestnicy programu tworzą sobie własny język opisujący poszczególne aspekty programu i własne doświadczenia. Sporządzone przez obserwatora notatki powinny wiernie oddawać język, jakim posługują się uczestnicy programu w odniesieniu do własnych doświadczeń, tak aby jego reguły mogły wyjść na jaw w analizie, a osoba czytająca raport końcowy mogła poznać niuanse języka używanego przez uczestników danego programu.

Dosłowne cytowanie wyrażeń używanych przez uczestników programu jest sposobem zrelacjonowania ich rozumienia doświadczenia, które stało się ich udziałem. Obserwator musi nauczyć się języka uczestników programu i zachować go w swoim raporcie wraz ze wszystkimi regułami użycia, aby przedstawić uczestników i ich świat za pomocą takich terminów i takich pojęć, jakimi oni sami posługują się do jego opisu.

Komunikacja niewerbalna

Badacze reprezentujący różne dyscypliny nauk społecznych już dawno dostrzegli znaczenie komunikacji werbalnej i niewerbalnej dla funkcjonowania grup ludzkich. Notując język charakterystyczny dla uczestników danego programu, ewaluator obserwuje także różne formy komunikacji niewerbalnej między nimi. W sytuacjach związanych np. z edukacją komunikacja niewerbalna może polegać na zachowaniach wyrażających stosunek słuchaczy do prowadzącego zajęcia. W czasie spotkań grupowych liczne objawy niepokoju, kręcenie się i szukanie wygodnej pozycji mogą świadczyć o tym, że uczestniczący zaabsorbowani

są przede wszystkim tym, co dzieje się wokół nich. Sposób ubierania się, objawy serdeczności, zachowywanie fizycznego dystansu względem innych uczestników dyskusji, sposób zajmowania miejsc w sali, gdzie odbywają się zajęcia — wszystko to są sygnały niewerbalne.

Własne odczucia ewaluatora i reakcje na sygnały niewerbalne są istotną częścią danych obserwacyjnych. Obserwator powinien notować własne reakcje na zachowania innych. Często warto również, łącząc obserwacje z wywiadem, zapytać uczestników spotkania o ich zachowania niewerbalne i reakcje na zachowanie niewerbalne innych osób. Śledząc wzory zachowań i opisując, co ludzie robią w różnych sytuacjach, obserwator będzie w stanie wyodrębnić zachowania niewerbalne znaczące dla konkretnych sytuacji.

Pomiar dyskretny

Ewaluacja wywołuje uczucie zaniepokojenia wśród wielu zarówno organizatorów, jak i uczestników programu. Bez względu na to, jak ostrożnie i delikatnie ewaluacja jest prowadzona, zawsze istnieje możliwość, że osoby, które o niej wiedzą, będą zachowywać się inaczej niż ci, którzy o niej nie wiedzą. W znanej książce *Unobtrusive Measures: Non-reactive Research in the Social Sciences* (Webb i współaut., 1966) czytamy:

„Świadomość, że się jest przedmiotem badania, nawet przy pozytywnym nastawieniu i chęci współpracy, może powodować zniekształcenie wyników. [...] Trzeba od razu zaznaczyć, że świadomość bycia pod kontrolą sama w sobie nie wpływa ujemnie na reakcje, ale prawdopodobieństwo pewnych odchyleń występuje we wszystkich badaniach, w których respondenci są świadomi bycia przedmiotem badania" (s. 13).

To spostrzeżenie dotyczące reakcji osób poddanych obserwacji doprowadziło do tego, iż niektórzy reprezentanci nauk społecznych rekomendują nieujawnianie prowadzenia badań (jak już była o tym mowa). Jeśli uczestnicy nie są świadomi, że zbiera się dane i że ktoś ich obserwuje, to ich reakcje mniej zagrażają trafności ewaluacji. Ale i wtedy, gdy uczestnicy są świadomi „bycia przedmiotem obserwacji", istnieje możliwość tzw. pomiarów dyskretnych. Pomiar dyskretny to pomiar dokonywany bez wiedzy obserwowanych osób.

Na przykład Robert L. Wolf i Barbara Tymitz (1977) posłużyli się dyskretnym pomiarem w toku badań prowadzonych w Muzeum Historii Naturalnej w Smithsonian Institute. Poszukiwali „śladów zużycia" jako wskaźników frekwencji w poszczególnych salach wystawowych.

Badacze ci przyjęli, że stopień wytarcia dywanów będzie wskaźnikiem popularności, jaką cieszą się różne działy tego muzeum.

Pomysłowy ewaluator może dowiedzieć się wiele o sytuacji poszukując materialnych śladów aktywności. Kurz na przedmiotach świadczy o tym, że nikt się nimi nie posługuje. Miejsca na terenie szkoły często uczęszczane przez dzieci będą wyglądały inaczej, będą bardziej zniszczone niż te, w których dzieci przebywają rzadko. Rozkład topograficzny programu zasygnalizuje ważne wzory interakcji i znaczące różnice statusów. Na przykład różnice w wyposażeniu wnętrza biurowego, lokalizacja biur i specjalne ułatwienia w dostępie do nich mogą ujawniać ważne zależności i pozycje różnych osób.

Poszukiwanie dyskretnych metod pomiaru dowodzi, że badania terenowe nie są postępowaniem rutynowym. Pomysłowy ewaluator, świadomy różnorodności sytuacji, może czerpać informacje studiując rozmieszczenie fizycznych i społecznych elementów otoczenia, może dowiedzieć się wiele dzięki dyskretnym wskaźnikom ludzkiej aktywności.

Dokumentacja programu

Jak już zostało powiedziane, obserwacja w terenie polega na patrzeniu i słuchaniu, obserwowaniu i przeprowadzaniu wywiadów. Celem badań terenowych jest dowiedzenie się możliwie jak najwięcej o wszystkim, co dzieje się w ramach programu. W wielu przypadkach kolejnym bogatym źródłem informacji są dokumenty i wszelkie materiały pisane związane z programem. Charakter tych materiałów i dokumentów będzie różny zależnie od programu, lecz we współczesnym społeczeństwie za każdym programem ciągnie się sznur papierów, które ewaluator może wykorzystać, by więcej wiedzieć i lepiej rozumieć, o co w programie chodzi.

Bardzo ważne jest, aby przystępując do ewaluacji zapewnić sobie dostęp do najrozmaitszych dokumentów i materiałów związanych z programem. Sytuacją idealną byłby dostęp do wszystkich notatek związanych z klientami programu, do całej korespondencji jego organizatorów, do sprawozdań finansowych, ustaleń organizacyjnych, regulacji prawnych, wykazów, wykresów i wszelkich innych oficjalnych i nieoficjalnych dokumentów powstałych w ramach lub dla potrzeb danego programu. Taka dokumentacja dostarcza ewaluatorowi informacji o wielu sprawach niedostępnych obserwacji: wydarzeniach, które miały miejsce przed rozpoczęciem programu, sprawach osobistych, do których ewaluator nie jest bezpośrednio dopuszczony, bądź organiza-

cyjnych, które pięknie przedstawiane w dokumentach oficjalnych nie są jednak realizowane w praktyce.

Dokumentacja programu dostarcza informacji cennych ze względu na ich źródłowy charakter, ale również dlatego, że może rodzić pytania, na które nie uzyska się odpowiedzi inaczej niż tylko za pomocą obserwacji i wywiadów. Tak więc materiały i dokumenty służą jako: (1) podstawowe źródło informacji o przebiegu programu; (2) źródło inspiracji badawczej, to znaczy pytań, na które ewaluator będzie poszukiwać odpowiedzi za pomocą obserwacji i wywiadów socjologicznych.

Jak w przypadku wszystkich informacji, do których ewaluator ma dostęp, musi być zachowana również poufność dokumentów, zwłaszcza dotyczących klienta, czyli osoby lub grupy osób zamawiających dany program i jego ewaluację. Biorąc za punkt wyjścia te dokumenty analizowanego programu, które zostały uznane za ogólnie dostępne i które mogą być opublikowane bez naruszenia poufności, należy wcześniej ustalić, w jakim zakresie ewaluator może powoływać się na dokumentację programu i cytować ją w raporcie końcowym. Dokumentacja związana z funkcjonowaniem współczesnych organizacji jest jednym ze źródeł informacji. Ewaluator prowadzący badania musi umieć z nich korzystać, studiować je i rozumieć.

Obserwowanie tego, co się nie zdarzyło

Dotychczas zajmowaliśmy się sprawami, które można obserwować. Obserwowanie różnych form aktywności, tego, co zdarzyło się w grupie, co zostało powiedziane i zrobione, jak przebiegały interakcje, a wreszcie istotnych cech otoczenia — wszystko to jest ważne przy wszechstronnym podejściu do badań terenowych. Jednakże równie ważne jest obserwowanie tego, *co się nie zdarzyło*. Zadanie to wymaga pewnej zręczności. Podejmując się obserwacji tego, co się nie zdarzyło, trzeba zdawać sobie sprawę, że obiektów, które można zakwalifikować jako „nie występujące", jest niemal nieskończenie wiele.

Można wymienić dwie sytuacje, w których wskazane jest zwrócenie uwagi na coś, czego brak. Po pierwsze, jeśli ewaluator wie, że pewne rzeczy powinny się zdarzyć lub oczekuje, że będą miały miejsce — to powinien zaznaczyć, że nie nastąpiły. Jeśli zgodnie z założeniami programu szkolnego dzieci powinny mieć możliwość zapoznania się ze środowiskiem społecznym, a nie zostały z nim zaznajomione, to obserwator powinien zanotować, że takie działania nie zostały podjęte. Jeśli ewaluator ograniczy się w swoim raporcie do tego, co zaobserwował, osoba czytająca ów raport może zastanawiać się, czy faktycznie nie po-

dejmowano żadnych innych działań poza zaobserwowanymi. Podobnie, jeśli program związany z resocjalizacją nieletnich przestępców przewiduje zapewnienie bezpośredniego dostępu do własnego doradcy, lecz w praktyce tak się nie dzieje, to ewaluator powinien wykazać w swoim raporcie brak tego typu kontaktów.

Druga sytuacja uzasadniająca wskazanie, że coś się *nie zdarzyło*, ma miejsce wtedy, gdy ewaluator — bazując na swoim doświadczeniu i zdrowym rozsądku — dojdzie do wniosku, że brak jakiegoś rodzaju aktywności jest wart odnotowania. Przykładowo, jeśli obserwator stwierdza trwałą harmonię i konsens między członkami grupy, to warto, aby odnotował, że, w interakcjach wewnątrzgrupowych nie zaobserwowano żadnych konfliktów ani animozji. Powiadomienie o tym czytelnika raportu jest istotne z tej przyczyny, że informuje go, iż nie selekcjonowano danych z zamiarem przedstawienia tylko tych, które świadczą o harmonii i zgodności. Brak osobistych animozji między ludźmi w grupie, w której potencjalnie istnieje zarzewie konfliktu, zasługuje na uwagę.

W wielu tego typu przypadkach obserwacja dotycząca tego, co się nie zdarzyło, jest po prostu innym sposobem zrelacjonowania tego, co się rzeczywiście zdarzyło. Taki sposób wypowiedzi może bardziej przyciągnąć uwagę czytelnika niż rutynowe sposoby obserwacji. Podajmy inny przykład: Jeżeli obserwowany program jest realizowany w społeczności wieloetnicznej, to być może jednym z jego celów jest uwrażliwienie części organizatorów na potrzeby, interesy i specyfikę kulturową mniejszości. Jednocześnie jednak założenia programu nie określają pożądanego składu etnicznego jego organizatorów. Jeśli więc ewaluator zaobserwuje, że zespół organizatorów programu składał się wyłącznie z „białych", to może poinformować o tym na dwa sposoby: (1) pisząc, że personel stanowili wyłącznie biali albo (2) zaznaczając, że „wśród organizatorów programu nie było przedstawicieli mniejszości". Ten drugi sposób polega na powiadomieniu, że czegoś nie zaobserwowano, że zaobserwowano coś, czego nie ma.

Podobnie, jeśli w założeniach programowych nie przewidziano, że sugestie jego uczestników mogą wpływać na jego realizację, to być może warto, aby ewaluator zaznaczył brak jakiejś formy oddziaływania uczestników na przebieg programu, co może być ważne dla jego kontynuacji lub przy planowaniu innych zamierzeń tego typu.

Proces zbierania danych

Notatki terenowe

Prowadzenie notatek terenowych wymaga pewnych wstępnych ustaleń natury technicznej, dotyczących: techniki zapisu; czasu i miejsca robienia notatek; opracowania przez obserwatora szyfru jako indywidualnej metody stenografii; sposobu archiwizacji notatek. Nie ma jakiejś uniwersalnej recepty odnośnie do techniki i procedur sporządzania notatek terenowych, ponieważ różne warunki narzucają różne sposoby postępowania, a szczegóły organizacji badań terenowych wiążą się przede wszystkim z indywidualnym stylem pracy osoby prowadzącej obserwacje. *Obowiązkowe jest tylko jedno: prowadzenie notatek!*

„Przy wszystkich innych obowiązkach — podstawowym zadaniem obserwatora jest robienie notatek, od nich bowiem zależy możliwość późniejszej analizy jakościowej. Notatki terenowe to *raison d'etre* obserwatora. Jeśli ich nie robi, to równie dobrze mogłoby go tu w ogóle nie być" (Lofland, 1971, s. 102).

Notatki terenowe to opis tego, co zostało zaobserwowane. Powinny zawierać to wszystko, co obserwatorowi wydaje się istotne do zanotowania. Nie można liczyć, że cokolwiek uda się odtworzyć później. Kiedy sytuacja zmienia się, stale dzieje się coś nowego, piszącego kusi nadzieja, że szczegóły czy jakieś detale sytuacyjne uda mu się zapamiętać i opisać później. Jeśli jednak chodzi o coś ważnego dla ciebie jako obserwatora — o informację pomocną w zrozumieniu kontekstu, programu, czegoś, co się dzieje, słowem, czegokolwiek — to najszybciej jak można powinieneś wciągnąć tę informację do notesu.

Co najważniejsze: notatki z pracy w terenie są opisem. Powinny zawierać podstawowe informacje, takie jak: data i miejsce przeprowadzenia obserwacji, kto był obecny, jaki rodzaj działań był podejmowany, w jakim otoczeniu realizowano dany program, jakie interakcje społeczne miały miejsce oraz inne informacje opisowe, które później umożliwią obserwatorowi odtworzenie w pamięci całej sytuacji. Dzięki temu wszystkiemu, co znajdzie się w notesie ewaluatora, czytelnik raportu końcowego będzie miał możliwość „doświadczyć" zaobserwowanych przez niego wydarzeń.

Niżej podane przykłady ilustrują różne rodzaje opisowych notatek terenowych. Po lewej stronie podano przykład zapisków niejasnych i nazbyt ogólnych, po prawej — sporządzonych na podstawie tych samych obserwacji notatek szczegółowych i konkretnych.

Notatki niejasne i zbyt ogólne	Notatki szczegółowe i konkretne
(1) Nowa klientka oczekiwała z niepokojem na rozpoczęcie wywiadu.	(1) Najpierw klientka usiadła bardzo usztywniona na krześle obok recepcji. Wzięła do ręki jakieś pismo i nerwowo przebierając palcami przerzucała strony nie zatrzymując uwagi na żadnej. Odłożyła pismo, spojrzała na zegarek, poprawiła sukienkę i sięgnęła ponownie po to samo pismo. Nawet go nie otworzyła, wyciągnęła z torby paczkę papierosów i zaciągnęła się dymem. Co chwila spoglądała z ukosa na recepcjonistkę, to znów na czasopismo, czasem zerkała na pozostałe dwie czy trzy osoby czekające w tym samym pokoju. Wodziła oczami od czasopisma do papierosa, od recepcjonistki do pozostałych czekających osób. Jednakże unikała z nimi kontaktu wzrokowego. Kiedy wywołano jej nazwisko, kobieta podskoczyła jakby ktoś ją przestraszył.
(2) Klientka była wrogo usposobiona wobec personelu.	(2) Kiedy ktoś z personelu powiedział jej, że nie będzie mogła zrobić tego, co chce, zaczęła krzyczeć. Wykrzykiwała, że nie ma wpływu na to, co dzieje się w jej życiu, że wszyscy mają ją za nic, że czuje się jak „nakręcana", że „sra już tym wszystkim" i wszyscy mogą „iść do diabła". Na koniec, grożąc całemu personelowi zaciśniętą pięścią, wybiegła, zostawiwszy swojego rozmówcę na środku pokoju z otwartymi ustami — stał, jakby go zamurowało.
(3) Następny student, który przyszedł na test, był ubrany bardzo niechlujnie.	(3) Następny student, który wszedł do pokoju, był ubrany zupełnie inaczej niż trzy poprzednie osoby. Strój tamtych można określić jako oficjalny, odpowiedni do biura. Ich włosy były starannie uczesane, ubrania czyste i wyprasowane, o odpowiednio dobranych kolorach i w dobrym gatunku. Nowy przybysz miał na sobie brudne spodnie, dziurawe i połatane na kolanach, wytarte na siedzeniu; jego flanelowa koszula była pognieciona. Miał dwa krawaty, jeden pod szyją, drugi niedbale przyczepiony do paska u spodni. Jego włosy były w całkowitym nieładzie, ręce zaś miał tak brudne, jakby właśnie skończył reperować silnik samochodu.

Przykłady te ilustrują posługiwanie się sformułowaniami ogólnymi do opisu konkretnych działań i sytuacji. Słowa takie, jak „biedny", „zły", „niełatwy", nie są opisowe. Takie interpretujące pojęcia, jak przytoczone wyżej, raczej wprowadzają w błąd, niż ujawniają szczegóły opisywanej sytuacji. Niewiele mówią osobie, która nie była świadkiem danego zdarzenia. Co więcej, użycie takich określeń w notatkach terenowych, bez towarzyszącego im szczegółowego opisu, świadczy, że badacz koncentrował się na interpretacji, a nie na opisie. Pojęcia tego typu mogą coś objaśniać jedynie przez porównanie. Aby opisać kogoś

jako „źle" ubranego, konieczny jest punkt odniesienia, to znaczy określenie, jak wygląda ktoś „dobrze" ubrany. Żadna inna umiejętność nie jest tak ważna w badaniach terenowych, jak właśnie ta — umiejętność relacjonowania zdarzeń w sposób opisowy, konkretny i szczegółowy. Jeszcze raz trzeba to zaznaczyć: notatki terenowe muszą być sporządzone w sposób opisowy.

Druga sprawa: notatki terenowe muszą zdawać relację z tego, o czym ludzie mówili. Powinny zawierać cytaty lub możliwie wiernie streszczenie ich wypowiedzi. Dotyczy to zarówno wypowiedzi w trakcie różnych działań czy zajęć, jak i w czasie wywiadów lub nieformalnych rozmów.

Trzecia sprawa: notatki terenowe powinny zawierać własne odczucia i spostrzeżenia ewaluatora, jego reakcje związane z tym, czego doświadczył podczas obserwacji, oraz jego refleksje dotyczące znaczenia i istotności tego, co zaobserwował. Trzeba jeszcze raz powtórzyć: jest to niezwykle ważne, aby nie oszukiwać samego siebie, iż można dokładnie przypomnieć sobie własne odczucia po przeczytaniu opisu wydarzeń. Dlatego trzeba niezwłocznie i skrupulatnie notować własne odczucia i spostrzeżenia. Ich rodzaj, charakter i intensywność. Przy badaniach sytuacyjnych własne doświadczenia ewaluatora są kluczowym elementem zebranych materiałów. Zbliżenie się do programu i do osób w nim uczestniczących poprzez badania terenowe ma m. in. na celu umożliwienie ewaluatorowi doświadczenia, jak to jest, gdy się uczestniczy w danym programie. Jeśli doświadczenia tego typu obserwator nie zarejestruje, ten ważny cel jego badania nie zostanie zrealizowany.

Na koniec, notatki terenowe powinny zawierać opinie, interpretacje, wstępne analizy oraz robocze hipotezy obserwatora dotyczące tego, co dzieje się w danym programie. Podstawowym jego obowiązkiem jest doświadczyć i opisać to, co się dzieje. Obserwator podchodzi do swojej pracy z surowym postanowieniem nieulegania powziętym z góry wyobrażeniom i pochopnym sądom, tym niemniej nie zamienia się w maszynę rejestrującą, włączoną w chwili rozpoczęcia badań. Gdy zacznie notować, pojawią się opinie, idee, inspiracje, i — jakże by inaczej — także oceny i wartościowanie.

Obserwatorzy muszą tak rozplanować i wyważyć technikę sporządzania notatek terenowych, aby nie ucierpiało na tym ani ich uczestnictwo, ani jakość prowadzonych obserwacji. Z tego względu podstawową praktyczną zasadą jest pisać niezwłocznie, sięgać po notes tak szybko, i tak często, jak jest to fizycznie i praktycznie możliwe. Sporządzanie notatek terenowych jest zadaniem wymagającym samodyscypliny oraz czasu — przynajmniej tyle samo czasu co sama obserwacja.

Korzystanie z tak zwanych informatorów kluczowych

Na każdym etapie pracy terenowej istotnym źródłem informacji mogą być ludzie. Kluczowi informatorzy są to osoby, które są w stanie udzielić szczegółowych i użytecznych informacji, osoby, które są lepiej poinformowane od innych i mogą pomóc obserwatorowi lepiej zrozumieć to, czego jest świadkiem. Prawdziwą podporą w pracy terenowej jest korzystanie z takich osób jako źródła informacji o tym, czego ewaluator-obserwator nie może bezpośrednio doświadczyć, mogą one również służyć pomocą przy interpretacji wydarzeń, których obserwator był świadkiem.

Osoby uznane za kluczowych informatorów muszą być dobrane bardzo starannie, tak aby nie dopuścić do powstania między nimi osobistych animozji i wrogości. „Pozycja" kluczowego informatora nie musi być w jakiś sposób sformalizowana. Kluczowym informatorem jest po prostu ta osoba, z którą obserwator chce spędzić więcej czasu rozmawiając o tym, co dzieje się w ramach programu. Często taki informator musi zostać przygotowany do tej roli.

Znalezienie informatorów jest szczególnie użytecznym sposobem dotarcia tam, gdzie obserwator nie ma lub nie może mieć bezpośredniego dostępu, na przykład do jakichś specyficznych podgrup działających w ramach całego programu. Niebezpieczeństwo związane z wykorzystaniem uzyskanych od nich informacji polega na tym, że ich opinie mogą być nieobiektywne, a zatem można uzyskać mylny obraz tego, co rzeczywiście miało miejsce. Ważne jest, aby informacje uzyskane od takich osób wyraźnie zaznaczyć w notatkach terenowych („informator X powiedział, że..."). Dane uzyskane od informatorów mają charakter subiektywny, nie ujawniają całej prawdy.

Wykorzystanie urządzeń technicznych w badaniach terenowych i obserwacji

Klasyczny obraz antropologa prowadzącego badania terenowe przedstawia go w afrykańskim szałasie otoczonego przez tubylców i piszącego wielostronicowe dzieło przy świetle księżyca. Tymczasem ewaluator w swoich badaniach terenowych korzystać może z licznych innowacji technicznych, które — rozsądnie użyte — sprawiają, że jego praca w terenie staje się bardziej efektywna i bardziej wszechstronna. Przede wszystkim dostępne mu są magnetofony i dyktafony. Nauka mówienia do mikrofonu wymaga praktyki, wysiłku i krytycznego spojrzenia na swoje pierwsze próby. Posługiwanie się magnetofonem wy-

maga także rozsądku, by nie stało się obsesją i barierą hamującą procesy społeczne i reakcje uczestników. Magnetofon przynosi więcej pożytku jako rejestrator osobistych notatek terenowych niż jako instrument trzymany cały czas pod ręką i uruchamiany przy okazji każdej rozmowy, w której obserwator uczestniczy.

Kolejnym narzędziem, którego użycie może usprawnić prowadzenie notatek terenowych, jest przenośny komputer. Również aparat fotograficzny może stać się częścią reporterskiego wyposażenia badacza terenowego, podobnie jak wideokamera, która stała się kolejną innowacją techniczną ochoczo wykorzystywaną w badaniach terenowych, zwłaszcza jeśli udaje się nią posłużyć w sposób dyskretny.

Bez względu na to, czy do sporządzania notatek terenowych użyte zostanie któreś z narzędzi współczesnej techniki czy też notatki te będą polegać na zwyczajnym spisywaniu dziejących się wydarzeń, pewne metody utrzymywania porządku obserwacji trzeba ustalić już na początku badań. Co więcej, sposób rejestracji dokonywanych obserwacji musi odpowiadać roli przyjętej przez ewaluatora i celowi jego obserwacji oraz uwzględniać wpływ procesu zbierania informacji na działania i osoby będące przedmiotem obserwacji. Wiele z tych procedur i pytań musi zostać opracowanych we wstępnej fazie badań terenowych, przy ich rozpoczynaniu. Niżej będzie mowa o paru zagadnieniach, które wymagają rozstrzygnięcia właśnie w tej wstępnej fazie badań.

Etapy badań terenowych

Do tej pory opisywaliśmy badania terenowe jako jedno integralne doświadczenie. Gdy praca w terenie przebiega sprawnie, staje się pewnym ciągiem następujących po sobie doświadczeń. Warto przyjrzeć się, jak się one zmieniają na poszczególnych etapach badań. Można wyróżnić trzy oczywiste okresy wszelkich badań terenowych: fazę wstępną, zasadniczy okres zbierania danych oraz fazę końcową.

Faza wstępna

Wchodzenie w strukturę sytuacyjną danego programu rozpoczyna się w chwili spotkania z osobą, która nadzoruje skład osobowy uczestniczących w programie. Jednakże z faktu, że „ktoś z góry" zaakceptował uczestnictwo obserwatora w programie, wcale nie musi wynikać, iż personel podległy tej osobie będzie chętny do współpracy z obserwatorem. Nie można również z góry zakładać pozytywnej reakcji klientów. Tak więc sam sposób wejścia w sytuację ma duże znaczenie i wymaga

wielostronnych, dyplomatycznych zabiegów. Jeśli ewaluator oczekuje współpracy ze strony personelu, uzyskanie akceptacji już na samym początku może mieć duży wpływ na ustanowienie wzajemnego zaufania i wzajemnego porozumienia w dalszej fazie jego pracy. Bywają jednak przypadki, że ewaluator spotyka się z niechęcią, a nawet wrogością.

Bez względu na to, jak wejście zostanie zaaranżowane, wstępna faza obserwacji jest „najmniej przyjemną fazą wszelkich badań terenowych" (Wax, 1971, s. 15). Jest to czas, kiedy obserwator musi znaleźć swoje miejsce w nowym otoczeniu, a ludzie z tego otoczenia muszą w jakiś sposób określić swój stosunek wobec obserwatora. Ten proces wzajemnego dostosowywania się ma doniosłe konsekwencje dla dalszego procesu badawczego.

Antropolog Rosalie H. Wax (1971) podkreśla znaczenie ukształtowania *wzajemnych* stosunków we wstępnej fazie badań. Oznacza to, że obserwator zajmujący się ewaluacją programu musi nauczyć się zachowywać w nowym dla niego otoczeniu, pozostali zaś uczestnicy muszą określić swój sposób zachowania wobec obserwatora. Wzajemne zaufanie, wzajemny szacunek i wzajemna współpraca zależą od ukształtowania się stosunków swoistej wymiany, polegającej na tym, że ewaluator otrzymuje potrzebne mu informacje, a osoby obserwowane coś, co wynagradza im ich współpracę z obserwatorem. Tym „czymś" może być poczucie ważności wynikające z faktu, że ktoś ich obserwuje, może być oczekiwany przez nie wpływ ewaluatora na ulepszenie programu, jakiś rodzaj satysfakcji z interakcji z obserwatorem czy też ze wspierania go w jego czynnościach obserwatora. Ten *model wzajemnych korzyści* zakłada, że można znaleźć jakiś powód, dla którego uczestnicy programu będą skłonni współpracować podczas ewaluacji i że może mieć miejsce pewien rodzaj obopólnej wymiany.

W sytuacji, gdy ludzie nie są przychylnie nastawieni do obserwacji, wzajemnej współpracy i wzajemnej wymiany, może być konieczna *infiltracja* (*infiltration approaches*). Douglas (1976, s. 167–171) opisał szereg strategii infiltracyjnych, jak: „wkradanie się", „sięganie po wytrych, który otworzy kogoś na nasze obserwacje", okazywanie „świętej prostoduszności"; powinny one rozbudzić w uczestnikach programu poczucie winy i w efekcie gotowość przyjścia z pomocą albo też przedstawić obserwatora jako „mięczaka" całkowicie nieszkodliwego dla obserwowanych. Douglas podpowiada również różne „ekspedycje okrężne", pozwalające badaczowi zmylić osoby biorące udział w badaniu co do jego celu rzeczywistego i przekonać je, że przedmiotem studiów jest

coś innego, niż im się wydaje. Istnieje również „taktyka wchodzenia fazami", polegająca na tym, że badacz, który ma trudności z wejściem do jednej grupy — tej, która jest przedmiotem jego studiów — rozpoczyna pracę w innej grupie, by później, gdy będzie to możliwe, powrócić do pracy w grupie będącej rzeczywistym przedmiotem jego zainteresowania.

Częstokroć najlepszą metodą przekonania nieprzychylnie nastawionych uczestników programu i uzyskania możliwości rozpoczęcia badań w tym środowisku jest powołanie się na znanego sponsora (*known sponsor approach*): obserwator powołuje się na racje i wiarygodność innej osoby, aby uprawomocnić i uwiarygodnić własne uczestnictwo w programie. Sięgając po tę strategię, trzeba mieć pewność, że znany sponsor jest rzeczywistym autorytetem.

Dobrze jest pamiętać, że bez względu na naturę badań terenowych, we wstępnej fazie badań, bardziej niż w jakimkolwiek innym momencie, *obserwator jest również obserwowany*, a ewaluator — również ewaluowany.

O względnej ważności słów *versus* czynów dla uzyskania wiarygodności we wstępnej fazie badań decyduje po części czas, jaki ewaluator zamierza spędzić w nowym otoczeniu. W przypadku prostych, powierzchownych obserwacji obecność badacza może się ograniczyć do konkretnych, parogodzinnych spotkań dziennie. W takich przypadkach problem „wejścia w grupę" oznacza coś innego niż wtedy, gdy obserwacja będzie wymagała uczestnictwa w działaniach podejmowanych w ramach programu przez długi okres czasu. Rosalie Wax opisała tę różnicę z dużą wnikliwością.

„Wszyscy badacze terenowi zastanawiają się, jak wytłumaczyć swoją obecność i swoją pracę swoim gospodarzom. «Jak powinienem się przedstawić?», zastanawiają się, lub: «Co powinienem powiedzieć o tym, czym się zajmuję?» Dla badacza, który ma w planie szybki i efektywny sondaż terenowy, pytania tego rodzaju są niezmiernie ważne. Sposób, w jaki prowadzący wywiad socjologiczny przedstawia siebie samego, słowa, jakich używa, w praktyce mogą stanowić o sukcesie lub porażce. [...]

Jeśli badacz terenowy zamierza zaangażować się w jakiś rodzaj obserwacji uczestniczącej, nawiązywać długotrwale związki, prowadzić studia wymagające poszerzenia własnych horyzontów, wtedy najlepszą rzeczą, jaką może zrobić, to rozluźnić się i powtarzać sobie, że nawet najrozsądniejsi ludzie nie uwierzą temu, co mówi ktoś obcy. Na dłuższą metę jego gospodarze będą go oceniać i wierzyć mu nie na podstawie tego, co mówi o sobie samym i swoich badaniach, lecz przede wszystkim ze względu na sposób bycia i postępowania oraz sposób, w jaki

się do nich odnosi. Na krótszą zaś metę mogą go akceptować lub tolerować, ponieważ został zarekomendowany przez jakiegoś krewnego, przyjaciela lub inną osobę, którą darzą szacunkiem" (Wax, 1971, s. 365).

Wax dowodzi, że w przypadku długotrwałej obserwacji stosunek ludzi do obserwatora kształtuje się na podstawie tego, co on robi, a nie tego, co mówi o swojej pracy. Gdy przychodzi mu powiedzieć coś o specyfice pracy ewaluacyjnej, są to wypowiedzi raczej formalne i grzecznościowe niż decydujące o tym, jak ewaluator będzie przyjęty. To, co rzeczywiście decyduje o reakcji na ewaluację i pracę ewaluatora, to przede wszystkim jego działania i stosunki, jakie wykształciły się między nim a tymi osobami, które obserwuje.

Nie można dać jakiejś uniwersalnej recepty na sposób rozpoczynania ewaluacji. Zależy on od charakteru ewaluacji, rodzaju programu i umiejętności obserwatora-ewaluatora. Wybierając strategię ewaluator powinien kierować się swoimi umiejętnościami, doświadczeniem społecznym, psychologicznym i politycznym. Oczekiwanie, że okaże się wrażliwy i świadomy tego, co robi, może narażać obserwatora właśnie w tym początkowym okresie na liczne frustracje i wątpliwości. Będzie się budził w środku nocy przerażony błędami i gafami, jakie popełnił w ciągu dnia. Będzie się czuł skrępowany, głupi i niepewny; będzie kwestionował zasadność całego projektu badawczego. Fakt, że jest przeszkolonym ewaluatorem prowadzącym duże projekty badawcze nie uodparnia go na normalne trudy uczenia się w nowych sytuacjach. Ale ów trud jest również częścią danych, jakich dostarcza mu badanie. Co więcej, ten wstępny okres badań może być dla niego również okresem radosnego ożywienia, zdobywania nowej wiedzy; może być okresem niezwykle inspirującym i podbudowującym go psychicznie. Może być to czas, w którym sprawdza swoje możliwości intelektualne, emocjonalne, fizyczne i społeczne. Dzieje się tak dlatego, że ta wstępna faza badań terenowych wyolbrzymia zarówno radości, jak i trudy związane z posługiwaniem się jakościowymi metodami badań.

Badania w toku: dynamika społeczna drugiej fazy

W drugiej fazie badań ewaluator wszedł już w swoją rolę, zna cel swoich badań i może przystąpić do zbierania danych. Nowość sytuacji już mu nie przeszkadza; obserwator zaczyna naprawdę *widzieć*, co dzieje się dookoła niego, a nie tylko niepewnie rozglądać się wokoło.

W przypadku badań terenowych często zdarza się, że obserwator zaczyna identyfikować się z ludźmi, których obserwuje. Zaczynając rozumieć zachowania, idee, obawy i uczucia uczestników i personelu

programu, możesz zacząć utożsamiać się z ich życiem, nadziejami i obawami. Taka identyfikacja może być naturalną i logiczną konsekwencją ukształtowania się między wami stosunków opartych na wzajemnym porozumieniu, zaufaniu i współpracy.

Identyfikowanie się z osobami uczestniczącymi w programie może być dla obserwatora doświadczeniem wstrząsającym, niejednokrotnie bowiem różni go od nich wykształcenie, doświadczenia, wiara w siebie czy też poziom dochodów. Ów proces utożsamia się z ludźmi będącymi przedmiotem badań stwarza nowe, swoiste problemy. Sytuacje społeczne rzadko kiedy są proste. Obserwator nie jest immunizowany na społeczną i polityczną dynamikę otoczenia, które obserwuje. Właściwie w każdym układzie istnieją jakieś grupy ludzi pozostające w konflikcie z innymi grupami. Te frakcje lub kliki mogą zaaprobować albo odrzucić obserwatora, rzadko kiedy jednak zachowują pełną neutralność.

Rzeczywiście ewaluator, aby uzyskać głębszy wgląd i lepiej zrozumieć jakąś konkretną grupę realizującą dany program, może chcieć się do niej przyłączyć. To, jak dokonuje się ów alians i jak jest interpretowany przez pozostałych uczestników programu, może mieć duży wpływ na całość pracy ewaluatora. Gdy ewaluacja jest jawna, obserwator, dzięki obowiązkom związanym z jego pozycją ewaluatora, może uzyskać dostęp do jakiejś frakcji czy kliki, nie stając się jej członkiem, a dystans ów uzasadniając koniecznością zachowania zasadniczej neutralności z uwagi na dobro całego procesu ewaluacji. Ilekroć obserwator nadmiernie utożsamia się z jedną podgrupą, dostęp do innych uczestników programu może być utrudniony.

Równocześnie naiwnością byłoby zakładanie, że możliwy jest ten sam stopień zażyłości lub dystansu w stosunku do wszystkich grup i frakcji. Ewaluatorzy jak wszyscy ludzie mają swoją osobowość i interesy: w naturalny sposób dla jednych będą bardziej pociągający niż dla innych. W gruncie rzeczy przeciwstawianie się tym naturalnym skłonnościom może sprawić, że obserwator będzie się zachowywał nienaturalnie, co utrudni mu pełną integrację z otoczeniem. Mając to na uwadze, obserwator prowadzący ewaluację stanie w obliczu bieżących decyzji w kwestii relacji personalnych i własnych związków z całą grupą oraz takiego podtrzymywania różnorodnych powiązań, aby nie stracić z oczu perspektywy, z jakiej patrzą na sprawy ci, którzy są luźniej z nim związani.

Przynależność do jakiegokolwiek środowiska pociąga za sobą konieczność osobistego wyboru relacji społecznych i aliansów grupowych, czyli opcji politycznych, kładąc jednak nacisk na projektowanie strategii

badań terenowych nie zamierzamy sugerować, że przeprowadzenie badań jakościowych w warunkach naturalnych jest rodzajem ekscytującej gry w szachy lub innej gry wojennej, gdzie swobodnie przesuwa się graczy lub pionki, by osiągać korzystny dla siebie cel. W badaniach terenowych są, oczywiście, okresy ekscytacji i okresy frustracji, lecz dominantą jest ciężka praca, niezwykła samodyscyplina oraz koncentrowanie się na sprawach bardzo przyziemnych, niejednokrotnie aż do znudzenia. Codzienna praca w terenie wymaga wielkiego wysiłku, a gromadzenie danych ciężkiej pracy.

Gdy badania terenowe mają się ku końcowi...

Gdy program badań terenowych zmierza ku końcowi i wydaje się, że większość danych została już skompletowana, coraz więcej uwagi trzeba poświęcić interpretacji zebranych materiałów. Kiedy obserwator czuje się coraz pewniejszy swojej wiedzy dotyczącej badanego środowiska, pojawia się coraz więcej pomysłów na temat spraw, do których trzeba się ustosunkować. Niektóre interpretacje podsunęły mu inne osoby, do pewnych doszedł sam. Krótko mówiąc, rozpoczyna się faza analizy, chociaż badania nie są jeszcze zamknięte.

Trzeba wyraźnie powiedzieć, że nie istnieje taki moment graniczny, w którym kończy się zbieranie danych, a zaczyna się analiza. W badaniach terenowych oba procesy nakładają się na siebie. Jeśli obserwator nabrał pewności siebie i zyskał rozeznanie co do specyfiki badanego programu — jego notatki terenowe mogą przybierać charakter coraz bardziej analityczny.

Gdy badanie zmierza ku końcowi, ewaluator coraz więcej zaczyna się zajmować *weryfikacją* danych, a coraz mniej gromadzeniem nowych. Sondując teren, ewaluator wystrzega się podciągania ludzi i sytuacji pod jakiekolwiek gotowe już kategorie konceptualne, to praktyczne doświadczenie terenowe podpowiada kategorie i wymiary pomocne w strukturalizacji wyników doświadczeń i obserwacji. Koncepcje te i wymiary również mogą podlegać weryfikacji w toku badań. Guba (1978b) opisał ten ruch wahadłowy między odkrywaniem a weryfikacją danych jako rodzaj „fali", jak bowiem przepływy i odpływy okresowo następują po sobie, tak i badacz koncentruje się na poszukiwaniu nowych danych, by następnie sprawdzać swoje domysły, idee i wyjaśnienia, a potem znów powraca do zbierania danych i kolejnej ich weryfikacji.

Kiedy praca badawcza w terenie ma się ku końcowi i wszystko przebiega dobrze, obserwator nabiera przekonania co do sensowności całego przedsięwzięcia. Ewaluator zaczyna wierzyć w użyteczność i ważność

swoich odkryć. Pod koniec pracy zebrane materiały i ich analiza układają się w coraz bardziej spójną całość. Powinna wyłonić się holistyczna struktura konceptualna ujmująca to, co w badanym programie istotne. Koncepcja ta musi być zrozumiała i spełniająca oczekiwania osób, które z ustaleń badacza będą korzystać. W tym momencie, ewaluator zaczyna wierzyć w swoje własne wnioski i ich przydatność dla usprawnienia programu.

Sprzężenie zwrotne jako element postępowania badawczego

Tym, co różni badanie ewaluacyjne od badań podstawowych, jest to, że ewaluator-obserwator musi być również zainteresowany reakcją, jaką wywoła jego badanie. Celem ewaluacji nie jest opublikowanie naukowego traktatu o życiu obserwowanych. Celem ewaluacji jest oddziałanie na decyzje i sposób realizacji programu. Tak więc, kończąc badania terenowe, ewaluator-obserwator musi zająć się uruchomieniem sprzężenia zwrotnego — zastanowić się, o czyją reakcję mu chodzi i co chce przez nią osiągnąć.

Uruchomienie sprzężenia zwrotnego *może być głównym składnikiem procesu weryfikacji ustaleń badawczych.* Ja sam opowiadam się za nieformalną procedurą ustnego przedstawienia uczestnikom i organizatorom programu opisu i analizy, i potraktowaniem ich reakcji jako części danych. Wzajemność w badaniach terenowych może polegać m.in. na porozumieniu, że uczestnicy programu zostaną zapoznani z wynikami obserwacji. Jak wynika z moich doświadczeń, zarówno uczestnicy, jak personel programu są żywo zainteresowani tego typu informacjami. Co więcej, doświadczenie dowodzi, że wiele nowego można dowiedzieć się z ich reakcji na przedstawione opisy i analizy, choć oczywiście nie jest możliwe przedstawienie wszystkiego, co zostało zaobserwowane. To nieformalne sprzężenie zwrotne uruchomione w fazie końcowej badań zasadniczo różni się od systematycznych analiz ostatniego okresu badań terenowych, lecz nie jest stratą czasu, można więc jeszcze przed zakończeniem badań podzielić się pewnymi wnioskami, by poznać reakcje tych, których opisują zgromadzone dane.

Wybór momentu takiej konfrontacji jest szczególnie ważny i trudny w przypadku ewaluacji formatywnej. Jeśli celem ewaluacji jest dostarczenie danych mających służyć ulepszeniu programu, jego organizatorzy często się niecierpliwią i chcą otrzymać informacje najszybciej jak można. Bywa, że ewaluator-obserwator ugina się pod tą presją i przedstawia swoje spostrzeżenia przedwcześnie, nie mając jeszcze całkowitej pewności co do przyjętego schematu rozumowania.

To ważne, aby ewaluatorzy prowokujący reakcję formatywną [taką, która może mieć wpływ na wynik końcowy ewaluacji] nie ulegali presji i nie przedstawiali interpretacji i analiz, zanim zyskają pewność, że wyselekcjonowane wyniki obserwacji naprawdę reprezentują ten model, który zasługuje na uwagę. Z drugiej jednak strony, nie można wskazać takiego momentu, w którym ewaluator już wie, że ma coś do przekazania. Ewaluator nieustannie potyka się o dylemat: jeśli przedstawi model niedopracowany, może spowodować ze strony organizatorów interwencje nietrafne, jeśli zaś spóźni się z weryfikacją poczynionych ustaleń, to przyjęte wzory mogą się utrwalić i ich zmiana okaże się trudna lub wręcz niemożliwa.

Nigdy nie natrafiłem na sytuację, żeby obserwacja z zewnątrz, z jednej strony, a interpretacje i zamierzone sprzężenie zwrotne [między wszystkimi zainteresowanymi], z drugiej, stanowiły układ idealnie zrównoważony. Dwustronne powiązania to kwestia wzajemnego osądu i zależą od charakteru stosunków między ewaluatorem a organizatorami programu. Kiedy się ma wątpliwości, a stosunki między ewaluatorem a organizatorami nie układają się na zasadzie pełnego zaufania, radzę ewaluatorowi-obserwatorowi polegać głównie na sobie. W stosunkach społecznych na ogół tak bywa, że wadliwe ujemne sprzężenie zwrotne* długo się pamięta i wielokrotnie analizuje. Z drugiej jednak strony — właśnie ścisłe wzajemne konsultacje mogą spowodować, że organizatorzy programu upewnią się w tym, co sami robią, i przestaną doceniać intuicje ewaluatora.

Tradycyjny model badań terenowych nie wymagał umiejętności inicjowania takiego sprzężnia zwrotnego, o jakim mowa. Zmienia ono rolę ewaluatora. Ci, którym przedstawia on swoje wnioski, zaczynają dużo lepiej orientować się w tym, jak ich zachowania i wypowiedzi są obserwowane. Ów dodatkowy wymiar wzmacnia normalne oddziaływania badawcze na obserwowane otoczenie i zwiększa wpływ obserwatora-ewaluatora na sytuację, w którą jest wciągnięty.

Gdy ewaluacja ma się ku końcowi, gdy ewaluator przygotowuje się do zakończenia badań, gdy porządkuje notatki i spostrzeżenia, aby sprawdzić, jaką wywołują reakcję, wtedy wpływ jego obecności na badane środowisko staje się szczególnie widoczny. Zamierzona we-

*Sprzężenie zwrotne — w teorii systemów oznaczające powrotne oddziaływanie skutku zjawiska na jego przyczynę — może być dodatnie lub ujemne. Z ujemnym mamy do czynienia wtedy, gdy skutek zjawiska przeciwdziała jego przyczynie. Ujemne sprzężenie zwrotne jest stosowane jako metoda regulacji procesów, m. in. także w układach społecznych — *przyp. red.*

ryfikacja przez konfrontację ukierunkowuje i doprowadza do punktu kulminacyjnego wszystkie nieuniknione efekty obecności obserwatora.

Obserwator i to, co jest obserwowane, czyli jedność i odrębność

Oddziaływanie ewaluatora na przedmiot obserwacji nie jest problemem swoistym jakościowych metod ewaluacyjnych. Zasada nieoznaczoności Heisenberga dotyczy tego samego problemu na gruncie nauk przyrodniczych. Zasada Heisenberga mówi, że narzędzia użyte do pomiaru położenia i momentu pędu elektronu wpływają na dokładność pomiaru. Kiedy badacz mierzy położenie elektronu, zmienia się moment jego pędu, kiedy zaś przedmiotem pomiaru jest moment pędu, nie można dokładnie określić położenia elektronu. Sam proces obserwacji wpływa na to, co jest obserwowane. Efekt ten jest rzeczywisty, nie wynika z błędnej percepcji ani z niedoskonałości narzędzi pomiaru. Wtargnięcie obserwatora zmienia sytuację.

To samo dotyczy programu poddanego badaniom ewaluacyjnym. Jaki będzie zakres ich wpływu, zależy od charakteru badań, osobowości obserwatora, zastosowanych procedur i od mnóstwa innych nie dających się przewidzieć okoliczności. Naukowcy prowadzący badania eksperymentalne i sondażowe, analitycy zajmujący się analizą kosztów i zysków, psychologowie stosujący standaryzowane testy psychologiczne, wszyscy wywierają wpływ na sytuacje, w których przystępują do gromadzenia danych. Fakt ten wydaje się bezsporny, powstaje natomiast pytanie, jak monitorować wszystkie te efekty, które wywołuje obecność badacza, i jak nie przeoczyć ich przy interpretacji danych.

Walor badań w warunkach naturalnych polega na tym, że obserwator jest dostatecznie włączony w sytuację, aby móc osobiście zrozumieć to, co się dzieje. Nie można dokładnie przewidzieć, jakie zmiany wywoła jego obecność. Można natomiast, podejmując decyzje co do roli, jaką będzie odgrywać obserwator, przewidywać sytuacje, jakie mogą powstać, i opracować strategie na użytek tych sytuacji. Zapoznając się z charakterem działań, które mają być realizowane, można określić stopień uczestnictwa w nich obserwatora i możliwe efekty tego uczestnictwa. Często rola obserwatora i wpływ wywierany przez jego obecność zmieniają się w miarę postępu badań. Ideałem jest taka strategia, która pozwoli ewaluatorowi zachować należytą niezależność myślową względem programu i równocześnie być obserwatorem — analizować i własne doświadczenie, i program.

Niezależnie od tego, czy ktoś podejmuje obserwację uczestniczącą czy jedynie obserwację z pozycji widza, to, co dzieje się w obserwowanym środowisku, będzie w pewnym stopniu zależeć od roli, jaką narzucił sobie obserwator. Od roli obserwatora, a zatem perspektywy, z jakiej prowadzona jest obserwacja, do pewnego stopnia zależeć będzie rodzaj gromadzonych danych. Osobisty charakter obserwacji ma swoje mocne i słabe strony: jej mocną stroną jest to, że osobisty udział może być źródłem doświadczeń i rozumienia „z pierwszej ręki", a jej słabością to, że osobiste zaangażowanie może ingerować w sytuację w sposób, który ją zniekształca. Wzajemna zależność między obserwatorem a przedmiotem jego obserwacji określa perspektywy badań w warunkach naturalnych.

Podsumowanie: wskazówki dla badaczy terenowych

Przedstawienie precyzyjnego zestawu reguł i procedur przydatnych w realizacji badań terenowych jest sprawą trudną, jeśli w ogóle możliwą. Patrząc wstecz na to, co zostało napisane, trzeba jeszcze raz podkreślić: wszystko, co robisz, zależy od sytuacji, celu twoich badań, charakteru środowiska, w którym je przeprowadzasz, twoich umiejętności, interesów, potrzeb oraz od twojego punktu widzenia. Wszystko jest ważne. Z tego wynika parę reguł przewodnich ułatwiających prowadzenie badań terenowych:

1. Pamiętaj, że twoje notatki terenowe muszą być opisowe.
2. Zbieraj różne informacje z różnych źródeł i różnych punktów widzenia.
3. Sprawdzaj wartość danych konfrontując je między sobą — posługuj się obserwacjami, wywiadami, analizą dokumentów, swoimi zapiskami i fotografiami.
4. Używaj cytatów; przedstawiaj uczestników programu ich własnymi słowami. Przedstawiaj punkty widzenia i doświadczenia uczestników językiem, w jakim je wyrażają.
5. Dobieraj kluczowych informatorów rozsądnie i wykorzystuj ich opinie z wyczuciem. Możesz czerpać z ich wiedzy, ale pamiętaj, że każdy z nich ma swój własny, subiektywny punkt widzenia.
6. Wiedz i pamiętaj, że kolejne fazy badań różnią się między sobą.
   a) We wstępnej fazie buduj wzajemne zaufanie i wzajemne porozumienie. Pamiętaj, że ewaluator prowadzący obserwacje jest również obserwowany i oceniany.
   b) Bądź czujny i zdyscyplinowany w środkowej fazie, gdy badania są w toku.
   c) Gdy badania mają się ku końcowi, skoncentruj się na porządkowaniu obserwacji i spostrzeżeń, zmierzaj ku syntezie.
   d) W sposób zdyscyplinowany i skrupulatny prowadź szczegółowe notatki terenowe na każdym etapie badań.
7. Staraj się jak najgruntowniej poznać program od wewnątrz w jego wymiarze praktycznym, ale nie zatrać perspektywy analitycznej, pamiętaj, że jesteś ewaluatorem.
8. Wyraźnie oddzielaj opis od interpretacji i własnych ocen.
9. Zaprezentuj swoje spostrzeżenia i potraktuj to jako proces weryfikacji wyników; starannie wybierz moment, obserwuj, jaką reakcję wywołują twoje wnioski.

10. W notatkach ewaluatora i w raporcie końcowym nie wahaj się uwzględnić własnych doświadczeń, myśli i odczuć. Są one częścią danych i dorobkiem badań.

Badania terenowe są zawsze doświadczeniem niezwykle osobistym. Są one doświadczeniem osobistym, ponieważ na procedury badawcze nakładają się indywidualne zdolności badacza i różnice sytuacyjne. Trafność i doniosłość uzyskanych rezultatów zależy przede wszystkim od umiejętności ewaluatora-obserwatora, jego samodyscypliny i przyjętej perspektywy badawczej. Wszystko to razem stanowi tak o sile, jak i słabości obserwacji jako metody.

tłumaczył *Michał Kowalski*

Herbert J. Rubin,
Irene S. Rubin

Jak zmierzać do celu nie wiążąc sobie rąk

Projektowanie wywiadów jakościowych

Herbert J. Rubin, Irene S. Rubin, „Keeping on Target While Hanging Loose. Designing Qualitative Interviews", rozdział z książki tychże autorów *Qualitative Interviewing. The Art of Hearing Data*, Sage Publications, Thousand Oaks – London – New Delhi 1995.

Projekt badania to nic innego niż zaplanowanie, kogo i o co zamierzasz pytać oraz dlaczego chcesz postawić te właśnie pytania (por. Marshall, Rossman, 1989). Przedsięwzięcie badawcze rozpoczynasz od projektu, który pozwoli ci zmierzać do celu, czyli do uzyskania odpowiedzi na stawiane sobie pytania. Nie oznacza to, że masz się kurczowo trzymać tego projektu. Pytania zmieniają się w toku badania, w miarę, jak otwierają się nowe kierunki dociekań.

Można powiedzieć, że projektowanie wywiadu jakościowego przypomina planowanie wakacji. Masz pewne ogólne wyobrażenie o tym, co chcesz zobaczyć i co chcesz robić, ale nie zakładasz, że na krok nie odstąpisz od wyznaczonej marszruty. Dopuszczasz zmianę planów, jeśli coś cię szczególnie zainteresuje lub gdy przydarzy ci się jakaś przygoda. Stale jednak będziesz pamiętać o ostatecznym celu podróży.

Projekt badania pomoże ci pamiętać o motywach, które skłoniły cię do jego podjęcia, zachęcając zarazem do sondowania zagadnienia na różne sposoby. Projekt pomoże ci zebrać wiarygodne dane, a także przekonać osoby czytające twoje opracowanie, że dane te uzyskałeś w sposób systematyczny i przemyślany. Dobry projekt osadza badanie w szerokim kontekście teorii i praktyki, co pozwoli ocenić, czy uzyskane wyniki są istotne; pomaga ci nie tracić z oczu badanego zagadnienia oraz chroni cię przed podejmowaniem w wywiadach zbyt wielu różnych wątków, których nie potrafisz potem ze sobą powiązać.

Nie można jednak z góry zaprojektować wszystkiego, w miarę bowiem uzyskiwania informacji od kolejnych rozmówców, projekt dalszych wywiadów może się zmieniać. Można natomiast rozpocząć od próbnego, szkicowego projektu, od rozmów z potencjalnymi respondentami, od porządkowania pomysłów, przeformułowywania badania i decydowania, z kim jeszcze i o czym należałoby porozmawiać. Na tym etapie możesz napisać propozycję, określając cel i przedmiot badań, precyzując ich znaczenie oraz przedstawiając to, co do tej pory, usłyszałeś. Możesz w projekcie wskazać, jak masz zamiar realizować swój plan (Locke, Spirduso, Silverman, 1987). W przypadku badań jakościowych sensowny projekt badawczy, nadający się do przedstawienia sponsorom lub komisji kwalifikacyjnej, może być opracowany tylko po wykonaniu tej wstępnej pracy. Jednak nawet wtedy, gdy projekt jest gotowy, dalej podlega on zmianom, w miarę jak zdobywasz wiedzę o przedmiocie badania i odkrywasz nowe perspektywy.

Projekt wywiadów jakościowych jest raczej elastyczny, etapowy i ewolucyjny (*flexible, iterative, continuous*) niż sztywny i niezmienny. Niżej wyjaśnimy, co rozumiemy przez określenia: elastyczny, etapowy i ewolucyjny. Opiszemy, w jaki sposób badacze znajdują i selekcjonują tematy badań, przystosowując je do własnych zainteresowań i osobowości, do zainteresowań oraz wiedzy swych rozmówców. Przedstawimy wreszcie, w jaki sposób na podstawie wywiadów buduje się sensowną teorię, oraz jak projektować badanie, aby cel ten osiągnąć.

Projekt elastyczny, etapowy i ewolucyjny

Projekt badań jakościowych posługujących się wywiadem przybiera swój kształt stopniowo, w miarę prowadzenia rozmów i notowania zawartej w nich treści. To, co wydaje się ważne na początku, później okazać się może mniej istotne, a z kolei kwestie, które na początku badania wyglądały na mniej ważne, okazać się mogą znaczące. Wobec tego — aby dostrajać go do tego, czego się dowiadujesz — twój projekt musi być elastyczny. Pracując nad projektem musisz pozostawić na boku własne przeświadczenia co do tego, jak się rzeczy mają, i czerpać pomysły i tematy od swoich rozmówców. Następnie powinieneś je przeanalizować i stopniowo wybierać te, które poddasz szczegółowym badaniom.

Jest to proces rozłożony na etapy — w jego wyniku powoli wyłania się projekt ostateczny. Na każdym etapie gromadzisz informacje, analizujesz je, selekcjonujesz i testujesz. W wyniku analiz i sprawdzianów

możesz ulepszyć lub zmienić pytania, dobrać inną grupę rozmówców i powtórzyć cały proces. Każda kolejna próba przybliża cię do sedna kwestii, które chcą ci zakomunikować rozmówcy. Ponawiając takie działanie kilkakrotnie, będziesz mógł zbudować teorię wyjaśniającą lub przedstawić opis dokładnie odpowiadający danym uzyskanym w wyniku wywiadów i mający za podstawę doświadczenia i sposób rozumienia badanych kwestii przez twoich rozmówców.

Projekt podlega ewolucji w tym sensie, że musisz go przeformułowywać w różnych fazach badania. Początkowo najwięcej czasu poświęcasz definiowaniu i redefiniowaniu przedmiotu badania i dociekaniu, jak interesujące cię kwestie rozumieją twoi informatorzy. W miarę postępów w pracy zaczniesz dostrzegać poszczególne idee i wyobrażenia twoich rozmówców i ustalać, czy są ze sobą zbieżne, czy nie, i co mogą oznaczać. Projekt winien skoncentrować się dalej na sposobach porządkowania tych koncepcji, zorientowaniu się, które są ważne, które mają związek z innymi, które występują powszechnie, a które są odosobnione. W środkowej fazie konstruowania projektu badawczego przychodzi czas na przemyślenie sposobów testowania tych koncepcji, a także na ponowne zastanowienie się, z kim należy rozmawiać, jak pytać, czego poszukiwać i jakie wyciągać wnioski. W końcowej fazie badania przedstawisz i przetestujesz teorię opartą na zebranych przez ciebie danych.

Elastyczność projektu

Poprawianie projektu w trakcie pracy jest normalnym, przewidywanym elementem badań jakościowych. Dowiadując się stopniowo, jak twoi rozmówcy rozumieją swój świat, możesz modyfikować przedmiot badań albo przemyśleć na nowo kwestionariusz wywiadu. Taka elastyczność jest znacznie lepsza niż trwanie przy projekcie, który się nie sprawdza bądź nie pozwala podążać w kierunku, którego nie przewidziałeś.

Wiele badań jakościowych dotyczy spraw stosunkowo mało zbadanych. Łatwo tu o przyjęcie założeń, które później okazać się mogą fałszywe. Jim, nasz kolega, przygotowując doktorat, badał, jak fundacje rządowe selekcjonowały projekty badawcze w kryminologii. Po pierwszych wywiadach z dwoma urzędnikami agencji federalnej Jim zorientował się, że przyjął zły tok myślenia: urzędnicy federalni nie postępują według zasad wykładanych w literaturze przedmiotu, przyjęte przez niego założenia okazały się naiwne. W rezultacie Jim przeformu-

lował temat badań i zajął się kwestią, jak fundacje dokonują wyboru między propozycjami.

Nie tylko błędne założenia wyjściowe zmuszają potem do dokonania zmian w projekcie. Zmiany takie mogą być potrzebne także wtedy, gdy w trakcie pracy wyłoni się lepszy sposób zbadania danego problemu. Jim próbując badać przestępczy rynek komputerowy postanowił zacząć od osób zajmujących się nielegalnym kopiowaniem i sprzedawaniem oprogramowania. Okazało się jednak, że komputerowy pirat, z którym Jim rozmawiał, nie chciał odpowiedzieć na pytanie, jak zarobił tysiące dolarów na swojej nielegalnej działalności. W czasie wywiadu człowiek ów bawił się komputerem i modemem. Jim zapytał go, co robi. Okazało się, że włamuje się do nie swojego komputera. Był też znacznie bardziej chętny do rozmowy o „hackerach" niż o komputerowym piractwie. Jim zmienił więc swoje pytania tak, by dotyczyły hackerów — innej grupy spośród interesujących go przestępstw.

Projekt musi być elastyczny w czasie całego badania, ponieważ mogą pojawiać się nowe kwestie i tematy. Swego czasu Irene Rubin badała sposób konstruowania budżetów miejskich. Okazało się, że urzędnicy i radni współdziałali w procesie tworzenia budżetu w sposób, którego ona w ogóle nie brała pod uwagę. Musiała więc przeformułować swój kwestionariusz, by uchwycić owe nieoczekiwane procedury (I. Rubin, 1992).

Elastyczność projektu jest pomocna z jeszcze jednego powodu. Zamiast ograniczać się do jednego zestawu pytań, które zadawać będziesz wszystkim swoim rozmówcom, możesz zbudować kwestionariusz tak, aby każdego z respondentów pytać o to, w czym jest zorientowany najlepiej. Przykładem może być badanie społeczności lokalnych przeprowadzone przez Herberta Rubina. Odkrył on, iż niektórzy z jego informatorów więcej mają do powiedzenia o filozofii społeczności lokalnej, inni z kolei o finansowaniu konkretnych projektów. Herbert nie zmuszał ludzi do mówienia o tym, co ich nie interesowało. Próbował natomiast sam dostosować się do tematów preferowanych przez respondentów. W rezultacie uzyskał bogatszy obraz badanej rzeczywistości (H. Rubin, 1994).

W wywiadach jakościowych elastyczność w przystosowywaniu kwestionariusza do wiedzy, jaką sam zdobywasz, oraz do wiedzy twoich rozmówców daje odkrywcze i interesujące wyniki. Możliwość łatwego przebudowania kwestionariusza pomoże ci głębiej wniknąć w sens wypowiedzi rozmówców bez potrzeby odrzucania tych fragmentów, które nie pasują do twojej początkowej koncepcji badawczej. Elastyczność

umożliwi ci także badanie tych obszarów interesującej cię problematyki, których istnienia nie przewidywałeś.

Etapowość projektu

Projekt wywiadów jakościowych realizuje się etapowo. Oznacza to, że powtarzając podstawowy proces gromadzenia informacji, ich analizowania, selekcjonowania i sprawdzania, przybliżasz się do klarowniejszego i bardziej przekonującego modelu badanego zjawiska. W początkowej fazie prowadzenia wywiadów projekt kładzie nacisk na gromadzenie wielu tematów i idei. Gdy badanie jest zaawansowane, bardziej koncentrujesz się na wybraniu pewnego ograniczonego zestawu kwestii, które będziesz drążył. W fazie końcowej skupisz się na analizie i testowaniu koncepcji, jaka ci się rysuje na podstawie tego, czego się dowiedziałeś, zaczniesz budować teorie, które poddasz ocenie twoich rozmówców oraz merytorycznych recenzentów.

W pierwszych wywiadach zależy ci na uzyskaniu szerokiej gamy pomysłów, tez i wyjaśnień, nie ograniczasz swobody wypowiedzi respondentów. Załóżmy, że tematem badania są „wakacje". Zamiast narzucać własne rozumienie wakacji, zaczniesz raczej od ogólnych pytań w rodzaju: „Gdy mówi się o wakacjach, co panu/pani przychodzi na myśl"? Jedni mogą wtedy opowiadać, gdzie byli lub co widzieli, inni zaczną mówić, na co mają ochotę albo czym są wakacje dla nich. Niektórzy z respondentów mogą zacząć opowiadać, jak drogie są podróże lub że nie mogą wyrwać się z pracy na urlop. Mogą opisywać takie formy działań, których nigdy nie skojarzyłbyś z odpoczynkiem czy rekreacją. Powinieneś zadać te same ogólne pytania możliwie wielu różnym ludziom, reprezentującym różne zawody i różne style życia.

Posługując się początkowo pytaniami otwartymi, będziesz mógł usłyszeć, co myślą twoi rozmówcy, zanim niepostrzeżenie zawęzisz zakres poruszanych zagadnień. Warto zaczynać od możliwie szerokiej gamy problemów z dwóch powodów: uzyskasz szerszą paletę odpowiedzi oraz pewność, że to, o co zapytasz później, ma znaczenie dla respondenta.

Kolejny etap to zawężanie poruszanych kwestii. Sposób, w jaki to zrobisz, zależeć będzie od twoich zainteresowań, od problematyki związanej z twoją dyscypliną oraz, oczywiście, od tego, co powiedział twój rozmówca. Socjolog być może szczególną uwagę zwróci na te wypowiedzi, w których mowa, jak wakacje integrują rodzinę. Może go to zainteresować zwłaszcza wtedy, gdy stara się ustalić, czy umacnianiu więzi rodzinnych lepiej służy wizyta u krewnych czy raczej wycieczki

do parków narodowych. Jeśli z kolei zajmujesz się polityką społeczną, bardziej może cię zainteresować, w jaki sposób na wybór miejsca spędzania wakacji wpływa postrzeganie przestępczości.

Następne zadanie to dopasowanie zakresu zagadnień do czasu, którym dysponujesz. Jeśli czas ten jest ograniczony, to skoncentrujesz się wyłącznie albo na tym, jaką wartość mają wakacje dla rodziny, albo na tym, jak wpływa na nie poczucie zagrożenia przestępczością. Jeśli masz więcej czasu, możesz spróbować dociec, od czego zależy wybór sposobu spędzenia wakacji z uwzględnieniem wartości rodzinnych, zjawiska przestępczości oraz atrakcyjności miejsc będących do wyboru.

Zdecydowałeś już, co skłonny jesteś uznać za istotne, ale czy twoi rozmówcy są tego samego zdania? Zanim więc przejdziesz do wywiadów pogłębionych, wróć do twoich pierwszych rozmówców i porozmawiaj z nimi w sposób nieformalny, by upewnić się, czy ich rozumienie problemu pokrywa się z tym, które ciebie interesuje. Upewnij się, czy znają dobre przykłady lub opowieści, które pozwolą pogłębić badany temat. Upewnij się, że to, czego chcesz się dowiedzieć, ma również znaczenie dla osób, z którymi rozmawiasz.

Opisana sekwencja: gromadzenie szerokich informacji, ich analiza, selekcja i testowanie, a wreszcie skoncentrowanie się na konkretnych pytaniach, które przyniosą poszukiwane odpowiedzi, powtarza się wielokrotnie. Stosujesz ją wtedy, gdy formułujesz temat, następnie wtedy, gdy poszukujesz koncepcji, jeszcze raz, gdy analizujesz kwestie badawcze, i znowu, gdy konstruujesz swoją teorię lub zespół hipotez wyjaśniających. Skąd wiadomo wobec tego, kiedy się zatrzymać? Zakończyć należy w momencie, kiedy z zebranych informacji wylania się kilka spójnych tematów, a kolejne wywiady nie wnoszą już nic nowego do problematyki, o którą pytasz. Glaser i Strauss (1967) moment ten nazywają punktem *nasycenia teoretycznego*.

Ewolucyjność projektu

Ewolucyjność wywiadu jakościowego oznacza, że pytania ulgają modyfikacjom w toku przedsięwzięcia badawczego. Jakaś konkretna odpowiedź może zasugerować nowy kierunek albo innych rozmówców niż przewidywani pierwotnie. Uzyskanie pełniejszego obrazu sytuacji w wyniku pierwszego etapu badań może sprawić, że w innym momencie zaplanujesz testowanie tego, co usłyszałeś.

Herberta Rubina interesowało, jak gminy rozwiązują problem mieszkań dla biednych. We wczesnej fazie badania skoncentrował się na pracy lokalnych organizacji działających w ekonomicznie zaniedbanych

dzielnicach większych miast. Przeprowadzając wywiady dowiedział się, że organizacje te w każdym z miast konkurują ze sobą zabiegając o ograniczone fundusze. Wobec tego chciał się dowiedzieć, jak owa konkurencja wpływa na ich pracę. Dlatego dokonał zmian w projekcie badania, uwzględniając w nim wywiady z aktywistami społeczności małych miast, gdzie istnieje mniej organizacji i mniejsza jest między nimi konkurencja.

Wprowadzanie zmian do projektu umożliwia podejmowanie nowych zagadnień bez rezygnowania z dotychczasowej struktury badania. Pozwala się upewnić, czy jest się na właściwej drodze oraz czy analiza jest wystarczająco pogłębiona, aby dostarczyć materiału do wyciągania wniosków, które znajdą potwierdzenie w zebranych danych. Nie wolno bowiem jedynie „dotykać" nowych zagadnień i posuwać się dalej. Należy raczej ocenić, czy wyłaniający się nowy temat jest istotny i jeśli okaże się istotny badać go dokładniej, wybierając odpowiednich rozmówców i opracowując nowe zestawy pytań. Opisane działania powtarza sie tak długo, dopóki nie uzyska się nasycenia teoretycznego.

Dalej omówimy, jak poszukiwać i selekcjonować zagadnienia badawcze. Projekt badania pomoże ci skupić się na tych kwestiach, które korelują z tym, co mówią respondenci, a równocześnie mających znaczenie teoretyczne.

Wybór zagadnienia

Badacz wybiera temat badania, który potem modyfikuje w świetle wypowiedzi respondentów. Przy wyborze owego tematu kierować się należy zainteresowaniami, możliwościami przeprowadzenia badania oraz tym, na ile problem może być analizowany za pomocą jakościowych narzędzi badawczych. Również precyzując temat upewniasz się, iż jest on wystarczająco ważny, aby poświęcić mu czas i wysiłek.

Zagadnienie nadaje się do badania jakościowego wtedy, gdy jest ugruntowane w doświadczeniach osób, z którymi zamierzasz rozmawiać. Jest to rzeczywistość ich życia, którą ty dopiero zamierzasz zrozumieć. Ty jednak także musisz być zainteresowany tą problematyką. Musisz być przekonany, że wciąga cię ona na tyle, że będziesz w stanie poświęcić jej czas potrzebny na przeprowadzenie badania jakościowego. Twoi rozmówcy zorientują się, jeśli będziesz znudzony, i przy braku zaangażowania z twojej strony oni także nie okażą większego zainteresowania badaniem. Zastanów się również, czy jakieś twoje nieprzychylne nastawienia mogą rzutować na badanie (złość, obawy, niechęć). Silne

zaangażowanie emocjonalne może być jednym z powodów, dla których podjąłeś daną problematykę, ale równocześnie może ono mieć wpływ zarówno na sposób formułowania i zadawania pytań, jak i na rozumienie odpowiedzi.

Źródła zagadnień badawczych

Skąd biorą się tematy badań jakościowych? Narzucają je pracodawcy, podsuwają je doświadczenia życiowe (Brown, 1991), mają one swoje źródło w osobie badacza, jego tożsamości rasowej, etnicznej lub seksualnej. Niektóre problemy budzą ciekawość badacza, inne pociągają go ze względu na wartości polityczne lub społeczne.

Niekiedy problematykę poddającą się badaniu metodą wywiadu jakościowego narzuca pracodawca. Irene Rubin zlecono przeprowadzenie wywiadów z biznesmenami. Chodziło o stwierdzenie, czy bodźce finansowe ze strony rządu rzeczywiście zachęcają ich do inwestowania w dziedziny, o które rządowi chodzi. Innym razem polecono jej zbadanie, dlaczego zawodzi system przepływu informacji w urzędzie sprawującym kontrolę nad zwolnionymi warunkowo z odbywania kary i jak można go naprawić. Herbert Rubin prowadził badania ewaluacyjne projektów rozwoju gospodarczego na Filipinach pod kątem celowości wydatkowania pieniędzy otrzymywanych w ramach pomocy zagranicznej.

Bywa, że taki zlecony temat po prostu wciąga badacza; kiedy indziej zaś brak należytego zaangażowania kompensuje on koncentrując się na pewnym interesującym go wycinku problematyki. Przypuśćmy, że burmistrz zlecił ci ewaluację programu selektywnej zbiórki śmieci w mieście. Twoją pierwszą reakcją może być myśl: „Ale nudy". Zaczynasz jednak przeprowadzać wywiady i stwierdzasz, że przeważa, zwłaszcza wśród młodzieży, postawa: „Cokolwiek zrobię, niczego to nie zmieni, po co więc mam sobie zawracać tym głowę". Dowiesz się także, iż wiele osób starszych przestrzega zasady segregowania śmieci, ponieważ „Wydaje się, że tak należy robić". Nagle okazuje się, że pojawiają się interesujące kwestie i problemy. Dlaczego niektórzy czują się częścią większej grupy, społeczności, gdy inni w ogóle nie poczuwają się do odpowiedzialności za jej losy? Możesz więc badać kwestie, które ciebie interesują, realizując równocześnie zamówienie zleceniodawcy.

Często temat badań podsuwa doświadczenie życiowe. Praca w zarządzie szkoły skłoniła pewnego badacza do szukania odpowiedzi na pytanie, dlaczego jedni obywatele akceptują wspieranie szkół z funduszów podatkowych, a inni nie. Ktoś inny pracował całe życie dla wojska

i doskonale orientował się, jaką politykę w tej dziedzinie prowadzi Kongres. Zainteresował się wobec tego nadzorem Kongresu nad agendami rządowymi. Zainteresowanie problematyką dyskryminacji ze względu na wiek może płynąć z własnych doświadczeń związanych z poszukiwaniem pracy; powrót na uczelnię celem dokończenia studiów po przerwie spowodowanej założeniem rodziny może skłonić do zajęcia się polityką naboru stosowaną przez uniwersytety.

Źródłem tematów jest też sam badacz jako osoba. Herbert Rubin często stanowi jednoosobową mniejszość w głosowaniach na radzie wydziału. Zainteresowała go więc kwestia, jak ludzie podporządkowują sobie innych. Znajomy, który przez większą część nocy pracuje, długo śpi i ma tylko jeden sportowy płaszcz, zainteresował się z kolei, jakie zachowania są określone jako dewiacje. Irene Rubin ciekawi, jak organizacje i jednostki aspirujące do kariery politycznej, które kładą nacisk na sukces i wzrost, dają sobie radę z niepowodzeniami i problemami trapiącymi społeczeństwo. Każda z tych kwestii może stać się tematem badań: Dlaczego ludzie obawiają się narzekać? Jak instytucje stojące na straży prawa traktują osoby, których działania nie rozumieją? Jak biurokracja reaguje na niedobory w budżecie?

Zdarza się, że tematyka badań jest pochodną tożsamości etnicznej badacza. Murzynów amerykańskich może szczególnie interesować, jak ruch obrony praw obywatelskich w latach sześćdziesiątych wpłynął na kształt obecnej polityki. Badacze pochodzenia żydowskiego mogą być zainteresowani ruchem politycznym kwestionującym śmierć 6 milionów Żydów w Europie w czasie II wojny światowej. Z kolei badaczy pochodzący z Meksyku lub Karaibów może przede wszystkim interesować obecna polityka imigracyjna.

Niektóre zagadnienia mogą przyciągać uwagę swoją niezwykłością, mogą być po prostu intrygujące. Na przykład prasa podała ostatnio opis zabójstwa na meczu baseballowym. Dlaczego zabójstwo przydarzyło się na meczu? Czy był tu jakiś związek między sportem — rywalizacją zastępczą — a przemocą, czyli rywalizacją spontaniczną? Czy zespoły biorące udział w meczu dzieliła rasa lub pochodzenie społeczne, a jeśli tak, to czy te antagonizmy odgrywały rolę na boisku? Wyjaśnianie odosobnionego przypadku może skłonić do zastanowienia się nad szerszymi kwestiami.

Źródłem zagadnień badawczych mogą być zainteresowania społeczne i chęć oddziałania na politykę społeczną. Można analizować przyczyny bólu i frustracji albo bezpieczeństwa i przyjemności w życiu osób niepełnosprawnych po to, by szukać sposobu przyjścia im z po-

mocą w samodzielnym radzeniu sobie z życiem. Ktoś inny chce zająć się inicjatywą oddania budynków mieszkalnych pod zarząd najemców lokali. Chce w ten sposób pomóc w zreformowaniu programów pomocy biednym. Z kolei inny badacz będzie studiować problematykę opieki zdrowotnej, chce bowiem dowiedzieć się, dlaczego jej koszty tak gwałtownie rosną.

Czy dane zagadnienie poddaje się badaniom jakościowym?

W procesie formułowania zagadnienia trzeba odpowiedzieć na pytanie, czy można je badać metodą wywiadu jakościowego. Dane o liczbie więźniów są ważną i użyteczną informacją, ale do ich uzyskania nie stosuje się badań jakościowych. Badania jakościowe stosuje się w odniesieniu do zagadnień wymagających głębszego zrozumienia, któremu najlepiej służą szczegółowe przykłady i rozbudowana narracja. W jaki sposób więzień uczy się przetrwania w zbrutalizowanych warunkach? Co robi miejscowa społeczność, aby przywrócić do stanu używalności dom, który zajmowali sprzedawcy heroiny? Jak dochodzi do powstawania deficytu w budżecie samorządu, mimo że jest niezgodny z prawem i trudno go ukryć? Problem można badać metodą wywiadu jakościowego wtedy, gdy pytasz, jak coś się dzieje, lub interesują cię szczegóły zagadnienia.

Wywiad jakościowy jest szczególnie przydatny, gdy pewne skomplikowane problemy trzeba ukazać w nowym świetle. *Outsiderzy* nie wiedzieli, dlaczego Murzyni amerykańscy uczestniczyli w marszach organizowanych w obronie praw obywatelskich, chociaż groziło za to aresztowanie. O wyjaśnienie poproszono Ivory'ego Perry'ego, czarnoskórego aktywistę z St. Louis. Perry odpowiedział: „I tak jestem w więzieniu, skoro są miejsca, do których nie mogę się udać, i rzeczy, których nie wolno mi robić" (Lipsitz, 1988, s. 80). Takie odpowiedzi pomagają wejrzeć w sytuację i wyjaśnić, dlaczego rzeczy mają się tak, jak się mają.

Wywiadem jakościowym należy się posłużyć wtedy, gdy celem badania jest rozwikłanie skomplikowanych zależności lub rozpoznanie powoli ewoluujących zdarzeń. Możesz po niego sięgnąć także wtedy, gdy chcesz dowiedzieć się, jakie decyzje bądź sytuacje były przyczyną stanu obecnego. Zaintrygowany faktem, że dobrze opłacani i dobrze traktowani pracownicy z rozdrażnieniem dyskutują o bieżącej pracy — możesz posłużyć się wywiadem jakościowym, by dociec, jakie wydarzenia pozostawiły tak mocny uraz. Być może dowiesz się, że właściciele fabryki polecili prywatnej policji spacyfikować strajk. I chociaż teraz jest inaczej — obawa i gniew pozostały.

Wywiady jakościowe są wskazane wtedy, gdy chodzi o głębsze zrozumienie problemów. Są również sposobem poznania szerszych implikacji problemu i pomagają umiejscowić go w kontekście historycznym, politycznym i społecznym.

Czy zagadnienie jest ważne?

Nie chcesz, oczywiście, prowadzić badań, których wyniki nikogo poza tobą nie zainteresują. Dlatego pracując nad projektem, musisz zastanowić się, dlaczego podejmowana przez ciebie kwestia jest ważna. Dla kogo może być ważna? Jak mogą być wykorzystane twoje wyniki? Częścią jakiej szerszej problematyki są twoje badania?

Czasami temat jest ważny, ponieważ ujawnia ukryty problem lub daje okazję do zabrania głosu ludziom pozbawionym tej możliwości. Badanie dotyczące maltretowania żon w Japonii może zwrócić uwagę publiczną na tę kwestię. Wywiady z pacjentami i lekarzami mogą pomóc wyjaśnić, dlaczego kobiety po pięćdziesiątce, równie często chore na serce jak mężczyźni, umierają z tego powodu dużo wcześniej. Wywiady z więźniami i ich żonami na temat, jak sobie radzą z dziećmi, mogą pomóc dzieciom — nie zamierzonym ofiarom istnej manii skazywania na karę pozbawienia wolności.

Zagadnienie badawcze może być ważne, jeśli pozwala lepiej zrozumieć jakiś powszechnie obserwowany problem (Broadhead, Fox, 1990). Badanie reakcji na zamykanie dużych przedsiębiorstw i utratę pracy pozwala stwierdzić, jak ludzie zachowują się w obliczu malejącej liczby względnie pewnych, dobrze płatnych zawodów w przemyśle (Lord, Price, 1992). Badania takie pokazują również związki zachodzące między utratą pozycji społecznej a nowymi ruchami politycznymi, które uciekają się do przemocy wobec cudzoziemców, mniejszości etnicznych i kobiet.

Zagadnienie może być ważne, jeśli dotyczy kwestii zagadkowej lub problemu znacznej wagi, którego nie można rozgryźć. Na przykład, dlaczego zimna wojna trwała tak długo mimo stanu zapaści, w jakim znajdował się Związek Radziecki. Stany Zjednoczone wydawały miliardy dolarów na obronę, ponieważ wywiad niezmiennie informował, że Związek Radziecki jest wciąż ogromnym zagrożeniem. Jakie zasoby zmarnowano, które priorytety błędnie określono? Co można powiedzieć o demokracji, jeśli polityka zdominowana jest przez agencje wywiadowcze zniekształcające informacje przez wzgląd na własne interesy?

Jeśli ewaluujesz program rządowy — twoje wyniki będą ważne o tyle, o ile raport z badań będzie czytany, a jego tezy wpłyną na

podejmowane decyzje. Jeśli badanie Departamentu Pomocy Dzieciom i Rodzinie, którego zadaniem jest ochrona praw i interesów dzieci, poprawi skuteczność jego działania, być może przyczynisz się do uratowania komuś życia. Jeśli zajmiesz się zmianami w realizacji budżetu federalnego, to niemal wszystkie twoje ustalenia mogą wywrzeć wpływ, ponieważ w grę wchodzić będą ogromne pieniądze.

Zagadnienie może być ważne, jeśli wskazuje sposób przerwania łańcucha negatywnych zdarzeń, takiego choćby, z jakim mamy do czynienia, gdy podupada jakiś kwartał miasta. W największym sklepie wybuchł pożar, interes upadł, a wypalony budynek stoi pusty. W efekcie w okolicy rosną stawki ubezpieczeniowe. Lokowanie się tutaj staje się droższe i mniej opłacalne. Puste budynki przyciągają różnych życiowych rozbitków, którzy piją i zaczepiają przechodniów. Dopóki nie wydarzy się nic, co przerwałoby taki łańcuch zdarzeń, choćby kupno i wyremontowanie budynku, zniszczenie postępuje dalej, aż okolica zaczyna wyglądać jak po bombardowaniu. Wskazanie, gdzie i jak interweniować w taki negatywny cykl, wymaga pogłębionych wywiadów.

Wywiady jakościowe dotyczą małego ośrodka, jednostkowej sytuacji, jednego podmiotu działania, jednego miasta czy okolicy, jednak wnioski wykraczać powinny poza te bezpośrednie okoliczności. Jeśli zajmujesz się znaczeniem tatuaży w środowisku więziennym, poznać możesz ogólniejsze kwestie buntu, przynależności do grupy, manifestowania swojego ja, recydywy lub reagowania na ograniczenie wolności. Uzyskana wiedza niekoniecznie musi dotyczyć tylko tego jednego więzienia, w którym prowadziłeś wywiady. Jeśli zajmujesz się likwidacją jednego przedsiębiorstwa, zadaj pytania, które pomogą ci rozeznać się w ogólniejszych aspektach sprawy, takich jak straty gospodarcze czy problemy przystosowania.

Jednym ze sposobów uzyskania pewności, że kwestia badana w mikroskali ma szersze implikacje, jest takie skonstruowanie projektu, aby uzyskane wyniki mogły sygnalizować jakieś ogólniejsze prawidłowości. Przypuśćmy, że masz zamiar prowadzić badania w więzieniu i sądzisz, że twoje ustalenia mogą dotyczyć reakcji ludzi na utratę wolności. Dla zweryfikowania tej szerszej kwestii przeprowadź kilka wywiadów w instytucjach innego rodzaju, na przykład w domach dziecka, szpitalach psychiatrycznych i koszarach. Fascynujące mogłoby być odkrycie, że wyniki badania w więzieniu znajdują potwierdzenie także w domu dziecka. Proponowałyby one ogólniejsze rozumienie reakcji ludzi na utratę wolności.

Jakich zagadnień unikać?

Nawet jeśli uznasz, że zagadnienie jest istotne i można je badać metodami jakościowymi, to możesz nie chcieć się nim zająć z przyczyn osobistych. Pamiętaj, że przed przystąpieniem do badań najpierw warto poznać siebie.

Umiejętnie prowadzony wywiad często wywołuje emocje i odsłania rozdzierające problemy, które okazać się mogą zbyt stresujące dla badacza. Rozmowy z ludźmi pracującymi w ośrodkach opieki nad dziećmi chorymi na AIDS mogą być wstrząsającym projektem badawczym, ale również to, co usłyszymy od wolontariuszy, może spowodować, że nie będziemy zdolni do dalszych badań.

Niektóre kwestie mogą wywoływać tak silne reakcje emocjonalne, że wysłuchanie jakichś przykładów czy historii okaże się dla badacza zbyt trudne. Ktoś, kto został zgwałcony, może nie być w stanie wysłuchać z uwagą opowiadania gwałciciela. Jeśli jesteś przekonany, że bankierzy i prawnicy są pasożytami, wywiad z nimi może cię krępować lub też sprowokujesz ich gniewną reakcję.

Czasami ważne jest, by poznać zagadnienie z dwóch stron i przeprowadzić rozmowy z osobami, z którymi nie zgadzasz się w zasadniczy sposób. Musisz sobie sam odpowiedzieć na pytanie, czy jesteś w stanie to zrobić. Nasz przyjaciel Jim miał kłopot z wywiadami z personelem więziennym, ponieważ odkrył, że są to ludzie wrogo nastawieni i rasiści. Musiał jednak pohamować na czas wywiadów swoją niechęć do nich, ponieważ bez wysłuchania ich nie da się zrozumieć więziennego życia.

Nastawienie bardzo pozytywne także może stwarzać problemy przy przeprowadzaniu wywiadów. Pracownicy socjalni chcą pomagać biednym, wiedzą jednak, że czasami działają oni na własną szkodę. Herbert Rubin, który swoimi tekstami staje w obronie ubogich, miał kłopoty z wysłuchaniem respondentów, których gniewa takie samodestrukcyjne zachowanie ludzi biednych.

Czy badanie da się zrealizować?

Innym kryterium wyboru zagadnienia jest pytanie, czy masz czas i pieniądze na realizację danego przesięwzięcia. Ile na nie potrzeba czasu? Czy można ograniczyć jego zakres, by dopasować do czasu, którym się dysponuje? Czy można je podzielić na etapy rozłożone w czasie? Ktoś, kto pracuje w dzień w pełnym wymiarze godzin, z większością respondentów nie będzie mógł przeprowadzić wywiadów

w ich miejscu pracy. Bez finansowego wsparcia nie będziesz mógł udać się do odległego miasta i przenocować tam, by zrealizować wywiady.

Niekiedy możliwość zrealizowania badania rozbija się o brak dostępu do odpowiednich osób. Możesz mieć doskonały pomysł — na przykład, jak sędziowie dochodzą do podjęcia decyzji, ale po prostu nie uzyskasz wstępu na ich obrady. Irene Rubin chciała przeprowadzić badanie porównawcze uniwersytetów publicznych i prywatnych, musiała jednak zmienić plany, ponieważ uniwersytety prywatne nie udzieliły jej zgody na badania.

Jeśli zaczynasz od ogólnie sformułowanego problemu, zazwyczaj znajdziesz teren, gdzie będziesz mógł badanie zrealizować. Irene Rubin, zastanawiając się nad tematem rozprawy doktorskiej, zdecydowała się zająć przypadkami ludzkich niepowodzeń, upadków i strat. Musiała zacząć od znalezienia osób, którym się w czymś nie powiodło — na przykład polityków, którzy przegrali w wyborach, lub studentów, którzy nie zdali egzaminów. Jednak odszukanie studentów po nie zdanych egzaminach byłoby bardzo trudne, jeśli w ogóle możliwe. Polityk przegrywający wybory może zniknąć ze sceny do czasu kolejnej elekcji. Nawet gdyby znalazło się wystarczająco wielu studentów i polityków, mogliby oni nie chcieć rozmawiać o tym, co zapewne uważają za swoją osobistą porażkę. Wobec tego Irene postanowiła badać organizacje przeżywające kłopoty ekonomiczne. Organizacje nie mogą się tak łatwo ukryć, a jednostki w nich zatrudnione nie obwiniają się z powodu strat i wobec tego są chętniejsze do rozmowy (I. Rubin, 1977).

Badacze, zwłaszcza ci, którzy nie ukończyli jeszcze studiów i nie pracują w ośrodku naukowym, podlegają niekiedy jeszcze innym ograniczeniom decydującym o tym, czy badanie da się zrealizować. Mogą mieć łatwy dostęp do miejsca, gdzie pracują — szpitala, więzienia, szkoły, sklepu itp. — a równocześnie może im brakować czasu na prowadzenie badań w innym miejscu. Mogą więc chcieć prowadzić wywiady w instytucji, do której mają dostęp. Zamiast wybierać zagadnienie, poszukują więc tematu, który mogliby podjąć w danym środowisku. [...] Należy jednak znaleźć temat o szerszym znaczeniu, nawet jeśli badanie nie wyjdzie poza obręb ściśle zlokalizowany.

Projektowanie w zamiarze budowania teorii

W wyniku wywiadów jakościowych badacze otrzymują obszerną charakterystykę jakiegoś obszaru tematycznego, jakiejś dziedziny kultury itp. Badacze o nastawieniu krytycznym prowadzą badania, aby

przekonać ludzi o potrzebie dokonania zmiany społecznej; wielu badaczy posługujących się metodami jakościowymi za właściwy i wystarczający cel badania uznaje wszechstronny opis sytuacji czy okoliczności.

Inni badacze stosujący metody jakościowe, w tym my, uważają, że celem tak realizowanych badań jest uzyskanie danych, które mogą posłużyć do zbudowania teorii opisującej sytuację lub wyjaśniającej zjawisko. Badacze zorientowani na metody jakościowe tworzą teorie inaczej niż pozytywiści. Według podejścia pozytywistycznego punktem wyjścia projektu badawczego jest istniejąca już teoria akademicka; zgromadzone dane służą do testowania i ewentualnie modyfikowania tej teorii. Odwrotnie rzecz ma się z badaczami stosującymi podejście jakościowe — budują oni teorię krok po kroku na podstawie materiału zebranego w wyniku wywiadów.

W pierwszych wywiadach badacz zaczyna testować przewidywane przyczyny danego zdarzenia oraz wybiera koncepcje i tematy, którymi się zajmie. Następnie projektuje kolejne wywiady, które pomogą sprawdzić i sprecyzować te wstępnie przyjęte tematy. Sugerują one pytania, jakie należałoby zadać; z kolei to, co usłyszymy, wskaże, jak przeformułować owe tematy i które z nich zbadać w sposób pogłębiony. Proces ten powtarza się wielokrotnie, dopóki zmodyfikowane koncepcje nie zostaną przetestowane i nie zarysuje się teoria wstępna, która ponownie jest testowana w kolejnych wywiadach.

Herbert Rubin w pierwszej fazie badania działaczy lokalnych usłyszał od swoich rozmówców, jak to lekceważą ich ludzie biznesu i bankowcy. Wobec tego postanowił dociec, jak działacze ci reagowali na lekceważenie i jakie to miało skutki. Opracował dotyczące tej kwestii pytania — na przykład pytał, czy zbuntowali się przeciwko takiemu traktowaniu, a jeśli tak, to co przez to osiągnęli.

Dzięki elastycznemu i ewoluującemu w miarę postępu badań projektowi badacz jest otwarty na różne nowe sytuacje i okoliczności, które pozwolą weryfikować wyłaniającą się teorię. Wywiady prowadzone w tych różnych okolicznościach mają na celu sprawdzenie, jak daleko sięga teoria oraz jak należałoby ją zmienić. Czy poczucie deprywacji jest udziałem działaczy głównie mniejszych, mniej znanych organizacji, czy może jest powszechne także w organizacjach większych, odnoszących sukcesy finansowe? Czy kolor skóry różnicuje to poczucie deprywacji? Czy działacze organizacji tworzących wpływową koalicję reagują inaczej na podważanie ich pozycji niż ci, których organizacje do takiej koalicji nie należą? W zależności od odpowiedzi na takie różne pytania umocnisz się w swojej teorii lub ją zmienisz.

Badacze zmierzający do uogólnień teoretycznych powinni uważnie przyjrzeć się *stanowiącym ich budulec przeświadczeniom, pojęciom i tezom*, które opisują kulturę i zarazem wyjaśniają, dlaczego coś się dzieje. Tezy to zdania, które wyjaśniają, dlaczego coś się stało lub co to coś oznacza. Tezy zbudowane są z pojęć. Pojęcia odzwierciedlają leżące u ich podłoża przeświadczenia i wyobrażenia, z ich pomocą ludzie etykietują swoje rozumienie świata (H. Rubin, 1983, s. 58–60). Pojęcia mogą reprezentować względnie proste wyobrażenia, jak choćby „fizyczny odpoczynek", lub bardziej skomplikowane przeświadczenia, na przykład „starzenie się" albo „atrakcyjność seksualna". „Starzenie się" może kryć w sobie różne komponenty: strach przed chorobami, utratę sił fizycznych, utratę dobrego wyglądu, niepewność materialną. „Atrakcyjność seksualna" może mieścić w sobie energię biologiczną, dobry wygląd, żywiołową osobowość, humor, ciepło, umiejętność kontaktowania się z ludźmi.

Tak w mowie, jak i w piśmie pojęcia występują zwykle pod postacią pojedynczych słów lub zwrotów. Zazwyczaj są to przymiotniki, równoważniki zdań lub rzeczowniki odsłowne. Słowa te są po prostu etykietami lub symbolami reprezentującymi głębiej ukryte przeświadczenia. Jednym z celów wywiadów jakościowych jest zrozumienie takich bazowych przeświadczeń oraz wskazanie słowa lub wyrażenia stanowiącego etykietę pojęcia.

Uwrażliwisz się na pojęcia słuchając nagranych wywiadów lub czytając notatki i wypisując te słowa i wyrażenia, którymi rozmówcy posługiwali się jako przykładami lub kluczowymi wyjaśnieniami. Słowa te mogą sprawiać wrażenie technicznego żargonu. Mogą mieć specjalne znaczenie w obrębie danej dziedziny. *Obsługa długu* znaczy coś innego dla rządu federalnego niż dla jakiejkolwiek innej agendy rządowej, ponieważ rząd federalny spłaca nie sam dług, a tylko odsetki. W związku z tym dla rządu federalnego obsługa długu oznacza obciążenie stałe, podczas gdy inne agendy są nią obciążone tylko doraźnie.

Pojęciem może być zwykłe, obiegowe określenie, które rozmówcy wciąż powtarzają, ponieważ jest ono dla nich ważne lub oddaje ważny składnik ich pracy albo życia codziennego. Działacze lokalni, mówiąc o swojej pracy, używają takich określeń, jak *upoważnienie, równość społeczna* i *dostarczenie mieszkań*; specjaliści od budżetu posługują się terminami *równowaga finansowa, subwencje celowe* i *wydatki pozabudżetowe*. Policja mówi *założyć obrączki* (kajdanki) i *załatwić papugę* (podejrzanemu adwokata), określając w ten żargonowy sposób czynności o istotnym dla siebie znaczeniu.

Powtarzane przez rozmówcę zwroty i opowieści często sygnalizują kryjące się za nimi, ważne dla niego pojęcia lub przeświadczenia. Słuchając opowieści o szefowej, usłyszysz być może, że jest sympatyczna i wiele uwagi poświęca pracownikom. Możesz z tego wywnioskować, że „życzliwość" jest pojęciem ważnym, daje wyraz przekonaniu respondenta, jaka powinna być szefowa.

Większość pojęć jest częścią języka, jakim posługują się badane osoby. Rozmówcy używają określonych słów lub zwrotów, które ty uznasz za wskaźniki szerszych idei. Niekiedy usłyszysz zbitkę pojęciową, którą potraktujesz jako wskazówkę sposobu rozumienia kwestii przez rozmówcę, choć on sam nie wyrazi swojej myśli za pomocą konkretnego słowa czy zwrotu. Przykładem mogą być dyskusje budżetowe, w których Irene Rubin często słyszała wzmianki o *subwencjach celowych* — pieniądzach przeznaczonych na konkretny cel, których nie można wydać na nic innego. Mówiono także o pieniądzach, które można wydać na dowolny cel. Środki te były szczególnie cenione, ponieważ dawały większą swobodę i mogły być użyte do rozwiązania różnych problemów. Jako że rozmówcy nie posługiwali się żadnym specjalnym określeniem na oznaczenie tej drugiej kategorii środków budżetowych, Irene nazywała je „subwencją bez przeznaczenia". Następnie spenetrowała sens tego określenia pytając, skąd bierze się owa subwencja, kto ją kontroluje i w jaki sposób jest wydawana.

Częścią pracy nad projektem, który podlega ewolucji, jest obmyślanie pytań mających służyć poznaniu sensu nowo odkrywanych pojęć. Proś o przykłady i doprecyzowanie związanych z nimi terminów po to, by słyszeć, w jaki sposób są one stosowane w różnych kontekstach.

W przytoczonym niżej fragmencie notatek Herbert Rubin zajmuje się pojęciem *umocowanie*, próbując zrozumieć, co ono oznacza dla tych, którzy budują domy dla ubogich. Herbert skonstruował pytanie wiedząc, że w przeszłości lokalni działacze mianem „umocowanie" określali także zwycięstwo nad administracją rządową — na przykład skłonienie miasta do ustawienia potrzebnych znaków „stop". Obecnie społeczności lokalne budują własne domy. Zamiast więc pytać „Co rozumiecie przez umocowanie?", które to pytanie sprowokować może abstrakcyjną, podręcznikową odpowiedź, Herbert poprosił o porównanie umocowania w dawnym i w nowym znaczeniu.

Herbert: Czym różni się w gruncie rzeczy poczucie umocowania, kiedy chodzi o dom i kiedy chodzi o znak „stop"?

Rozmówca: Jest większy, po pierwsze. I myślę, że oddziałuje na ludzi w inny sposób. Uczciwie mówię, jeśli ludzie mają przyzwoite miejsce do mieszkania, codziennie mają gdzie wrócić do domu, ich możliwości zajmowania się innymi sprawami stają się

dużo większe. Gdy nie muszę się martwić, gdzie będę mieszkał, ile zapłacę, to mogę zrobić następny krok i zacząć się przejmować masą innych rzeczy. Jeśli ludzie uzyskali ustawienie znaków „stop", ale wciąż martwią się, gdzie będą mieszkać, to ten znak nie sprawi, że zrobią następny krok. To jest sprawa samorealizacji. Najpierw troszczysz się o najbardziej podstawowe potrzeby, a potem wchodzisz na kolejny szczebel, w górę.

Nie zadając wprost pytania o definicję, Herbert ją otrzymał: „umocowanie" to rosnące możliwości rozwiązywania własnych problemów.

W wypowiedziach swoich rozmówców możesz poszukiwać pojęć i wtedy dopiero zadawać pytania, które pomogą ci zrozumieć, co one znaczą i w jaki sposób są używane. Następnie możesz stopniowo zestawiać z pojęć tezy, dłuższe zdania wyjaśniające lub wypowiedzi. Czasami sami rozmówcy, chcąc wyjaśnić zjawiska, wypowiadają tezy. Na przykład rozmówcy Herberta Rubina mogli powiedzieć: „Bankowcy nie szanują społeczności lokalnych, dlatego musimy iść czasami na kompromisy, żeby otrzymać fundusze". Pojawiają się tutaj dwa pojęcia: *szacunek* i *kompromis*. Łączą się one w tezę: brak szacunku wymusza kompromisy. Badacz buduje teorię, jak się rzeczy mają i dlaczego tak właśnie, zestawiając tezy w całość, która jest wyjaśnieniem odnośnej kwestii. Implikacje wyłaniającej się teorii analizowane są za pomocą kolejnych pytań zadawanych zarówno pierwotnym, jak i nowym rozmówcom.

Przechodzenie od ustalania i modyfikowania pojęć do budowania tez, a następnie prawdopodobnych teorii zilustrujemy przykładem. Załóżmy, że prowadziłeś wywiady z profesorami na temat ich problemów ze studentami. Profesorowie skarżyli się na studentów zalegających z terminami i przedstawiających różnym profesorom te same prace. To, co usłyszałeś, podsumujesz dwoma pojęciami: *spóźnianie się* oraz *oszukiwanie*.

W kolejnych wywiadach poprosisz o następne przykłady ilustrujące te problemy. Jednak na podstawie tego, co słyszysz, zaczniesz podejrzewać, że chodzi o coś więcej. Inni profesorowie narzekają na studenckie wykręty: a to zapomnieli wziąć ze sobą pracy, a to nie mogli przyjść na zajęcia z powodu kłopotów rodzinnych. Tym razem sądzisz, że profesorowie mówią o „sposobach wymigania się od pracy", czyli o innym przeświadczeniu. Oddawanie tej samej pracy różnym profesorom może być nie tyle „oszukiwaniem", ile inną formą „wymigiwania się od pracy".

W wywiadach z profesorami usłyszysz również, że czują się wykorzystywani, że studenci nie traktują wykładowców z szacunkiem, na jaki zasługują ze względu na osiągnięcia intelektualne. Będą także potępiać studentów za to, że tak dużo czasu spędzają „na piciu i zabawie".

Badacz zastanawiając się, jak kontynuować badanie, będzie teraz poszukiwał tez, które mogłyby wyjaśniać, jak profesorowie interpretują swój świat. Jedna z nich może brzmieć: „Studenci wymigują się od pracy, ponieważ nie szanują profesorów jako intelektualistów", a inna z kolei: „Studenci próbują wymigać się od pracy, ponieważ zdominowani są przez kulturę, która kładzie nacisk na zabawę". Takie prowizoryczne tezy sugerują pytania, które powinny być zadane celem sprawdzenia, czy i kiedy są one prawdziwe.

W kolejnych wywiadach badacz testuje wyłaniające się hipotezy. Tym razem pytania zadane zostają zarówno profesorom, jak i studentom, ponieważ po to, aby zweryfikować tezy, trzeba dowiedzieć się, jak obie strony widzą zagadnienie. Badacz zapyta studentów, czy dużo uczą się w domu i dlaczego czasami nie wykonują zadanych prac. Pojawią się także pytania, czy studenci szanują profesorów i co ten szacunek oznacza.

Przeprowadzając te wywiady zwróć uwagę, jakie związki zachodzą między pojęciami: szacunek, wymigiwanie się od pracy i zabawa. Gdy, na przykład, studenci opisują profesorów, których darzą szacunkiem, zapytaj, jak bardzo trzeba pracować na ich zajęciach. Jeśli zauważysz tu jakiś związek, będziesz mógł zacząć myśleć o teorii. Musisz jednak badania prowadzić dalej, żeby lepiej zrozumieć, co znaczy szacunek, o jaką zabawę chodzi i skąd takie właśnie rozumienie tych pojęć się bierze.

Budowanie i przebudowywanie teorii może poprowadzić cię w różnych kierunkach. Który z nich wybierzesz, zależy zarówno od rodzaju twoich zainteresowań, jak od przewidywanych adresatów wyników twoich badań. Czy interesują cię wyniki nauczania, skoro profesorowie uważają, że studenci wymigują się od pracy? Czy niektórzy z profesorów ulegli demoralizacji i minimalizują wysiłek wkładany w nauczanie, czy też mają pretensje do administracji, która przyjmuje studentów nie zainteresowanych nauką? Czy profesorowie próbują wybrać tych studentów, którzy chcą się uczyć, i poświęcają im więcej uwagi, angażując się w ich sukcesy?

Innym kierunkiem, w którym można podążać, jest pytanie, czy upowszechniający się na uniwersytecie obyczaj szukania wymówek przeniósł się również na inne dziedziny życia studentów. Czy studenci sądzą, że znajdowanie wymówek jest w porządku? Czy posługiwali się nimi także w szkole średniej? Czy stosują je w pracy? Każdy z możliwych kierunków prowadzenia badań podsuwa kolejne koncepcje, których

należałoby poszukiwać, pytania, które trzeba postawić, i rozmówców, z którymi trzeba przeprowadzić wywiady.

Posługując się przykładem z uniwersytetu pokazaliśmy, jak budować teorię, wybierając najbardziej obiecujące kierunki badań, i jak zmieniać projekt badania, by podążać za nowymi wątkami. Innym sposobem budowania teorii jest wybranie i zestawienie różnych punktów widzenia na badane zagadnienie. Podejmując próbę wyjaśnienia różnic, sformułujesz nowe tematy.

Irene Rubin na podstawie wywiadów w Rochester (w stanie Nowy Jork) na temat reformy budżetowej sformułowała dwie tezy wyjaśniające, dlaczego reforma ta nie powiodła się (I. Rubin, 1992). Po pierwsze, usłyszała, że reforma upadła, ponieważ szefowie wydziałów (policji, straży pożarnej, robót publicznych, parków) nigdy właściwie nie zrozumieli, o co w tej reformie chodzi. Po drugie, dowiedziała się, że politycy bali się jej, gdyż dzięki niej ludzie dowiedzieliby się, czy dotrzymane zostały obietnice wyborcze.

Pamiętając o tych dwóch tezach, Irene kontynuowała swoje badania w Phoenix w Arizonie. Urzędnicy byli tu bardziej profesjonalni, a politycy spokojniejsi o opinię obywateli znających osiągnięcia wybranych przez siebie przedstawicieli. Duże różnice między odpowiedziami uzyskanymi w Rochester i w Phoenix wskazywały na to, że nie zrozumie się kwestii związanych z reformą budżetu bez wcześniejszego zbadania ogólniejszego problemu reformy politycznej. W jaki sposób obywatele włączani są w sprawy budżetu? Czy urzędnicy zajmujący się budżetem to profesjonaliści czy ludzie z nadania politycznego? Co oznaczają pojęcia *włączenie obywateli* i *reforma polityczna* w kontekście tworzenia budżetu? Irene, dążąc do zbudowania teorii, zmieniła projekt badania — poszukiwała teraz odpowiedzi na pytanie, jak reforma i profesjonalizm wpływają na zmiany budżetu.

Teorię buduje się i testuje krok po kroku. Testowanie pomaga selekcjonować pomysły — słabsze z nich oraz nie znajdujące potwierdzenia w wywiadach zostają odrzucone. Testowanie pomaga również modyfikować teorię odpowiednio do nowych przypadków i sytuacji, tak by można ją było poprzeć dowodami, a także w miarę możliwości nadać jej walor uogólniający. Badanie kontynuujesz analizując inne okoliczności i sprawdzając implikacje wyłaniającej się teorii w nowych warunkach. Jeśli potwierdzają się, nabierasz pewności co do swojej teorii; jeśli nie — zmieniasz teorię odpowiednio do wiedzy zdobytej w nowych okolicznościach.

Załóżmy, że badasz więzienie, w którym doszło do buntu, i zarówno więźniowie, jak strażnicy i naczelnik twierdzą, że jego przyczyną było złe jedzenie. Musisz więc przetestować tezę łączącą „buntowanie się" z „jakością jedzenia". Sprawdzasz dodatkowe przypadki — więzienie znane ze złego wyżywienia, ale względnie spokojne, oraz inne, w którym dobrze dają jeść, a mimo to doszło tam do buntu.

Oba przypadki potencjalnie przeczą twojej pierwotnej tezie. Jeśli teza, ogólnie rzecz biorąc, jest słuszna, to złe jedzenie powinno być przyczyną buntu, gdy zaś dają dobrze jeść, w więzieniu nie powinno być buntów. Wybrałeś więc dwa przypadki, które nie potwierdzają tezy. Musisz teraz dowiedzieć się, dlaczego tak jest i co wynika z tego dla twojej teorii. Jakie warunki zapewniają pokój społeczny pomimo złego jedzenia? Jakie warunki skłaniają do buntu nawet wtedy, gdy jedzenie jest dobre? W obu przypadkach możesz modyfikować i rozszerzać teorię dodając nowe warunki. Tak w jednym, jak i w drugim więzieniu powinieneś rozmawiać z osadzonymi, strażnikami i naczelnikiem na temat jedzenia i ładu społecznego. Zapytasz też ogólnie o życie więzienne. Testując znaczenie, jakie ma „jakość jedzenia", nie powinieneś pomijać innych przyczyn buntów.

Teraz przychodzi czas na porównanie trzech środowisk. Przypuśćmy, że w obu więzieniach, w których doszło do buntu, dowiedziałeś się, że strażnicy i więźniowie okazywali sobie wzajemną pogardę na tle rasowym. W jednym bunt wyzwoliły wrzaski strażników: „Ty (epitet rasowy) zasługujesz na jedzenie tych pomyj". W drugim więzieniu strażnicy pozamykali cele na klucz krzycząc: „Powinniście (epitet rasowy) być trzymani w klatkach jak zwierzęta". Więźniowie odpowiedzieli: „P... (epitet rasowy)". Możesz więc przyjąć, że do buntu doszło na tle napięć rasowych, a nie kwestii jedzenia. Jedzenie było po prostu wymówką. Upewnisz się w tym przekonaniu, jeśli dowiesz się, że zanim doszło do buntu jedzenie się poprawiło, niewiele natomiast włożono wysiłku w usunięcie napięć na tle rasowym. Wniosek okaże się tym trafniejszy, jeśli stwierdzisz, że w więzieniu, w którym jedzenie było kiepskie, strażnicy współczuli więźniom, a antagonizm rasowy był względnie wyciszony.

Przykład powyższy pokazuje, że projektowanie wywiadów jakościowych z zamiarem zbudowania teorii wymaga nie tylko zastanawiania się nad tym, co chcesz testować, ale też gdzie i z kim powinieneś rozmawiać. Herbert Rubin przeprowadzając pierwsze wywiady w kilku miejscach dowiedział się, że organizacje działające na rzecz rozwoju lokalnego obawiały się bezpośrednio naciskać na rząd, by udzielił im

dodatkowego wsparcia. Obawiano się, iż rząd zemści się i obetnie wszystkie środki. Zamiast bezpośrednio wywierać nacisk na fundatorów, działacze lokalni stworzyli koalicję sił wystarczająco niezależnych od funduszów rządowych, aby zdecydowanie działać na rzecz uzyskania od miast finansowanego wsparcia dla swoich projektów.

Herbert chciał przetestować hipotezę, iż działacze lokalni stworzyli koalicję, by zabezpieczyć swoje zależne finansowo organizacje przed represjami. Wybrał więc kilka miast różniących się między sobą pod względem siły koalicji i przeprowadził wywiady. Wszędzie pytał o współpracę między agendami rządowymi i organizacjami lokalnymi. W miastach, gdzie koalicje były słabsze, rozmówcy twierdzili, że rząd ignoruje inicjatywy lokalne. Z kolei w miastach, gdzie koalicje były silne, rozmówcy radośnie opisywali kampanie koalicji, które odmieniły agendy rządowe. Wyłaniającą się teorię o znaczeniu koalicji podbudowały wywiady z różnych porównywalnych miejscowości. Proces budowania teorii na podstawie wywiadów, jej testowanie, modyfikowanie, kolejne testowanie trwa tak długo, aż w kolejnym cyklu zmiany będą już nieznaczne.

Budulcem teorii konstruowanych na podstawie wywiadów jakościowych są same wywiady, a nie pomysły czerpane z dzieł akademickich. Osiągając szersze znaczenie, nie przestają być mocno ugruntowane w doświadczeniach i sposobie rozumienia problemu przez respondentów. Stworzoną i przetestowaną przez siebie teorię możesz skonfrontować z literaturą przedmiotu, żeby zobaczyć, jak twoje badanie ma się do tego, co napisali inni. Jeśli jednak zaczniesz od teorii sformułowanych przez innych i będziesz je tylko testował — nie odkryjesz niczego nowego, jeśli w ogóle będziesz w stanie dostrzec cokolwiek. Teoria jako punkt wyjściowy badania może działać jak końskie okulary i ograniczać pole twojego widzenia. W wywiadach jakościowych końskie okulary są niedopuszczalne. Musisz mieć pełną swobodę podążania tropem uzyskiwanych informacji i tam, dokąd cię one prowadzą.

tłumaczył *Marcin Ciechomski*

Rozdział V

Społeczny kontekst ewaluacji

Ernest R. House

Demokratyzacja ewaluacji

Ernest R. House, „Democratizing Evaluation", rozdział z książki tegoż autora *Evaluating with Validity*, Sage Publications, Beverly Hills – London 1980.

Polityka wyboru

Historycznie rzecz biorąc, rządy demokratyczne w krajach zachodnich poprzedzało państwo liberalne. Społeczeństwo liberalne zrodziło się na gruncie koncepcji wolnego rynku, na którym jednostki konkurują ze sobą dokonując możliwie najlepszych transakcji w każdej niemal dziedzinie. Fundamentalną ideą liberalizmu była zasada wolnego wyboru jako podstawa organizacji systemu społecznego. „Demokracja liberalna to polityka wyboru" (Macpherson, 1966, s. 33).

„Istotą liberalizmu [...] jest wizja społeczeństwa składającego się z niezależnych, autonomicznych jednostek, które współpracują tylko wtedy, kiedy warunki współpracy zbieżne są z celami zainteresowanych stron. Paradygmatem takiej współpracy są stosunki rynkowe, co trafnie oddaje określanie przejścia od feudalizmu do liberalizmu jako zmiany polegającej na zastąpieniu «pozycji społecznej umową [społeczną]», proces ten, który w połowie XIX w. osiągnął apogeum, w toku późniejszego rozwoju uległ odwróceniu. Umowa stanowi model nawet dla tak nie sprzyjających jej relacji, jak relacje polityczne, w których prawo przyznaje korzyści jednym kosztem innych. Twierdzi się, że system jako całość jest korzystny dla wszystkich, zatem każdy z góry powinien go akceptować" (Barry, 1973, s. 166).

Liberalizm naprawdę wyzwalał. W okresie między XVII a XIX stuleciem oswobodził ludzi od rygorów obyczaju i autorytetu. Pozycję społeczną zastąpił bezosobową umową i relacjami rynkowymi. Chociaż produktem ubocznym wyzwolenia stały się wielkie nierówności mająt-

kowe, to traktowano je jako koszt nieunikniony i rozsądny. Nierówności nie były niczym nowym.

W społeczeństwach liberalnych nawet rządy, znalazłszy się w sytuacji rynkowej, stały się bardziej odpowiedzialne. Były postrzegane jako dostawca dóbr politycznych, spośród których konsument wybiera preferowane przez siebie. Konkurujące ze sobą partie polityczne stały się odpowiedzialne przed wyborcami dzięki procedurom elekcji. Zatem i rządy poczuły się odpowiedzialne za interesy elektoratu, ale elektorat nie był w żadnym sensie demokratyczny. Początkowo stanowiła go wąska elita bogaczy, mężczyzn z klas wyższych. Tylko tej elicie oferowano wybór rządu, a rząd liczył się jedynie z interesami elity.

Demokracja nadchodziła stopniowo i niepewnie. Wykorzystując mechanizmy wolności słowa i zgromadzeń, klasy niższe domagały się prawa głosu. Powołując się na równe prawa jednostek oraz na równość szans, żądały praw i wolności podobnych do tych, które zdobyły klasy wyższe. Zgodnie z logiką liberalizmu, byłoby nie *fair* odmawiać im udziału w wyborze rządu mającego reprezentować ich interesy. Po uzyskaniu prawa głosu klasy niższe wykorzystały to prawo nie do odrzucenia klas wyższych, lecz do zajęcia konkurencyjnej pozycji w służbie własnych interesów.

Liberalna forma demokracji istotnie różniła się od form nieliberalnych, które później pojawiły się w krajach komunistycznych i rozwijających się. Zarówno państwa komunistyczne, jak rozwijające się odrzuciły ideę społeczeństwa rynkowego. Społeczeństwa komunistyczne traktowały demokrację jako rządy ludu bądź rządy dla ludu. Rządzić miał proletariat, prowadząc do społeczeństwa bezklasowego. Pod jego rządami mało było miejsca dla wolności liberalnych.

Kraje rozwijające się często odrzucały zarówno rynek, jak idee klas, traktując demokrację jako rządy „woli ogólnej" wszystkich ludzi danego kraju. Zamiast wyborów, które kwestionowano, często w rządzie wykonującym wolę ogólną dominowała jedna partia, w tradycyjnym i nierynkowym społeczeństwie brakło miejsca dla konkurencyjnych partii politycznych. W tych trzech odmiennych pod względem politycznym społeczeństwach wybór miał różne znaczenie.

Tak więc równość była ideą fundamentalną, ale w demokracjach liberalnych wiązała się z ideami wyboru, konkurencji i społeczeństwa rynkowego. Wprawdzie owe liberalne idee demokratyzowały się na skutek rozszerzania praw obywatelskich na różne grupy oraz zwiększania zakresu decyzji publicznych to zasadą fundamentalną pozostawała zasada wyboru. Jako podstawę dokonywania ewaluacji w demokracjach

liberalnych proponuję zatem zasadę *równości wyboru*, zasadę uwzględniającą zarówno równość, jak wybór.

Chociaż równość wyboru jest zgodna z obiema fundamentalnymi zasadami demokracji liberalnej, to znacząco różni się od stanowiska czysto liberalnego, z którym aktualnie mamy do czynienia i w polityce, i w ewaluacji. Czysto liberalna zasada wyboru wymagałaby maksymalizowania możliwości wyboru bez względu na dystrybucję. Właściwa liberalnej demokracji zasada równych szans domaga się nie tylko maksymalnych możliwości wyboru, ale i tego, aby były one w jakimś stopniu równo dystrybuowane. Liberalna demokracja stwarza zatem warunki maksymalizacji nie tylko asortymentu, ale również jego dystrybucji.

Ewaluacja liberalnodemokratyczna

Koncepcja formalnej procedury ewaluacji publicznej, służącej decyzjom w kwestii programów publicznych i polityki, sama w sobie wywodzi się z liberalnego pojęcia wyboru. Jest zaś na tyle demokratyczna, na ile odnosi się do klas niższych i do każdego obywatela. Wybory dotyczące programów i polityki często mają za podstawę prywatne preferencje i prywatne interesy. Ewaluacji liberalnej (maksymalizowania możliwości wyboru) nic nie chroni przed wykorzystaniem do celów prywatnych.

Równie dobrze można przeprowadzać ewaluację dla króla, jak i dla celów publicznych, doktryna liberalna nie stwarza tu żadnych przeszkód. Prawdę mówiąc, można by popierać ewaluację w interesach prywatnych. W skrajnych przypadkach ewaluator dokonywałby ewaluacji w interesie kogokolwiek, kto by za nią zapłacił. Tego rodzaju liberalną ewaluację można oceniać z punktu widzenia jej przydatności dla odbiorcy, kimkolwiek by on był. Jednak ewaluacja programów publicznych zmienia wybór w przemyślaną *decyzję publiczną*. Ostatecznym kryterium ewaluacji nie może być służenie interesom prywatnym. Przydatność nie może być wartością pierwszoplanową. Ewaluacja liberalnodemokratyczna stawia interesy społeczne i publiczne przed prywatnymi interesami jednostek.

Wszystkie główne współczesne podejścia do ewaluacji zakładają wolność wyboru. Zakładają również indywidualistyczną metodologię, orientację praktyczną oraz idee wolnego rynku, na którym konsumenci będą „kupowali" to, co najlepsze. Wszystkie te podejścia są liberalne, ponieważ opierają się na idei społeczeństwa konkurencyjnego, indywidualistycznego, rynkowego. Istnieją jednak między nimi znaczące

różnice dotyczące tego, jakie wybory będą dokonywane, kto będzie ich dokonywał, na jakiej podstawie. Inaczej mówiąc, różnią się one swoimi demokratycznymi tendencjami.

W praktyce można dalej demokratyzować ewaluację — rozszerzając podlegające ewaluacji możliwości wyboru na wszystkie grupy, a ewaluacją publiczną obejmując wszystko, co jest przedmiotem wyboru na użytek publiczny. Można to osiągnąć uzupełniając rodzaj gromadzonych danych o nowe typy danych, ześrodkowując ewaluację na wyższym poziomie decyzji, rozszerzając grono odbiorców i grup odniesienia, a w końcu możliwości wyboru metod ewaluacji. Tego rodzaju działania zmierzałyby do wprowadzania zasady równości wyboru w ewaluacji, tak jak wprowadza się równość szans w szerszych kręgach społecznych.

Fundamentalne pojęcie równości oznacza traktowanie każdego (czy każdej desygnowanej grupy) jako odrębnej grupy odniesienia. Jeżeli ludzie traktowani są niejednakowo, to trzeba mieć w zanadrzu zasady czy mocne argumenty zdolne usprawiedliwić owo niejednakowe traktowanie (Barry, 1965). Ewaluacja z konieczności wychodzi od punktu widzenia jakiejś konkretnej grupy. Mówiąc bardziej precyzyjnie, grupa, na którą się powołujemy, jest grupą ludzi, których interesy nie mogą być ignorowane.

Grupa odniesienia musi być zidentyfikowana przed ewaluacją. Może ona — ale nie musi — być tożsama z adresatami ewaluacji. Na przykład ewaluator może kierować ewaluację do kluczowych decydentów w rządzie, ale jako grupę odniesienia traktować pozbawionych reprezentacji formalnej tych, których interesy są rozważane. Grupa odniesienia nie zawsze musi wiedzieć o ewaluowanym programie ani o samej ewaluacji.

Grupa odniesienia może być bardzo mała, jak rodzina kupująca samochód. Może być jednak bardzo duża, jak ogół konsumentów w przypadku niektórych ewaluowanych programów publicznych. Ponieważ fundamentalne pojęcie równości sugerowałoby traktowanie każdej osoby jako grupy odniesienia, stawiałoby to każdą ewaluację w obliczu dość trudnych wymagań. Jestem zdania, że w ewaluacji właściwą grupą odniesienia są ci wszyscy, których dany program czy polityka dotyczy. W sposób praktyczny zawęża to zakres rozważań i pozwala zastanowić się, które grupy dany program czy polityka w różny sposób dotyka.

Interesy

Aktualnie ewaluacja sama w sobie nie jest procedurą podejmowania decyzji społecznych, jest częścią procedury podejmowania decyzji społecznych dotyczących alokacji środków. Antycypuje pewien rodzaj sytuacji, w których podejmuje się decyzje społeczne. W tym procesie rola ewaluacji jest ograniczona do racjonalnej perswazji na podstawie wspólnych zasad i wartości.

Ewaluacja odwołuje się do faktów i argumentów celem oceny lub określenia wartości czegoś. Zestaw zasad lub wartości służy jako podstawa formułowania sądów. Kryteriów może być niewiele, nawet jedno, jak kryterium maksymalizacji przydatności u pragmatyków, lub możemy mieć do czynienia z mieszaniną kryteriów bez wcześniej określonych priorytetów, jak w przypadku zwolenników pluralizmu.

Przy formułowaniu sądów politycznych w demokracji liberalnej bierze się pod uwagę dwa rodzaje zasad: zasady dotyczące potrzeb i zasady dotyczące ideałów. Zasady dotyczące potrzeb nie ustosunkowują się do samych potrzeb, przyjmują życzenia i oczekiwania ludzi jako dane i podpowiadają jak można je zaspokajać lub jak dystrybuować ich zaspokajanie. Zasady dotyczące ideałów, przeciwnie, stwierdzają że pewne potrzeby są lepsze od innych i powinno się popierać ich zaspokajanie w działaniu publicznym. Na przykład potrzeby artystyczne mogą być uznane za ważniejsze niż potrzeby sportowe i z tego względu ich zaspokajanie zasługuje na publiczne poparcie (Barry, 1965).

Cechą charakterystyczną liberalizmu jest podejmowanie decyzji przede wszystkim na podstawie zasad dotyczących potrzeb. Potrzeby są akceptowane *en bloc* jako wartość pierwszoplanowa określająca cele publiczne. Publicznie moc obowiązującą mają tylko sądy dotyczące potrzeb. Oczywiście w życiu prywatnym jednostki mogą według własnej woli kierować się ideałami.

Liberalizm uznaje za prawomocne tylko te potrzeby, których nie wymusza państwo. Ponieważ nie uważa się, że jedni ludzie mają lepszy gust, a inni gorszy, jedni ludzie formułują trafniejsze sądy niż inni, więc dla celów politycznych opinie każdej jednostki traktuje się jako równorzędne. Jedynym kryterium dobra staje się zaspokajanie potrzeb (Barry, 1965). Dla kontrastu, przeciwnicy liberalizmu pewne potrzeby uznają za ważniejsze od innych i bardziej zasługujące na poparcie publiczne. Perfekcjoniści czy marksiści wierzą, że pewne potrzeby ludzkie powinny być zaspokajane, a inne nie. Punktem odniesienia sądów jest ideał. Czasem kwestionuje się autentyczność wyrażanych potrzeb. To stanowisko przemawia na rzecz ewaluacji nieliberalnej.

Tradycja liberalna jednakże respektuje wszystkie zgłaszane potrzeby, a kwestia polega albo na agregacji, albo na dystrybucji stopnia ich zaspokojenia. Z zasadami agregacji mamy do czynienia wtedy, gdy maksymalizuje się stopień zaspokojenia, ideą kluczową jest działanie w czyimś „interesie". Zasady dystrybucji uwzględniają takie pojęcia, jak słuszność, sprawiedliwość, uprawnienie, równość i wolność.

Podstawową zasadą agregacji jest zasada maksymalnego zaspokajania potrzeb. Mówiąc o kimś, że należy do grupy odniesienia danej ewaluacji, chcemy powiedzieć, iż ewaluacja ta w jakiś sposób bierze pod uwagę jego „interesy". Program czy polityka są prowadzone w interesie osoby, jeżeli zwiększają możliwości zaspokajania jej potrzeb. Mówiąc inaczej, „interesy" reprezentują uogólnione środki służące realizacji celów, jakie sobie stawia dana osoba. Prawdopodobnie można chronić lub pomnażać interesy osoby, np. jej bogactwo czy władzę, nie wiedząc, jaki jest jej cel ostateczny (Barry, 1965).

Ludzie mogą również mylnie pojmować swoje interesy. Mogą opowiadać się za programem czy polityką, które nie dadzą oczekiwanych przez nich rezultatów, w tej sytuacji ewaluacja jest szczególnie przydatna. Mogą też wybierać program czy politykę sprzeczną z własnymi interesami. Ogólnie rzecz biorąc, pojęcie „interesu" służy jako wskaźnik sumy satysfakcji i stopnia zaspokojenia potrzeb i z tego względu jest praktycznym kryterium ewaluacji.

Koncepcja „interesu" służy również do porównań. Program czy polityka leżą w czyimś interesie tylko w porównaniu z innym programem czy polityką bądź ze *status quo*. Ludzie mogą się zgadzać na rezultaty danego programu czy polityki nie godząc się co do tego, czy leżą one w interesie określonej grupy. Ponieważ mogą świadomie lub nieświadomie porównywać dany program z różnymi programami konkurencyjnymi, czyli przyjmować różny układ odniesienia, więc ostateczne oceny sukcesu danego programu mogą się różnić.

Wiele dyskusji w kwestii ewaluacji dotyczy kategorii porównawczych ewaluowanego programu czy polityki, a nie konkretnych rezultatów empirycznych. To, czy program lub polityka leży w interesie grupy odniesienia, zależy od wyboru układu odniesienia, a wybór ten często przesłaniają dyskusje metodologiczne.

Wielu czołowych teoretyków ewaluacji zakłada *implicite*, że celem ewaluacji jest pomoc decydentom lub samym grupom odniesienia w określeniu istotnych interesów poszczególnych grup (Cronbach, 1979). Jeżeli ewaluacja ma na tym polegać, to musi ona jakoś reprezentować interesy grup odniesienia. Ogólną zasadą jest, że interesy wszystkich

stron, których dotyczy program lub polityka, powinny być objęte ewaluacją, natomiast kwestią otwartą pozostaje, jak to robić.

W ewaluacji czysto liberalnej potrzeby traktowane są jako dane, a każda z nich zasługuje na zaspokojenie. Potrzeb nie traktuje się jako derywatów konkretnych grup bądź klas społecznych. Tak więc instrumentem „oceniania potrzeb" są zazwyczaj zgłaszane braki. Jeżeli oczekuje się od ludzi bezpośredniej oceny programu, to prosi się, aby dali bezpośredni wyraz własnym interesom. Sondaże dotyczące tych kwestii są prawie zawsze skrzywione klasowo ze względu na dobór respondentów i metodologię. Prosi się w nich o informacje operując narzędziami i kategoriami stosowanymi przez profesjonalistów z klasy średniej (House, 1973b).

Dla kontrastu, ewaluacja nieliberalna byłaby oparta na sądach wartościujących pewne potrzeby i interesy. Najbardziej powszechna dewiacja polega na tym, że profesjonaliści mają silnie ugruntowane przekonania w kwestii, kim powinna być osoba wykształcona. Czasem ten ideał jest w konflikcie z publicznie wyrażanym zapotrzebowaniem na edukację. Ewaluacja może inkorporować wiele ideałów środowisk profesjonalnych, ale ewaluator może się również kierować własnym rozumieniem interesów publicznych. Oczywiście, w wielu kwestiach sądy profesjonalistów mogą być trafniejsze od opinii wyrażanych przez ogół.

Inna możliwość dotycząca ideałów polega na promowaniu interesów poszczególnych klas, takich jak klasy niższe. Tego rodzaju ewaluacja byłaby oparta na sformułowanej *explicite* koncepcji ideału społecznego i obejmowałaby wartości i kryteria mające za podstawę określoną wizję społeczeństwa. Pewne rzeczy z góry byłyby określane jako ważne. Podejścia reprezentujące nieliberalny stosunek do ideałów nie muszą nakłaniać do dokonywania wyboru.

Propagowany tutaj pogląd głosi równe prawo wyboru. Ludziom powinien być dany wybór, aby żaden stan rzeczy nie był im narzucany, nawet w ich własnym interesie. Ale wybór powinien być udostępniony tak, aby wszystkie grupy i klasy społeczne miały pod tym względem równe możliwości. Dotyczy to szczególnie niższych grup społecznych, które powinny mieć możność samodzielnego dokonywania wyboru.

Interesy publiczne w ewaluacji

Interesy jednej lub kilku osób są to interesy „prywatne". O ile w przypadku ewaluacji reprezentującej interesy prywatne brak jakichkolwiek ograniczeń, o tyle w przypadku ewaluacji programów i polityki

publicznej trudno się rozeznać, jakim może ona podlegać ograniczeniom.

Polityka i programy publiczne prawie zawsze wiążą się z interesami dużych grup, takich jak osoby niepełnosprawne, poszkodowane czy utalentowane. Na poziomie lokalnym mogą to być interesy konkretnej szkoły lub grupy uczniów. Są to „interesy specjalne". Aczkolwiek wiele programów rządowych dotyczy interesów grup specjalnych, to częściej program czy polityka dotyczą interesu kilku grup łącznie. Programy dla osób niepełnosprawnych dotyczą nie tylko uczniów, ale i ich rodziców, nauczycieli a także pracodawców. Można powiedzieć, że te grupy mają „interesy wspólne". W demokracji liberalnej za większością programów rządowych stoją wyodrębnione interesy specjalne, ale są one rezultatem porozumienia zawartego w imię interesów wspólnych. Ewaluatorzy zwykle mają do czynienia z programami czy polityką reprezentującą kilka interesów.

Sama ewaluacja może być postrzegana jako polityka będąca pochodną interesu wspólnego. Strony zgadzają się co do tego, że wszyscy na ewaluacji zyskają. Jeżeli program odpowiada pewnym kryteriom, ewaluacja będzie wiążąca. Chociaż jednej ze stron rezultaty mogą się nie podobać, a jej interesy mogą w istocie ucierpieć, to o ewaluacji jako całości można powiedzieć, że ma na względzie interes wspólny.

Wreszcie istnieje interes publiczny. Jest to interes wspólny określonej publiczności, a „publiczność" to nie ileś określonych osób, lecz nieokreślona liczba „bezimiennych" jednostek (Barry, 1965). Interes publiczny może być (choć nie musi) silniejszy lub wyżej ceniony niż interes prywatny, specjalny czy wspólny, tyle że w różnym stopniu w nim uczestniczymy.

Załóżmy, że w przypadku dwu interesów specjalnych, np. interesu fabryki samochodów i interesu związków zawodowych, osiągnięto porozumienie co do interesu wspólnego, polegającego na podwyższeniu ceny samochodów. Co jednak z interesami tych ludzi, którzy nie stanowią strony umowy, zwłaszcza z konsumentami? Interes publiczny przesłania interesy tych bezimiennych członków publiczności, których mogą dotknąć skutki takich porozumień. W tym sensie przynależenie do publiczności nie jest czymś trwałym, lecz zmienia się w zależności od kwestii i kontekstu. Często publiczność jest utożsamiana z konsumentami, którzy na ogół chcą mieć więcej lepszej jakości dóbr za niższą cenę. Jednak nie trudno sobie wyobrazić taką sytuację, kiedy konsumenci są w konflikcie z innymi segmentami publiczności, choćby z rzecznikami ochrony środowiska.

Jak można więc zdefiniować interes publiczny? Decyzja nie jest tak prosta, jak w przypadku jednostki ważącej i bilansującej swój własny zysk netto — w grę wchodzi tu wiele różnych jednostek. W tym przypadku interes publiczny określa się często jako sumę wszystkich interesów wchodzących w rachubę. Stanowi on równoważnik maksymalizacji zaspokajania potrzeb bez względu na dystrybucję asortymentu. Alternatywą jest tylko popieranie interesów wspólnych (tam, gdzie istnieje zgoda co do interesów) i ignorowanie interesów rozbieżnych.

Różne podejścia do ewaluacji w różny sposób uwzględniają interesy. Podejścia te można podzielić generalnie na utylitarne i pluralistyczno-intuicyjne. Utylitaryści dążą do sformułowania ogólnej oceny użyteczności społecznej opartej na jednej dominującej zasadzie czy kryterium; jest to zasada agregatywna i polega na maksymalizacji zaspokajania potrzeb.

Utylitaryści dzielą się na odłam *menedżerski*, dla którego odbiorcami pierwszoplanowymi i/lub grupą odniesienia są menedżerowie, oraz odłam, dla którego odbiorcami i/lub grupą odniesienia są konsumenci. Odłam menedżerski, stosując analizę systemową, konstruuje takie wskaźniki społeczne, jak standaryzowane, punktowane testy, rodzaj pomiaru przydatności społecznej utożsamianej z interesem publicznym. Przyjmuje się, że liczba punktów uzyskanych dla każdej z grup społecznych oddaje stopień zabezpieczenia jej interesu. Gdy występuje konflikt interesów — wyniki sumuje się. Im więcej uzyskanych punktów, tym większa prawdopodobnie satysfakcja ze spełnionych żądań. Interes publiczny jest rozumiany jako uzyskanie wyższej punktacji przy najniższych kosztach. Warunek ten spełniają programy „najlepsze".

Podejście zoperacjonalizowanych zadań reprezentuje interesy definiowane przez cele. Jeżeli realizacja zadań jest zmaksymalizowana lub ustalone minimum zostaje osiągnięte, to służy to interesowi publicznemu. Ewaluator mierzy stopień realizacji zadań. W podejściu ukierunkowanym na podjęcie decyzji zakłada się, że interes publiczny reprezentuje oficjalne stanowisko decydentów. Zadaniem ewaluatora jest dostarczenie publiczności informacji celem podjęcia trafniejszych decyzji. Czy interes publiczny sygnalizują wskaźniki społeczne, publicznie sformułowane zoperacjonalizowane zadania czy decyzje urzędników państwowych — wszystko są to podejścia utożsamiające interes publiczny z oficjalną strukturą menedżerską.

Ewaluację opierającą się na tych podejściach różni to, czy traktuje ona menedżerów jako wyłącznie odbiorców czy także jako grupę odniesienia. Jeżeli menedżerowie stanowią grupę odniesienia, to ewaluatorzy służą raczej interesom menedżerów niż publiczności. Niestety, zdarza

się to zbyt często. Ewaluatorzy pozostają na usługach (specjalnych) interesów menedżerów, co jest nie do przyjęcia nawet dla teorii leżących u podłoża tych podejść.

Odłam ukierunkowany na konsumentów dzieli publiczność na producentów i konsumentów. Jego adresatami mogą być zarówno menedżerowie, jak konsumenci, ale grupą odniesienia są konsumenci. Ewaluator reprezentuje interesy konsumentów. Modelem jest Unia Konsumentów. Interes publiczny jest również często określany jako najlepszy produkt za najniższą cenę.

Pluraliści-intuicjoniści kierują się przy ewaluacji wieloma zasadami bez ustalania priorytetów. Dzielą się na zorientowanych na profesjonalizm i na uczestnictwo. Zorientowani na profesjonalizm wierzą, że najwięcej do powiedzenia powinni mieć ci, którzy w danej dziedzinie dysponują wiedzą i informacją. Interesom publicznym najlepiej służą decyzje ekspertów. W tym sensie do podejmowania decyzji powołane są grupy o najwyższym poziomie wiedzy.

Jak długo grupa nastawiona na profesjonalizm postrzega siebie jako działającą na rzecz publicznie określonych celów, konfliktu można uniknąć. Ale, oczywiście, grupa ta ma swoje niezachwiane przekonania dotyczące człowieka wykształconego. Ponieważ standardy ewaluacyjne często są standardami profesjonalnymi, mogą się one różnić od standardów publicznych. W tym sensie ewaluacja nastawiona na profesjonalizm może naruszać bezwzględne zasady zaspokajania potrzeb, dając pierwszeństwo ideałom. W dodatku standardy trudno oddzielić od interesów własnych profesjonalistów. Ostatnio wielokrotnie domagano się respektowania żądań publicznych w ewaluacjach tego rodzaju — począwszy od programów testowania kompetencji, a na komisjach weryfikacyjnych w medycynie skończywszy.

W końcu grupa zorientowana na uczestnictwo wierzy, że interesom publicznym najlepiej służą ludzie do pewnego stopnia uczestniczący w ewaluacji. Podejście *quasi*-sądowe korzysta z opinii tych, którzy tak czy inaczej w nim uczestniczą, często dosłownie cytuje je w raporcie końcowym. Podejście *quasi*-sądowe pozwala uczestnikom dzielić się na grupy prezentujące dowody i argumenty na wzór procedury sądowej. To podejście jest pluralistyczne nie tylko ze względu na posługiwanie się w ewaluacji różnymi kryteriami, jest pluralistyczne także w sensie politycznym, ponieważ bierze pod uwagę różne interesy.

Koncentracja władzy versus rozproszenie władzy

Pouczające jest porównanie dwu podejść do ewaluacji, krańcowo różnie definiujących interes publiczny: jedno stoi na stanowisku koncentracji władzy, drugie — na stanowisku rozproszenia władzy. Przykładem pierwszego podejścia jest polityka ewaluacyjna Biura Edukacji Stanów Zjednoczonych (U.S. Office of Education) w ostatnim okresie (McLaughlin, 1975; House, 1978b). Przyjmuje ono, że interes publiczny najlepiej może określić rząd centralny. Reprezentanci pochodzą z wyboru, a oni powołują administrację rządową, która określa linię postępowania, podejmuje decyzje i egzekwuje ich realizację. Może ona zatrudniać grupy ekspertów wspomagających jej przedsięwzięcia (Barry, 1965).

W praktyce rząd federalny często działa tak, jak gdyby chciał decydować, który program socjalny jest najlepszy, i stosuje analizę systemową do ewaluacji programów publicznych. Wskaźniki społeczne, w edukacji prawie zawsze standaryzowane testy, są używane jako wykładnik interesu publicznego. Władza i decyzje są skoncentrowane po stronie administracji centralnej, ewaluatorów i wskaźników.

Z drugiej strony, idea rozproszenia władzy powraca co najmniej od Hume'a, który twierdził, że „każdy człowiek może okazać się *łotrem* i we wszystkich swoich poczynaniach mieć na względzie wyłącznie interes prywatny". W świetle tej tezy nie można wierzyć nikomu ze sprawujących władzę. Muszą istnieć mechanizmy kontroli i równoważenia interesów, takie jakie przewidziano w konstytucji Stanów Zjednoczonych. Zasadę rozproszenia władzy w edukacji prawdopodobnie najlepiej reprezentuje „demokratyczna ewaluacja" MacDonalda, najbardziej demokratyczne ze wszystkich podejść zorientowanych na uczestnictwo.

W ewaluacji „demokratycznej" ewaluator zbiera opinie uczestników programu i wiernie prezentuje ich poglądy w raporcie sporządzonym jako studium przypadku. Ewaluator jest „pośrednikiem" w wymianie informacji dostępnej dla niespecjalistów. Reprezentuje on wiele różnych interesów i nie manipuluje informacjami (MacDonald, 1974). Ewaluator nie formułuje rekomendacji, lecz raczej prezentuje informacje adresatom, którzy wykorzystają je tak, jak uznają za stosowne. W sytuacji idealnej ewaluator przedstawia raport z ewaluacji ludziom, wśród których zbierał dane, i pozwala im sprzeciwić się włączeniu pewnych informacji do raportu.

Jak się wydaje, ewaluacja MacDonalda przewiduje taki sposób podejmowania decyzji, jak w demokracji bezpośredniej, kiedy to obywatele decydują o czymś w wyniku dyskusji *face-to-face*. To z kolei opiera się na klasycznym poglądzie liberalnym, zgodnie z którym jednostka bez-

pośrednio współdziała z innymi w dążeniu do własnych celów, bez pomocy instytucji czy rządu. Wybór indywidualny nie podlega ograniczeniom. Rząd, a w tym przypadku ewaluator, występują tylko w roli arbitra.

Ten pogląd na ewaluację budzi większe zaufanie ze względów politycznych i moralnych aniżeli podejście rządu federalnego. Surowo krytykowałem federalną politykę ewaluacyjną za nieuwzględnianie zróżnicowanych interesów, za postępowanie nie *fair*, za usterki techniczne (House i wspólaut., 1978). Nawet jeśli jest przeprowadzona dobrze, to supozycja, że wskaźniki społeczne rzeczywiście reprezentują interes publiczny, wzbudza wątpliwości. Częściej ewaluacja reprezentuje interesy grupowe (specjalne). Częste arbitralne akcje rządu na rzecz jednej strony nie podbudowują ewaluacji moralnie.

MacDonald w swoim podejściu intencjonalnie uwzględnia różne interesy, pozwala ludziom przedstawiać je i opiera się na idei wzajemnego porozumienia. Koncepcja demokracji MacDonalda jest bliska, jak się wydaje, rządzeniu „w porozumieniu z rządzonymi" lub, w tym przypadku, z ocenianymi. Są oczywiście inne koncepcje demokracji, w tym, zajmujące się podziałem dóbr między klasy społeczne. Podejście opierające się na porozumieniu bardziej przemawia do przekonania niż polityka federalna, mająca swoje sposoby narzucania działań bez porozumienia czy konsultacji.

Z drugiej strony, podejście opowiadające się za rozproszeniem władzy nie jest wolne od problemów. Można mieć wątpliwości, czy w postindustrialnym społeczeństwie masowym, złożonym ze sfragmentaryzowanych grup, demokracja bezpośrednia jest możliwa. Decyzje są zazwyczaj podejmowane albo na poziomie grupy, albo na poziomie centralnym, co może wymagać całkiem innych informacji. Oczywiście, można twierdzić, że nie na tych poziomach *powinno się* podejmować decyzje, że daje to złe rezultaty. Można również pytać, czy ubieganie się o zgodę każdej osoby na rezultaty ewaluacji w celu ustalenia interesu wspólnego uczestników programu nie zagraża interesom publicznym. Czy wartościowe informacje nie zostaną wykluczone? Czy ewaluacja nie okaże się tendencyjna na rzecz zachowania *status quo* i niepodejmowania bardziej radykalnych rozwiązań?

W praktyce, troska o zgodę każdej osoby zwiększa koszty negocjacji. Negocjacje z każdą osobą mogą pochłaniać ogromną ilość czasu i energii. W gruncie rzeczy, z tego powodu wiele tego rodzaju ewaluacji nie skończono we właściwym czasie lub nie dokończono w ogóle (Simons, 1977). Jeżeli koszty porozumienia są zbyt wysokie, ludzie odmawiają

udziału w negocjacjach i podejmują pracę niedostatecznie przemyślaną. Negocjowanie rezultatów również niesie ze sobą możliwość ich niewłaściwego wykorzystania przez osoby bez skrupułów (Elliott, 1977). Jednym ze sposobów redukowania kosztów negocjacji jest porozumienie z przedstawicielami, ale, oczywiście, to jest znowu powrót do koncentracji władzy.

Ostatnia trudność w tego rodzaju procedurach to koszty informacji (Barry, 1965). Przy gromadzeniu i przekazywaniu informacji od grup do decydentów przydać się mogą małe grupy dobrze artykułujące swoje interesy. Grupy duże, amorficzne, nie umiejące wyartykułować swoich interesów, jak zbiorowości należące do klas niższych, uznają, że sformułowanie ich punktu widzenia jest bardzo kosztowne. Z tego względu, nawet jeżeli ilość informacji dostarczanych decydentom zwiększa się, to informacje te mogą być tendencyjne z odchyleniem raczej w kierunku interesów grupowych (specjalnych) niż publicznych. Rozproszenie władzy może w rzeczywistości zwiększać wpływ interesów specjalnych (Barry, 1965).

Oczywiście, z wszystkimi podejściami do ewaluacji wiążą się trudności. Akcentuję te problemy, by pokazać, że nawet ewaluacja demokratyczna nie stanowi *panaceum*. „Demokratyczna" ewaluacja MacDonalda słabo reprezentuje interesy klas społecznych, ale wnosi wyjątkowo ważną ideę moralną: wzajemne porozumienie. Wzajemne porozumienie ujawnia się w wyborze. Opowiadam się za tym, aby to dochodzenie do zgody dotyczyło porozumienia się w sprawie ewaluacji raczej niż w sprawie jej wyników.

Zasady dystrybucji

Ludzkie pragnienia i interesy w sposób nieunikniony wchodzą ze sobą w konflikty i konieczne są pewne decyzje, aby te konflikty rozwiązywać. O ile zasady agregacji biorą pod uwagę tylko rozmiar zaspokojenia potrzeb grupy odniesienia, o tyle zasady dystrybucji, równość, słuszność i sprawiedliwość, bierze się pod uwagę celem osądzenia, jak zaspokajane są potrzeby poszczególnych jej członków (Barry, 1965). Lub, bardziej precyzyjnie, zasady stosuje się, aby osądzić *procedury* podejmowania decyzji o alokacji.

Procedury podejmowania decyzji społecznych, służące rozwiązywaniu konfliktów, nabierają w społeczeństwie liberalnym decydującego znaczenia. W liberalizmie niewiele jest spraw ważnych, co do których wszyscy się zgadzają, nie można zatem spodziewać się, że podstawą

decyzji będą wyniki czy jakiś idealny wzór dystrybucji, np. każdemu tyle samo. Słuszności decyzji muszą dowodzić raczej procedury, jedną z nich jest ewaluacja. Zasady dystrybucji zwykle odnoszą się raczej do procedur podejmowania decyzji aniżeli do samych decyzji. Ewaluacja jest dla mnie procedurą przygotowawczą, polegająca na stwierdzaniu faktów lub „ustalaniu wartości". Jest to procedura podejmowania decyzji tylko w tym sensie, że publicznie określa stan spraw, na podstawie którego mają być podjęte decyzje.

Fundamentalne pojęcie równości precyzuje, że wszyscy ludzie są traktowani jako należący do jednej grupy odniesienia. Jak dzielić rzeczy między ludzi, jeśli grupa odniesienia może być definiowana węziej lub szerzej? Jeszcze bardziej skomplikowaną sprawą jest dystrybucyjna równość. Jej „mocna" interpretacja wymaga, by dobra były dzielone równo bez względu na osobiste właściwości ludzi z grupy odniesienia. Interpretacja „słaba" poprzestaje na tym, że tylko *możliwości* zaspokojenia potrzeb powinny być równe. Forma słaba, raczej równych możliwości niż równego udziału, jest ideałem obowiązującym w społeczeństwach liberalnych. Z tego względu łatwiej znaleźć argumenty przemawiające za tym, że ewaluacja powinna brać pod uwagę interesy wszystkich wchodzących w grę grup, aniżeli za tym, że należy w równym stopniu czynić zadość wszystkim interesom. Stwarza się tylko możliwość.

Nierówność możliwości jest skutkiem niesprawiedliwych procedur lub warunków. Jeżeli nie przestrzega się odpowiednich procedur lub przy podejmowaniu decyzji bierze się pod uwagę nieodpowiednie czynniki, jak rasa czy klasa społeczna, wówczas procedura jest niesprawiedliwa. Prawidłowe procedury i warunki stwarzają prawdopodobnie równe możliwości.

Grupy mają prawo protestować nie tylko przeciw temu, że w ewaluacji nie są reprezentowane ich interesy, ale także przeciw tendencyjności spowodowanej zastosowaniem nieodpowiednich narzędzi, analiz itp. Poprawność procedur i warunków staje się wyjątkowo ważna w takich sytuacjach społecznych, kiedy ktoś ma prawo ubiegać się o zaspokojenie potrzeb, ale wynik zależy od tego, czy zwycięży w konkurencji z interesami innych.

To, jakie zasady dystrybucji zastosować, zależy od konkretnych procedur podejmowania decyzji społecznych. Barry (1965) wymienia siedem „czystych" typów takich procedur: dyskusja merytoryczna; walka: „na chybił trafił"; głosowanie; przetargi; konkurs; autorytatywne orzeczenie. Różnego rodzaju ewaluacje antycypują różne typy procedur podejmowania decyzji społecznych.

Dość trudno pokazać, jak ewaluacja może sprzyjać takiej procedurze decyzyjnej, jak walka. Można sobie wyobrazić, że jest to raczej metafora manewru politycznego i przetargów, a rzeczywista walka (narzucenie woli siłą) będzie rzadkością. Zwykle dochodzi albo do targu, albo do konkursu. Rzadkością jest również procedura podejmowania decyzji na chybił trafił. Pozostają przetargi, dyskusja merytoryczna, konkurs i autorytatywne orzeczenie.

Zarówno przetargi, jak dyskusje merytoryczne mogą być częścią negocjacji. Ale wąsko pojmowane przetargi to sytuacja, w której jedna strona oferuje drugiej korzyści w zamian za podjęcie pewnych działań lub ostrzega przed stratami, jakie spowoduje podjęcie działań przeciwnych. Wiele decyzji społecznych zależy od targów, ale nie ma to nic wspólnego z ewaluacją. Można sobie wyobrazić ewaluację mającą służyć utrudnieniu przetargów, ale trudno sobie wyobrazić jakiekolwiek moralne podstawy ewaluacji, w której planowo stosowano by utrudnianie i zachęty materialne. Rola ewaluacji w społecznym procesie podejmowania decyzji powinna być ograniczona do perswazji. Jest to kwestia zasadnicza, wytyczająca granicę pomiędzy polityką i moralnością. By mieć jakikolwiek autorytet moralny, ewaluacja nie może być postrzegana jako przetarg. Jej rezultatów nie można kupić, posługując się naciskiem czy zachętami materialnymi.

Pozostała dyskusja merytoryczna, konkurs i autorytatywne orzeczenie jako uprawnione procedury decyzyjne, które ewaluacja powinna antycypować. Dyskusja merytoryczna prowadzi do porozumienia w kwestii moralnego prawa dzielenia dóbr. Porozumienie osiąga się na podstawie tego, co stanowi interes wspólny, co zaspokoi najwięcej potrzeb itp. Nie ma w tym przypadku utrudnień ani zachęt. Jeżeli porozumienie zostaje osiągnięte, wówczas strony zmieniają swoje zamierzenia. Nawet jeżeli jedna ze stron będzie miała dość siły, aby zmienić sytuację, nie zrobi tego. Dla kontrastu, w przetargach każda strona stara się osiągnąć wszystko, co tylko może, stosownie do własnej siły. W wyniku dyskusji merytorycznej każda strona dochodzi do przekonania, że rozwiązanie jest właściwe. To dla ewaluacji wydaje się idealną procedurą podejmowania decyzji społecznych.

Czasem dyskusja merytoryczna będzie się wiązała z pytaniem, kto jest w czymś lepszy. By na nie odpowiedzieć, można zorganizować konkurs. Może to być rozwiązanie łatwiejsze, a jego wynik bardziej obiektywny. Takimi konkursami są ewaluacje posługujące się eksperymentami porównawczymi i projektami wariantywnymi. Ponieważ decyzja zależy od tego, jak wypadnie porównanie rezultatów, wszyst-

kie strony muszą wcześniej znać kryterium, na podstawie którego ich osiągnięcia będą oceniane. I, oczywiście, rezultat musi odzwierciedlać jakość, którą konkurs ma mierzyć (Barry, 1965). Ewaluatorzy nazywają to trafnością. Na przykład w ewaluacji *Follow Through*, mającej postać konkursu, było dużo wątpliwości co do tego, czy konkurs mierzy tę jakość, którą miał mierzyć. W tym sensie konkurs był nie *fair* (House i współaut., 1978).

I na koniec mamy autorytatywne orzeczenie. Gdy strony nie mogą się zgodzić, mogą poprosić kogoś, by rozstrzygnął dysputę, kogoś, kogo wszystkie uznają za uprawnionego do wydania orzeczenia. Strony wybierają arbitra. Przy autorytatywnym orzekaniu rezultaty, na podstawie merytorycznych walorów poszczególnych przypadków, ustala arbiter, podczas gdy w konkursie biorący w nim udział sami decydują o wyniku dzięki swoim umiejętnościom i kompetencjom. Procedura autorytatywnego orzekania potrzebuje arbitra, który ustala wynik, konkurs zaś potrzebuje sędziego, który stoi na straży przestrzegania zasad konkursu. W *Follow Through* rząd federalny działał w roli sędziego określającego zasady, a później w roli arbitra, który wskazał zwycięzcę. Są to role niespójne.

Chociaż większość ewaluatorów będzie utrzymywała, że dyskusja merytoryczna, w trakcie której wszystkie strony dochodzą do porozumienia, jest ideałem procedury podejmowania decyzji społecznych, to prawdopodobnie większość ewaluatorów nie unika orzeczeń autorytatywnych. Ewaluator, urzędnik państwowy, grupa profesjonalistów podejmują decyzje. Oczywiście, nawet w tych sytuacjach ogłoszenie decyzji poprzedza zwykle dyskusja merytoryczna — chociaż ograniczona do wybranej grupy profesjonalistów.

Przeciwnego rodzaju ewaluacja, która odwołuje się do jury lub sędziów, w rzeczywistości modeluje siebie na legalnych procedurach autorytatywnego orzekania, w ten sposób wymusza formalną decyzję. Innego rodzaju ewaluacje, w rodzaju studium przypadku, starają się zapobiegać autorytatywnym decyzjom części ewaluatorów wychodząc z dyskusją merytoryczną na szersze forum. Inni zachęcają ewaluatora do przedstawienia wyrazistej interpretacji wydarzeń w uspołecznionej procedurze podejmowania decyzji. W praktyce procedury decyzyjne są mieszaniną czystych typów.

Procedury podejmowania decyzji społecznych, jeżeli mają być uznane za prawomocne i stwarzające równe możliwości, muszą być postrzegane jako uczciwe. Prawidłowa procedura decyzyjna zwiększa równość szans. Uczciwa ewaluacja stwarza równość wyboru. Do samych

procedur decyzyjnych odnosi się zasada bezstronności (bezstronnego porównywania możliwości).

Sprawiedliwość proceduralna wymaga rzeczywistego przestrzegania przewidzianych formalności. W uczciwej dyskusji merytorycznej na ogół nie ma takich formalności (aczkolwiek mogą być reguły dyskusji) z wyjątkiem tego, że nie można stosować przymusu ani zachęt. W przeciwnym razie procedura zmienia się w zwykły przetarg. W uczciwym konkursie muszą być przestrzegane zasady, a autorytatywne orzeczenia muszą być poprzedzone ustaleniem odpowiednich na tę okoliczność procedur. Ogólna sprawiedliwość doskonali się poprzez reguły sprawiedliwości proceduralnej i wymaga, by zaangażowane strony miały odpowiednią pozycję startową. Również przy autorytatywnym orzekaniu ma znaczenie, czy arbiter przestrzega stosownych zasad. Przestrzeganie zasad jest warunkiem konsekwencji postępowania, podstawowego wymogu sprawiedliwości.

Wszystko to, co składa się na uczciwość i bezstronność, prowadzi do „prawidłowych" rezultatów procedury decyzyjnej. Razem stanowią one o równości szans. Ogólnie, bezstronność procedur decyzyjnych jest sprawą niezwykłej wagi w społeczeństwie liberalnym, ponieważ tym, co ludzie uzgadniają, są procedury a nie ich rezultaty. Ponadto, jeżeli ktoś zgadza się na daną procedurę i akceptuje korzyści, jakie ona zapewnia, to jest w porządku tylko wtedy, jeżeli postępuje zgodnie z tą procedurą, nawet gdy nie jest to korzystne dla niego osobiście (Rawls, 1994).

Ewaluacja jako procedura podejmowania decyzji moralnych

Ewaluację można rozumieć jako procedurę podejmowania decyzji społecznych, aczkolwiek sądzę, że należałoby ją raczej traktować jako składową procedur decyzyjnych. Ewaluacja rzadko decyduje o kwestiach społecznych, ale może decydować. Najczęściej zasila ona procedury decyzyjne dotyczące faktycznej alokacji dóbr. Ważny jest sposób, w jaki ewaluator przewiduje te procedury podejmowania ostatecznych decyzji.

Jako ta część łańcucha decyzyjnego, która polega na stwierdzaniu faktów i „ustalaniu wartości", sama ewaluacja antycypuje określone procedury i przyjmuje ich cechy. Na ile jest organizowana jako dyskusja merytoryczna, jako konkurs czy autorytatywne orzekanie, na tyle jest przedmiotem analogicznych uwag o bezstronności. Szczególna troska

o metodologię, o „należyte postępowanie", odzwierciedla przekonanie o decydującym znaczeniu właściwych procedur decyzyjnych.

Obecna definicja „właściwej" metodologii wiele zawdzięcza środowiskom profesjonalistów i techników. Za pomocą swojej metodologii technicy starają się wyeliminować „odchylenia", to znaczy zapewnić powtarzalność wyników. Jednak powtarzalność wyników nie gwarantuje ewaluacji demokratycznej i moralnie akceptowalnej.

Z demokratycznego i moralnego punktu widzenia powtarzalność rezultatów może być wyjątkowo niewskazana. Metodologia pozytywistyczna, bazująca na metodologii nauk przyrodniczych, prowadzi tu ewaluatora na błędną drogę. Bywa, że urzędnik państwowy, antycypując poważny sprzeciw wobec jego autorytatywnego sposobu podejmowania decyzji, usilnie namawia ewaluatora do sięgnięcia po „najmocniejsze", najbardziej naukowe metody, celem poparcia autorytetu rządu. Jednak zaangażowany ewaluator może dojść do wniosku, być może za późno, że metoda obliczona na uzyskanie powtarzalnych wyników jest w tych okolicznościach niewłaściwa.

Teoria demokracji zakłada, że polityka, jeśli ma być akceptowana, musi odzwierciedlać interesy członków społeczności (Care, 1978). Demokratyczne procedury decyzyjne są pomyślane tak, aby uwzględniały i wyrażały te interesy. Dotyczy to również ewaluacji jako procedury decyzyjnej w społeczeństwie demokratycznym. Ale nawet to może nie wystarczać. Czy procedura dzieląca coś równo między wszystkie osoby lub automatycznie przyznająca większość zasobów klasom niższym będzie uznana za procedurę poprawną? Sądzę, że nie, przynajmniej nie w społeczeństwie liberalnym, choć można wyobrazić sobie społeczeństwo, w którym takie procedury uznaje się za słuszne.

Nie każda procedura i nie każda polityka odzwierciedlająca interesy członków społeczności jest z definicji moralna (Care, 1978). Może być bowiem rezultatem przetargów lub kompromisów, w których zastosowano zachęty materialne czy naciski. W ewaluacji, która z założenia ma dostarczyć podstaw do dyskusji merytorycznej, jest to moralnie nie do przyjęcia. Jak zatem konstruować bezstronną i moralnie akceptowalną procedurę ewaluacyjną?

Odpowiedź znowu zawiera się w wyborze, w daniu zainteresowanym stronom możliwości wypowiedzenia się w samej ewaluacji. Uczestniczenie w planowaniu procedury ewaluacji jest sposobem zapewnienia akceptowalności moralnej (Care, 1978). Traktuję akceptowalność moralną jako ściśle związaną z autonomią moralną i porozumieniem. Nikt nikomu nie może narzucać swojej woli. Uważam, że dobrowolne

porozumienie co do procedury decyzyjnej zobowiązuje wobec niej moralnie, nawet jeżeli komuś jej rezultaty się nie podobają. Jednak nie każde porozumienie jest wiążące moralnie.

Care (1978) wprowadził pojęcie „moralnej akceptowalności procedury", zgodnie z którym uczestniczenie w porozumieniu powoduje że rezultaty porozumienia są moralnie akceptowalne. Jeżeli osoby zaangażowane w porozumienie spełniają pewne kryteria i warunki, rezultaty są moralnie akceptowalne — jak przy przestrzeganiu zasad gry.

Innymi słowy, strony włączone do ewaluacji wcześniej dochodzą do porozumienia, jak ewaluacja będzie przebiegała. To porozumienie będzie podstawą oceny ewaluacji jako przeprowadzonej uczciwie. Jednak nie każde porozumienie w sprawie ewaluacji gwarantuje taką ocenę. Porozumienie może być tylko przetargiem. By zapewnić akceptowalność moralną, porozumienie musi odpowiadać pewnym warunkom, dającym gwarancję, że uczestnicy będą w stanie zidentyfikować swoje interesy w danej sprawie. Jest dwanaście takich warunków koniecznych, aby porozumienie ewaluacyjne było uczciwe.

Czy warunkom tym można kiedykolwiek zadośćuczynić w pełni to godna uwagi kwestia praktyczna. W każdym jednak przypadku stanowią one ideał moralny, punkt odniesienia przy ocenie ewaluacji. Można powiedzieć, że porozumienie ewaluacyjne jest uczciwe i moralnie wiążące w takim stopniu, w jakim spełnia te warunki. Należy sądzić, że jest to prognostyk uczciwości ewaluacyjnej jako takiej.

Tak więc uczciwe porozumienie ewaluacyjne stanowi sposób na zapewnienie równego wyboru. Porozumienie może być zupełnie nieformalne, nie zapisane, chociaż jest komponentem idei „umowy". W społeczeństwie postrzeganym jako zbiór autonomicznych jednostek współpracujących tylko ze względu na własne cele, a jest to istota liberalizmu, umowa jest środkiem, za pomocą którego jednostki dobrowolnie przyjmują na siebie pewne zobowiązania. Umowa społeczna jest tą trwałą podstawą moralną, którą liberalizm ma do zaoferowania.

tłumaczyła *Eugenia Potulicka*

Eleanor Chelimsky

Praktyka ewaluacyjna dzisiaj
Kilka refleksji

Eleanor Chelimsky, *Where We Stand Today in the Practice of Evaluation: Some Reflections*, „Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization" 1995, nr 3.

Wydaje się, że badania ewaluacyjne wkroczyły w niezwykle obiecującą i interesującą fazę swojego rozwoju. Pozostawiając za sobą raporty z lat osiemdziesiątych, wskazujące na regres tej dyscypliny, dzisiaj, w roku 1994, obserwujemy odradzanie się profesji ewaluacyjnej i jej ekspansję na nowe obszary zainteresowań tematycznych, większy zasób doświadczeń i praktyki badawczej, i — przynajmniej w Stanach Zjednoczonych — pewien rodzaj legislacyjnego wsparcia (wraz z daleko idącą inicjatywą wykonawczą nazwaną National Performance Review).

Równocześnie ewaluatorzy w innych częściach świata rozwijali *swoje* metody badawcze i swoje sposoby wykorzystywania wyników badań przy podejmowaniu decyzji związanych z polityką społeczną — rzecz nie do pomyślenia jeszcze dwadzieścia lat temu. Dawne metody, takie jak badania kwestionariuszowe, eksperyment badawczy lub analiza „koszt — efekt" zostały zaadaptowane i twórczo przetworzone tak, aby można było je zastosować w różnych instytucjach i dla różnych potrzeb we Francji, Szwecji, Kolumbii, Pakistanie, Tunezji itd. Co więcej, wszędzie tam, gdzie dorobek badawczy może być poddany syntezie (w Wielkiej Brytanii, Kanadzie, Holandii), zaczęły być szeroko stosowane pewne współcześnie rozwijane metody, jak na przykład synteza badawcza (*research synthesis*) (znana również jako metaanaliza).

Równolegle z rozpowszechnianiem się tych metod powstają stowarzyszenia ewaluacyjne: w Australii, w Ameryce Południowej i Środkowej, w Europie. Bank Światowy i inne organizacje międzynarodowe odegrały znaczącą rolę w doskonaleniu powiązań między planowa-

niem, programem lub zakładaną efektywnością, a rozdziałem zasobów finansowych. W Stanach Zjednoczonych biura Inspektora Generalnego* powołują coraz to nowe oddziały zajmujące się ewaluacją. Urzędy skarbowe na całym świecie zaczęły uwzględniać ewaluację, którą nazywają „kontrolą wydajności" (*performance audit*), kontrolą „opłacalności" (*value-for-money audit*) lub kontrolą ogólną (*comprehensive audit*), jako jeden z kluczowych obszarów ich zainteresowania. Z zadowoleniem stwierdzam, że osoby zajmujące się księgowością finansową coraz częściej postrzegają ewaluację jako coś *swoistego*, odmiennego od sprawozdania finansowego.

Skoro jednak ewaluacja w szybkim tempie przyjmuje się na całym globie, powstaje pytanie co do rzeczywistych możliwości jej adaptowania — już nie do wielkich organizacji *międzynarodowych*, gdzie kontrola wydatków jest warunkiem przeżycia, ale do tak bardzo różnych instytucji i temperamentów. Ostatecznie narody bardzo różnią się między sobą historią, stylem sprawowania władzy, wartościami i kulturą. Pamiętajmy chociażby, że w Europie filozofowie są znakomitościami, podczas gdy w Ameryce znakomitości są filozofami. Tak więc zasadne jest oczekiwanie, że *również* ewaluacja będzie różnić się zależnie od miejsca, w którym jest realizowana. Dotyczy to sposobu jej organizacji, sposobu jej konstruowania oraz sposobu wykorzystania jej wyników.

W 1992 r. w Waszyngtonie na międzynarodowym kongresie przedstawicieli instytucji zajmujących się sprawozdawczością można było obserwować ogromną różnorodność koncepcji funkcji sprawozdawczości. Obecnie obserwujemy podobną różnorodność odnośnie do ewaluacji — chociażby w zakresie rozwiązań organizacyjnych: powiedzmy w Kanadzie ewaluacja jest w pełni zdecentralizowana, każda agenda rządowa organizuje ewaluacje, przesyłając protokoły do odpowiedniego ministra; we Francji istnieje centralna rada naukowa uprawniona do zatwierdzenia lub odrzucenia danego projektu ewaluacyjnego; w niektórych zaś organizacjach międzynarodowych często występuje dwuszczeblowy system sprawozdawczości kładący nacisk na uczestnictwo i zachowanie ciągłości.

Na Międzynarodowej Konferencji Ewaluacyjnej (International Evaluation Conference), planowanej na listopad 1995 r. w Vancouver, będziemy mogli zyskać znacznie lepsze rozeznanie, jak narody, tak różne, jak chociażby rozwijające się i rozwinięte, wyobrażają sobie i organizują ewaluacje. Jak dobierają metody i kryteria, jak gromadzą dane i jak

*Inspector Generals' Offices — urząd o zbliżonych kompetencjach jak w Polsce Najwyższa Izba Kontroli — *przyp. tłum.*

wykorzystują wyniki badań. Mówiąc krótko, będziemy mogli zobaczyć nie tylko, jakie mogą być sposoby, za pomocą których narody i międzynarodowe organizacje przystosowują ewaluację do swoich potrzeb i instytucji, lecz przede wszystkim, jak ewaluacja *powinna się* zmieniać, aby wnioski z niej płynące mogły mieć znaczenie w określonym kontekście kulturowym i politycznym.

Fakt, że organizujemy światową konferencję poświęconą ewaluacji właśnie *teraz*, jest niezwykle wymowny; świadczy o znaczeniu jej nowych wymiarów. W chwili obecnej ewaluatorzy czują się gotowi do zajęcia się znacznie szerszym spektrum tematów, związanych z interesami o bardziej globalnym charakterze niż te, które do tej pory były przedmiotem ich studiów. Możemy obecnie odnotować rosnącą liczbę metodologicznie udanych ewaluacji, które wykorzystano w polityce społecznej z dużym sukcesem. Możemy również znacznie lepiej przewidywać i omijać niektóre z tych przeszkód, które zawadzały nam w przeszłości.

W tym kontekście nasze oczekiwania dotyczące tego, co ewaluacja może zdziałać w danym otoczeniu politycznym, stały się znacznie bardziej realistyczne, a nadto lepiej rozumiemy znaczenie nie tylko bycia wiarygodnym, lecz przede wszystkim bycia *postrzeganym* jako ktoś, komu można zaufać.

Chciałabym zatem wymienić trzy nowe wymiary ewaluacyjne:

- po pierwsze: czego — w sposób uzasadniony — możemy oczekiwać od ewaluacji;
- po drugie: jakimi zmianami charakteryzują się metody badawcze obecnie przez nas stosowane;
- po trzecie: czego nauczyliśmy się o wiarygodności ewaluacji i jej znaczeniu dla aplikacji wniosków z niej płynących.

Co ewaluacja może, a czego nie może

Logicznie rzecz biorąc należałoby uznać, iż ewaluacja powinna być użyteczna dla osób podejmujących ważne decyzje publiczne. Od czasów Johna Stuarta Milla przyjęto, że rolą nauk społecznych jest wspieranie debaty zmierzającej do rozeznania rzeczywistych lub możliwych efektów prowadzonej polityki. Ewaluacja robi to na wiele sposobów, przede wszystkim poprzez:

- syntezę tego, co aktualnie wiemy o danym problemie społecznym i o propozycjach lub programach zmierzających do jego rozwiązania;

• demistyfikowanie stereotypowych prawd lub popularnych mitologii związanych albo z problemem, albo ze sposobami jego rozwiązywania (możemy tutaj przywołać starą ideę *Entzauberung*, „odczarowywania" świata Maxa Webera);

• uzyskiwanie nowych informacji o programie lub skutkach polityki;

• wyjaśnianie aktorom sceny politycznej wniosków płynących z informacji uzyskanych w wyniku ewaluacji.

Przy tak wyraźnie określonej funkcji ewaluacji dostarczanie obiektywnych informacji osobom zajmującym się polityką, traktowane jako pomoc przy podejmowaniu decyzji, powinno być działaniem rutynowym. Lecz faktycznie *nie jest* to regułą. Doświadczenie amerykańskie i to, co obserwowałam w praktyce przez ponad dwadzieścia pięć lat pracy zawodowej, nie świadczy, aby tak było. Przeciwnie, często między środowiskiem podejmującym decyzje polityczne a procesem ewaluacji tak dalece brak łączności, że nie następuje żaden przepływ informacji. Rząd lub wysoko postawieni funkcjonariusze organizacji międzynarodowych bywają tak głęboko zaangażowani w jakieś szczególne działania polityczne, że niezależny ewaluator ma małe szanse odegrania jakiejś istotniejszej roli, a nawet niewielką możliwość przedstawienia swoich uwag.

Mam tu na myśli, na przykład, amerykańską politykę w czasie wojny w Wietnamie, kiedy to rezultaty ewaluacji, specjalnie zamawiane przez Departament Obrony, były całkowicie ignorowane przy podejmowaniu decyzji. Jest to problem wyników każdej ewaluacji, które muszą stawić czoło nieprzepartej sile celów politycznych lub instytucjonalnych w rodzaju zachowania kontroli, twarzy albo projektowania własnego *image'u*. W takich generalnej wagi przypadkach i we wszystkich specyficznych dziedzinach, gdzie debata polityczna jest bardzo gorąca — jak aborcja czy przestępczość w Stanach Zjednoczonych, planowanie rodziny w Chinach lub sytuacja kobiety w Afryce bądź problemy Bliskiego Wschodu — dzisiejsi ewaluatorzy napotykają mur niechęci, nie tak znowu różny od tego, z jakim miał do czynienia Alexis de Tocqueville w 1847 r., kiedy to przekonywał deputowanych francuskiego Zgromadzenia Narodowego, że rewolucja stoi za drzwiami, której to myśli właśnie najbardziej nie chcieli dopuścić do świadomości. Z podobną sytuacją miał do czynienie Max Weber, gdy ostrzegał niemieckich wojskowych przed reakcją amerykańską na ataki łodzi podwodnych podczas I wojny światowej.

Innym powodem, dla którego ewaluacja często nie może być pomocna, jest fakt, że w przypadku wielu problemów społecznych brak

podstawowych badań, które wskazywałyby, jak należy podejść do problemu, a przynajmniej takich, które wskazywałyby kryteria wyboru metod jego rozwiązywania. Mam tu na myśli wątpliwą użyteczność oceny wpływu podejmowanych działań — powiedzmy, na rozładowanie tłoku ulicznego — w sytuacji, gdy hipotetyczny związek między interwencją a jej możliwymi skutkami nie tylko nie został oszacowany, ale nawet nie stwierdzono jego istnienia. Kiedy indziej nie wiemy, które *oczekiwania* są zasadne: do jakiego stopnia powinniśmy uważać za pewne, że porady psychologiczne redukują depresję i rozpacz pacjentów chorych na AIDS? Jakie jest prawdopodobieństwo, że szkolenie zawodowe pomoże znaleźć pracę? Dlaczego powinniśmy oczekiwać, że ogłoszenie zasady „trzy razy kogoś uderzysz i wylatujesz" spowoduje zmniejszenie przestępczości? W tych wszystkich przypadkach nie wiemy, nie mamy żadnej podpowiedzi, gdyż nie dysponujemy konkretnymi badaniami. Tymczasem pewne programy zaczęto już realizować, niektóre z nich są bardzo kosztowne. Tak więc potrzebujemy podstawowych badań, *zanim* określimy ogólną politykę, która będzie wdrażana w skali całego kraju lub nawet całego świata.

W przypadku, gdy przeprowadzenie ewaluacji *jest* wykonalne i użyteczne, powstaje pytanie, jakich decyzji powinna ona dotyczyć? To, czego się nauczyliśmy, wskazuje, że istnieją trzy typy decyzji, którym ewaluacja może się dobrze przysłużyć i wyraźnie zwiększyć szanse ich sukcesu. Pierwszym typem decyzji jest sformułowanie lub określenie *nowych* działań lub programów społecznych. Z jakich względów ewaluacja może być tutaj użyteczna? Są trzy powody. Ewaluacja wcześniejszych inicjatyw może uchronić osoby zajmujące się polityką społeczną przed wyważaniem otwartych drzwi i późniejszą publiczną kompromitacją, dostarczając już na wstępie ich działań dowodów, że nawet najbardziej obiecujące i nowatorskie propozycje nie gwarantują realizacji danego programu. Taka ewaluacja umożliwi działaczom społecznym tworzenie programu zbudowanego na tych wcześniejszych doświadczeniach, które *okazały się* efektywne lub co najmniej obiecujące. Rolą ewaluatora jest tutaj dostarczenie decydentom najlepszych dostępnych informacji o poprzednich doświadczeniach związanych z danym problemem i o strategiach jego rozwiązywania.

Drugim właściwym sposobem wykorzystania ewaluacji przez tych, którzy decydują o polityce społecznej, jest posłużenie się nią do monitorowania realizacji programu oraz określenia, czy założenia, które legły u podstaw danej polityki lub danego programu, okazały się słuszne, a przynajmniej czy nie okazały się *błędne*.

Na koniec, decydenci używają ewaluacji do ustalenia rezultatów swoich inicjatyw, konsekwencji zarówno bieżących, jak długofalowych skutków własnych pomysłów. Nie tylko dlatego, że chcą mieć wpływ na ewentualne modyfikacje danego programu, lecz również dlatego, aby móc bronić własnego stanowiska w kolejnych kampaniach wyborczych.

My, ewaluatorzy, możemy obecnie wspomagać te trzy sposoby podejmowania decyzji w ramach szerokiego spektrum działań politycznych — nie ograniczając się wyłącznie do polityki społecznej jak to miało miejsce w przeszłości. Możemy wspomagać decyzje dotyczące obronności, komunikacji, rolnictwa, ochrony środowiska, a więc wszystkie decyzje ważne społecznie, choć oczywiście tylko wtedy, gdy zostały przeprowadzone badania podstawowe dotyczące danego problemu, a klimat polityczny i intelektualny pozwala osobom decydującym zarówno na *zapoznanie się*, jak i *wykorzystanie* wyników ewaluacji.

Ta ostatnia uwaga jest ważna nie tylko ze względu na implikacje związane z poprawą i ulepszeniem polityki lub programu: oczywiście, jeżeli politycy nie będą chcieli nawet wysłuchać wniosków płynących z przeprowadzonej ewaluacji, szanse na to, że sugestie dotyczące konkretnych działań zostaną wykorzystane w praktyce, są niewielkie. Jednakże groźniejszą i demobilizującą sytuacją jest brak gotowości do dialogu. Ewaluatorowi bardzo trudno szczerze dzielić się swoją wiedzą, gdy nie ma do tego odpowiedniej motywacji. Thomas Huxley napisał, iż uśmiercanie pięknych hipotez przez brzydkie fakty jest wielką tragedią nauki. To prawda, lecz gdy badacz pracuje w środowisku bardzo upolitycznionym, to uśmiercanie pięknych hipotez może nawet nie zostać odnotowane, o opłakiwaniu nie wspominając: nikt nawet o nich nie *usłyszy*, są bowiem nadmiernie uwikłane w politykę.

Trudność, z jaką mamy tu do czynienia, polega na tym, że podstawowym celem ewaluacji jest powiedzenie prawdy o tym, czego się dowiedzieliśmy. Pomijając już względy etyczne — instytucjonalnym *obowiązkiem* rzetelnej ewaluacji jest odwaga powiedzenia tego, czego osoby zamawiające ewaluacje mogą nie chcieć usłyszeć. Oczywiście oddziałuje ona *post factum*, dzięki wnioskom, ale pozostawia także swój ślad poprzez zwykłą obecność: już sama wiedza o pojawieniu się ewaluatora — lub zapowiedź, że może się zjawić w każdej chwili — sprawia, że inicjatorzy nowego programu stają się bardziej rozważni i umiarkowani w swoich zapędach, bardziej ostrożni w rachunkach, *znacznie* bardziej świadomi ponoszonych kosztów i — co rzadko zdarza się w polityce — skłonni zawracać z obranej drogi. Kiedy jednak ewaluator straci nie-

zależność, ewaluacja sama w sobie traci sens, swoją wartość doradczą, swoją wiarygodność i — oczywiście — swoją siłę.

Nie twierdzę, że badania są niepotrzebne, że nie należy zawracać sobie nimi głowy przy kwestiach doraźnych lub kontrowersyjnych. Po pierwsze, pewne badania są ważne przez wzgląd na nie same, niezależnie od tego, czy zostaną natychmiast wykorzystane. Po drugie, sytuacja polityczna zmienia się i wnioski z badań, które są trudne do przełknięcia dla jednej grupy decydentów, dla ich następców mogą okazać się manną z nieba. Nie ulega jednak wątpliwości, iż zawsze może być trudno znaleźć osoby skłonne wysłuchać wniosków niezgodnych z intencją aktualnej opcji politycznej. Tak więc, jeśli chcemy robić ewaluacje tak, by zapewnić im sukces, powinniśmy koncentrować uwagę na tym, co jest naprawdę ważne, starając się jednak uniknąć kwestii tak obciążonych emocjonalnie, że poważna debata nad ich wynikami nie jest możliwa.

Z własnego doświadczenia wiem, że nawet wtedy, gdy możemy posłuchać tej rady, ewaluatorzy zawsze mogą natknąć się na zaskakujące naciski, zdolne pokrzyżować najbardziej rutynowe, konwencjonalne i niekwestionowane ewaluacje. Naciski mogą zmierzać do całkowitego wyeliminowania ewaluacji, do zmiany zadawanych pytań na bardziej „nieszkodliwe", do stonowania opinii lub wyeliminowania osobistych wniosków i konkluzji. Mogę powiedzieć, że te naciski będą się nasilać pod koniec badań — zarówno jeśli chodzi o ich liczbę, jak i ich częstotliwość — proporcjonalnie do tego, jak poważnie traktuje się społeczne i polityczne konsekwencje zarówno ewaluacji, jak i innych form analizy.

Przeciwstawienie się takim naciskom wymaga do ewaluatora odwagi, a także wystarczającej kompetencji, aby obronić to, co najczęściej staje się celem atakujących „nieprzyjemne" wnioski — to jest metody badawcze użyte przez ewaluatora. Na szczęście jesteśmy do tego lepiej przygotowani niż dwadzieścia lat temu. Spostrzeżenie to przybliża mnie do drugiego tematu, który chciałabym poruszyć.

Zmiany w metodach ewaluacji

W latach dziewięćdziesiątych patrzymy na metody ewaluacyjne nieco inaczej niż robiliśmy to wcześniej — prawdopodobnie dzięki wielu praktycznym doświadczeniom ewaluacyjnym, które mogą obecnie uzupełnić wcześniejsze perspektywy teoretyczne. To spotkanie teorii z praktyką wywołuje reakcję podobną w pewien sposób do tej, jaka jest rezultatem zetknięcia się dwóch substancji chemicznych: *obu* składniki

ulegają zmianie, ale w sposób nie do końca oczywisty. Wciąż istnieją liczne obszary, na których — jak sądzę — czujemy się pewnie po prostu dlatego, iż *praktyka* badawcza uległa zmianie. Pozwólcie mi wskazać pięć takich obszarów.

Pierwszy, związany jest z tym, że dziś wiemy już, iż każda metoda badawcza stosowana w naukach społecznych ma swoje słabe strony, lecz słabości te niczego nie przesądzają. Nauczyliśmy się, że zamiast walczyć, starając się wykazać wyższość jednej metody nad innymi, trzeba raczej *wykorzystywać kilka metod łącznie*, aby zalety jednej rekompensowały niedoskonałości pozostałych. Oznacza to, że obecnie *mniej* myślimy o zaletach jednej metody na tle pozostałych, a *więcej* uwagi poświęcamy temu, czy i jak je użyć jednocześnie, aby uzyskać jak najbardziej wiążące wnioski.

Druga ważna zmiana dotyczy skoncentrowania się w większym stopniu na pytaniach, na które chcemy uzyskać odpowiedź. Nauczyliśmy się, że wybór metod badawczych (a także narzędzi i mierników) zależy znacznie bardziej od typu pytania niż od jakości samych metod. Czy odpowiedź na pytanie ma mieć charakter opisowy? Czy wymaga rozumowania przyczynowo-skutkowego? W badaniach opisowych nie musimy troszczyć się o metody badań eksperymentalnych, ponieważ nie mają one w tym przypadku zastosowania, w przypadku zaś badań przyczynowo-skutkowych mamy do dyspozycji wiele metod, które można zastosować zależnie od charakteru zagadnienia.

Koncentrując się raczej na pytaniu niż na metodzie, przyjmujemy ten sam kierunek myślenia, jak podpowiadają ułomności poszczególnych metod, to znaczy myślenia w kategoriach *komplementarności* metod i tego, jak się one wzajemnie wspierają. Powierzchowność badań kwestionariuszowych kompensujemy za pomocą studium przypadku; badanie panelowe humanizujemy prowadząc dodatkowo badania kwestionariuszowe lub wywiady socjologiczne; integrujemy proces ewaluacji na etapie analizy jej wyników. W ten sam sposób myślimy o metodach jakościowych i ilościowych, mając raczej na uwadze to, jak wzajemnie się wspomagają, a nie to, która z nich jest lepsza.

Koncentrowanie się raczej na pytaniu niż na metodzie uwalnia nas jakoś od dylematów metodologicznych oraz wymusza pluralizm, który jeszcze niedawno trudno było wypatrzyć w świątyniach ewaluacji. Świątynie te wznoszono na chwałę jakiejś szczególnej, zawsze jednak zwodniczej metody. Każda z nich miała swoich własnych gorących orędowników i wyznawców. Dzisiaj niestety możemy obserwować skumulowany efekt ich działań, rodzaj muru wzniesionego między

ewaluatorami a obszernym zestawem narzędzi metodologicznych, które zawsze były w ich zasięgu.

Trzecia zmiana w sposobie naszego dzisiejszego myślenia o metodach ewaluacji to koncentrowanie się nie tylko na pytaniu warunkującym odpowiedź dotyczącą konkretnej kwestii społecznej, lecz także na perspektywicznych potrzebach informacyjnych osoby korzystającej z ewaluacji. Oznacza to, że ewaluatorzy coraz bardziej angażują się w badania dotyczące przyszłości (ponieważ osoby podejmujące decyzje bardziej interesuje przyszłość niż przeszłość), przestajemy więc traktować naszą pracę jako z natury retrospektywną. Jest to zmiana *znacząca*. Zapożyczyliśmy się bardzo lub nawet jawnie podkradliśmy różne metody używane dotychczas przez badaczy terenowych, matematyków, teoretyków i innych, i stosując te nowe dla nas narzędzia stopniowo powiększamy potencjalny zakres naszej pracy, jej adekwatność społeczną, jak również zmieniamy czasowe koordynanty naszych interwencji. Podsumowując, dzisiaj ewaluujemy zarówno *ex ante*, jak *ex post*, przy czym nawet badanie *ex post także* nastawia się na to, jak minione doświadczenia wykorzystać w przyszłości.

Ewaluacja zorientowana na przyszłość, zwłaszcza jeśli jest podejmowana dla potrzeb polityki społecznej, może w dużym stopniu usprawnić realizację nowego programu. Metoda coraz częściej stosowana przy tego typu ewaluacji nazywana jest *syntezą badawczą* lub metaanalizą. Patrząc z punktu widzenia osób wykorzystujących wyniki ewaluacji, postęp ten jest bez wątpienia jedną z ważniejszych zmian, jakie zaszły w ewaluacji na przestrzeni lat. Umożliwia im w stosunkowo krótkim czasie zrozumienie tego, co zostało ustalone odnośnie do konkretnego problemu lub działania, a czego wcześniej nie wiedzieli; umożliwia określenie tych obszarów, na których występuje prawomocny brak zgody, a także wskazanie potrzeb, które należy poddać odpowiednim badaniom.

Czwarta zmiana jest bezpośrednio związana z trzema nowymi problemami, o których już wspomniałam: pytaniem dotyczącym danej polityki; komplementarnością stosowanych metod; potrzebami informacyjnymi użytkowników ewaluacji. Zmiana ta wiąże się ze znacznym wzrostem zainteresowania *środowiskiem* użytkownika ewaluacji i wynika z troski o wykorzystanie wniosków w praktyce społecznej. Próbuje się określić, co w kulturze organizacyjnej i politycznej użytkownika ewaluacji sprzyja lub przeszkadza takiemu wykorzystaniu wyników badań. Skuteczna ewaluacja wymaga więc od nas analizy środowiska, w jakim działa decydent, po to, aby zrozumieć, gdzie, kiedy, jak i czy w ogóle

ewaluator może interweniować, aby ułatwić korzystanie z wyników swojej pracy.

Przyczyną, dla której zaczęliśmy zwracać tak wiele uwagi na otoczenie adresata wyników ewaluacji, jest to, że staliśmy się bardziej świadomi faktu, iż politycy i ewaluatorzy znacznie różnią się pod względem celów, standardów oczywistości oraz tolerancji na niepewność. Wiedza — podstawowy cel ewaluatora — może się znajdować na odległym miejscu listy priorytetów użytkownika ewaluacji. Podobnie jak zgromadzone przez niego materiały bywa zazwyczaj używana wyłącznie jako instrument w negocjacjach lub przy podejmowaniu decyzji. Z drugiej strony, dla ewaluatora pewność jako taka może być ostatnim z celów jego badań. Nie mówimy o oczywistości, dopóki nasze dane nie *uprawniają* do takich stwierdzeń. Z kolei z punktu widzenia decydenta wartość pozytywną ma „podejmowanie decyzji", bez względu na to, czy istnieją obiektywne dowody potwierdzające ich słuszność.

Oznacza to, że nasza demistyfikująca praca może okazać się niezwykle trudna. Musimy przekonać zagorzałych zwolenników, że to, o czym „wiedzą", iż jest absolutnie prawdziwe, jest w istocie hiperbolą, mitem, czymś, co nie znajduje potwierdzenia lub jest wręcz całkowitym kłamstwem. Spinoza pisał, że wytłumaczenia wymaga popełnienie błędu, a nie odkrycie prawdy. Lecz w przypadku podejmowania decyzji prawdą jest, iż często popełnia się błędy, w które zaangażowani są sami decydenci. Tak więc dla ewaluatora próbującego ustalić fakty jest to, niestety, prawda, która wymaga nie tylko wyjaśnienia, ale bardzo przekonujących argumentów oraz danych wspierających wnioski.

Nauka, jaka płynie z powyższych spostrzeżeń, sprowadza się do tego, że jeśli wnioski z naszych ewaluacji mają być użyte w konkretnych działaniach, to dla ewaluatora prawie równie ważne jak metodologiczna poprawność i niepodważalność samych wyników jest zrozumienie tych, którzy będą je wykorzystywać, a tym samym wyłożenie swoich wniosków w sposób zrozumiały dla ich adresatów.

Na koniec pozwólcie mi wskazać na parę pomniejszych zmian, jakie obecnie obserwujemy w metodach ewaluacji. Po pierwsze, narasta tendencja uciekania się do panelowych dyskusji ekspertów nad kryteriami, gdy nie ma formalnej zgody na to, jakie powinny być te kryteria. Po drugie, często spotykamy ewaluatorów odwołujących się do schematów przy wyjaśnianiu zależności między zestawem skomplikowanych założeń leżących u podłoża jakiegoś programu lub polityki. Po trzecie, od czasu, gdy zaczęliśmy zwracać większą uwagę na potrzeby informacyjne użytkowników ewaluacji, zauważyliśmy, że pytania „co" i „jak"

mogą być tak samo ważne na arenie politycznej, jak pytanie „dlaczego", a odpowiedź na *te* pytania jest w posiadaniu wielu różnych grup będących pierwszorzędnym źródłem informacji opisowych, co w sposób zadziwiający przeoczyliśmy w przeszłości. Tak więc powstaje teraz pytanie: czy jesteśmy *praktykami* (tak jak policjanci i nauczyciele) będąc równocześnie menedżerami tych programów; czy poszukujemy naszych *pacjentów* tak jak lekarze, *farmerów* jak ekonomiści, *ofiar* jak złoczyńcy czy wreszcie *beneficjentów* programów jak ci, którzy te programy planują i realizują.

To, co się zdarzyło, polega na ześrodkowaniu naszych zainteresowań na uzyskiwaniu bardziej złożonego obrazu rzeczywistości, bardziej konkluzywnych odpowiedzi dotyczących spraw publicznych w związku z podstawowym celem, tj. zwiększeniem naszej zdolności dostarczania informacji osobom podejmującym decyzje.

Z drugiej jednak strony musimy przyznać, że o ile lepsze informowanie tych, którzy decydują, jest z pewnością rzeczą godną uwagi, o tyle zmuszenie ich, aby nas słuchali, jest czymś zupełnie odmiennym. W tym miejscu wchodzimy na obszar wiarygodności ewaluacji jako osoby, które mają obiektywnie i kompetentnie przedstawić użytkownikom ewaluacji ich pracę. Tę dziedzinę zainteresowań pozostawiłam sobie na koniec naszego dzisiejszego spotkania.

Znaczenie wiarygodności dla wykorzystania wyników ewaluacji

Moim zdaniem, istnieją dwa typy wiarygodności, które powinny znaleźć odzwierciedlenie w każdym raporcie ewaluacyjnym. Pierwszy z nich to wiarygodność *substancjalna* (*credibility of substance*), odnosi się ona do tego, *czego* dowiedzieliśmy się w toku ewaluacji i *jak* się tego dowiedzieliśmy — na przykład, czy zastosowane metody miały jakiś wpływ zniekształcający [obraz stanu rzeczy] i jak te zniekształcenia trzeba oszacować i uwzględnić. Wiarygodność substancjalna stanowi więc o obiektywności ewaluatora, dowodzi, że ewaluator zachowuje bezstronność.

Drugi typ wiarygodności odnosi się do *prezentacji* wyników; dotyczy nie treści tego, co zostało zrobione, lecz relacji o tym, co zostało zrobione — jej czytelności, rozsądnej objętości, precyzji zarówno logicznej, jak stylistycznej oraz języka, który wszędzie tam, gdzie jest to możliwe, powinien być prosty.

Oczywiście, żadnego z tych aspektów wiarygodności nie można oddzielić od drugiego. Przeciwnie — są one ze sobą ściśle związane w tym sensie, że to, *co* jest opisywane, oraz, *jak* to jest opisywane, oddziałuje na recepcję obiektywności i kompetencji ewaluatora. Innymi słowy, obiektywność i kompetencja ewaluatora nie będą miały znaczenia, jeśli jego praca zostanie przedstawiona tak niejasno, że może spowodować całkowite niezrozumienie prezentowanego problemu.

Mimo że nie można oddzielić tych dwóch typów wiarygodności, pozwólcie, że omówię je kolejno dla zachowania jasności mojego wywodu.

Wiarygodność substancjalna

Jeśli celem przeprowadzenia ewaluacji jest dostarczenie nowej wiedzy pracownikom zajmującym się polityką społeczną, to wiarygodność substancjalna sprowadza się do pytania, jaka jest ta wiedza i jak została uzyskana. Innymi słowy, wiarygodność substancjalna dotyczy tego, jak w procesie ewaluacji ustala się to, co jest *wiadome*, to, co jest *niepewne*, i to co *nie jest* wiadome, jak również, jaki jest stopień pewności w odniesieniu do wszystkich trzech typów ustaleń.

Obecnie używamy wielu metod wzmacniających wiarygodność, trzy z nich jednak mają — jak sądzę — znaczenie kluczowe, albowiem pokazują wysiłek włożony w to, aby uzyskać wiedzę należytej jakości. Pierwsza z tych metod to przegląd literatury przedmiotu. Uwiarygodnia on ewaluację na cztery sposoby: pomaga ewaluatorom ustalić konkretną ekspertyzę; sytuuje daną ewaluację w stosunku do poprzednich prac w tej samej dziedzinie; stwarza układ odniesienia dla dokładnego wykazania, jakie *nowe* informacje zostały zgromadzone w wyniku przeprowadzonej ewaluacji; pomaga zaprezentować jej obiektywizm dzięki temu, że ewaluatorzy uwzględniają prace charakteryzujące się odmiennymi punktami widzenia i odmienną metodologią niż ta, którą sami stosowali w swojej pracy. Mówiąc inaczej, jednostronna prezentacja wykonanej pracy u tych, którzy znają literaturę przedmiotu, zawsze wzbudza podejrzenie, że kierowano się jakimiś ukrytymi motywami. Pierwszym sprawdzianem jakości wykonanej pracy ewaluacyjnej jest więc jakość przeglądu literatury zawartej w raporcie.

Drugim sposobem uwiarygodnienia substancjalnego jest *stosowny* samokrytycyzm w dyskusji metodologicznej. Odwołajmy się, na przykład, do metod eksperymentu badawczego. Wielką zaletą tych metod jest jednoznaczność uzyskiwanych wyników, ale mają one pewne słabości, obecnie już dobrze rozpoznane. Należy do nich zwłaszcza brak ela-

styczności i powiązania między naszym rozumieniem programu i jego efektywnością. Tym niemniej, jak wspomniałam wcześniej, nadal stosujemy te projekty badawcze, jak też inne metody *quasi*-eksperymentalne, neutralizując ich ograniczenia zaletami innych metod. Budujemy wiarygodność tych projektów badawczych otwarcie mówiąc zarówno o ich niedoskonałościach, jak i o zaletach.

Na przykład, *mówimy* czytelnikowi raportu, jak uporaliśmy się z zagrożeniem dla trafności, mówimy także, czego nie można było zrobić, choć było to potrzebne lub pożądane. Na przykład, jeśli termin wyznaczony przez osobę zamawiającą ewaluację lub niedostatek przeznaczonych na ten cel środków uniemożliwił nam, powiedzmy, staranną selekcję wyjaśnień potrzebnych do sformułowania wniosków, tak że raport końcowy zawiera pewne prawdopodobne warianty alternatywne, to *mówimy* o tym otwarcie, zaznaczając również, jakie mogą być praktyczne konsekwencje wykorzystania naszych wniosków. Nauczyliśmy się, że niewielki odsetek odpowiedzi w badaniu sondażowym lub niemożność sformułowania uogólnień w nietypowych badaniach jakościowych w mniejszym stopniu ogranicza wiarygodność badawczą niż brak jasnej charakterystyki problemu i jego implikacji w analizie potrzeb polityki społecznej.

Trzeci sposób uwiarygodnienia wiąże się dzisiaj z *przewidywaniem* możliwych ataków metodologicznych i politycznych wymierzonych w wyniki ewaluacji. Pozwala nam to ubezpieczać się już na etapie *projektu badawczego*, aby móc odpowiedzieć na te ataki, wykazując zasadność naszych wniosków.

Krótko mówiąc, wiarygodność substancjalna jest pochodną odbioru, *czytelnik* raportu ma przekonać się, że jesteśmy ekspertami w swojej dziedzinie i uczciwie mówimy o naszych wnioskach i naszej metodologii. Jednakże kompetencja i szczerość są tylko połową naszego zadania: musimy, oczywiście, uzewnętrznić je w naszych wnioskach, ale równocześnie musimy uniknąć tego, co Cronbach nazywa „samoudaremniającą się dokładnością" (*self-defeating thoroughness*). Narażeni bowiem jesteśmy na istotny dylemat: dla ewaluatora wszystkie dane są ważne, a wszystkie techniczne detale interesujące, lecz dla polityka niekoniecznie. Przestaje czytać raport, wierząc jak Voltaire, że „sekret nudziarza polega na tym, iż mówi o wszystkim". Jak więc rozwiązać dylemat i mówić o wszystkim tak, aby być czytanym? Pytanie to zbliża nas do wiarygodności prezentowania wyników, łatwo się o tym mówi, trudniej stosować to w praktyce.

Wiarygodność prezentowania wyników

Machiavelli już około 450 lat temu powiedział, co należy przedstawić w raporcie z ewaluacji naszym użytkownikom i jak to ująć, aby nasze wnioski były dla nich zrozumiałe. Raport powinien być zwięzły, bez rozwlekłych zdań i górnolotnych słów, choć nauczenie się tego może nas kosztować lata ciężkiej pracy. Musimy przekazać w nim naszym czytelnikom przynajmniej cztery sprawy:

- dokładną treść pytań, na które poszukiwaliśmy odpowiedzi;
- metody (czasem bardzo złożone), którymi posługiwaliśmy się szukając odpowiedzi;
- główne zalety i słabości projektu badań i sposobu jego realizacji oraz starania czynione dla skompensowania tych słabości;
- i najważniejsze — wnioski z badań: do jakiego stopnia są wiążące, czy i dlaczego są ważne, jak się mają do spostrzeżeń poprzedzających daną ewaluację oraz gdzie i jak powinny być zastosowane.

Ideałem jest, oczywiście, połączenie wszystkich tych rzeczy ze sobą, a przy tym respektowanie przepisu Pascala na elokwencję, to jest mówienie o tych sprawach tak, aby odbiorcy nie tylko mogli bez trudu (ba, wręcz z satysfakcją) je zrozumieć, ale także, by to, co napisaliśmy, okazało się rzeczywiście zajmujące i w ich własnym interesie nakłaniające do dalszych refleksji.

Pewne napięcie jest tu nieuniknione. Nikt nie zakwestionuje zależności między wiarygodnością a *precyzją* wypowiedzi, choć chcemy również, z uwagi na wiarygodność, do minimum ograniczyć język specjalistyczny. Jak się okazało, dobrym rozwiązaniem technicznym są suplementy do raportu. Pozwalają obszernie omówić na przykład to, jak ewaluatorowi udało się uniknąć różnego typu tendencyjnych danych, a zarazem nie ingerują w tok wywodu w jego warstwie merytorycznej, traktującej o przedmiocie ewaluacji.

Innym sposobem rozwiązania tego problemu jest poprzedzenie raportu krótką prezentacją ustną. Działacze polityczni korzystają z niej bardzo często; jest to skuteczny, krótki i klarowny sposób wyjaśnienia tego, czego dowiedzieliśmy się w trakcie badań; bez wdawania się w techniczne zawiłości jesteśmy w stanie przekazać komunikat polityczny — jeśli nie w pełni satysfakcjonujący, to przecież wychodzący naprzeciw potrzebom informacyjnym użytkownika ewaluacji.

Mówiąc w skrócie, wiarygodność jest ważna i choć nie przychodzi łatwo zasługuje na to, aby się o nią starać, ponieważ zapewnia: rzetelne wysłuchanie naszych racji; zainteresowanie odbiorców; przetrwanie do następnego przemówienia; znacznie większe prawdopodobieństwo, że

nasza praca będzie wykorzystana i będzie mieć *znaczenie* w kształtowaniu praktyki różnych dziedzin życia.

Wnioski

Podsumowując, chciałabym powiedzieć, że mimo pewnych trudności, jakie mają ewaluatorzy zwłaszcza wtedy, gdy ewaluowana polityka jest bardzo kontrowersyjna, *generalnie* rzecz biorąc ewaluacja przezwycięża obecnie dawne problemy, doskonali swoją praktykę i uzyskuje większe wsparcie ze strony decydentów. Mamy teraz większe możliwości analizy problemów społecznych dotyczących wielu kluczowych obszarów w sposób produktywny i użyteczny dla naszych sponsorów. Co istotne, możemy również zmagać się z bardziej złożonymi i ambitnymi problemami niż w przeszłości. Nie pytamy już, czy *możemy robić* ewaluacje, pytamy raczej, dlaczego nie możemy ich robić więcej.

Wydaje się, że tak jest przynajmniej w Stanach Zjednoczonych, gdzie znacznie więcej ewaluacji *będzie* wykonywanych dzięki wysiłkom National Performance Review oraz dzięki ustawodawstwu, które wspiera tą inicjatywę. Tak jak powiedział Nils Bohr: „Przewidywanie jest rzeczą trudną, zwłaszcza przewidywanie przyszłości", a tym bardziej w obliczu tej politycznej zawieruchy, jakiej jesteśmy świadkami dziś, na przełomie wieków. Ale faktem jest, że rozwijamy środki służące określeniu stopnia realizacji zadań — na przykład w około dziewięćdziesięciu badaniach pilotażowych, co pozwala oszacować indywidualne osiągnięcia poszczególnych programów i agend rządowych. Poza tym każda agenda rządowa jest zobligowana do określenia swoich zadań ewaluacyjnych w 1997 r.

Z opisanych wyżej względów, jak również z powodu ogólnoświatowych procesów rozwoju, moja wizja ewaluacji zmienia się i pozwala mi optymistycznie widzieć jej przyszłe perspektywy. Nie są one łatwe do realizacji i nakazują nam mówić o potrzebie wytrwałości. Wielkie zmiany w instytucjach zwykle przychodzą powoli, wymagają wielu przygotowań, ustaleń i niewdzięcznej ciężkiej pracy. Ewaluacja *jest* wielką zmianą. Sukces ewaluacji wymaga: zgromadzenia utalentowanych ewaluatorów i przyjścia im z pomocą we wzajemnym zrozumieniu; wybierania takich tematów do ewaluacji, które są ważne z punktu widzenia realizacji konkretnej polityki i jednocześnie możliwe do oceny; ustalenia czasu, zapewnienia środków na finansowanie ewaluacji i zobowiązania użytkowników ewaluacji do pomocy w jej realizacji, to znaczy zapewnienia niezależności jej wykonawcom; wsparcia ewaluatorów w ich

walce z wypaczaniem ich wniosków; zapewnienia że ewaluatorzy rozumieją kulturę i potrzeby informacyjne użytkowników, którym niekiedy ewaluowana problematyka jest dość obca, ale od których zależeć mogą losy jej wyników.

Nie jest łatwo ewaluatorom i działaczom politycznym nauczyć się razem pracować. Chociaż ostatnio dostrzegamy wiele przykładów ich owocnej współpracy i choć wygląda na to, że przyszłość rysuje się bardzo obiecująco, to jednak wciąż muszę radzić każdemu ewaluatorowi, aby pamiętał, że cielę nigdy nie może spać spokojnie, jeśli mieszka z lwem pospołu.

Prawdopodobnie największym dokonaniem współczesnej ewaluacji jest to, że nie musimy wciąż powtarzać — tak jak to było w latach osiemdziesiątych — że nasza praca nie jest wykorzystywana w kształtowaniu polityki społecznej. Przeciwnie, chlubą dzisiejszej ewaluacji jest to, że ma trwałą podstawę, że wnioski z niej płynące są wykorzystywane w konkretnych działaniach oraz że ewaluacje stały się składową rutynowej praktyki organizacyjnej. Tak więc ewaluacja upowszechnia się w różnych krajach jako strategiczne wsparcie rozwoju i może tak się zdarzyć, że będziemy mieli szansę zobaczyć ewaluację, zgodnie z naszymi pierwotnymi oczekiwaniami, jako rodzaj pomocy dla instytucji, aby stały się bardziej efektywne, bardziej odpowiadające istniejącym potrzebom, bardziej wiarygodne, bardziej odpowiedzialne i nawet — kto wie? — lepiej zarządzane.

Tłumaczył *Michał Kowalski*

Bibliografia

Adelman C. (red.) (1984) *The Politics and Ethics of Evaluation*, Croom Helm, London.

Adelman C., Alexander R. J. (1982) *The Self-Evaluating Institution: Practice and Principles in the Management of Educational Change*, Methuen, London.

Adelman C., Kemmis S., Jenkins D. (1976) *Rethinking Case Study: Notes from the Second Cambridge Conference*, „Cambridge Journal of Education" nr 3.

Alkin M. C. (1969) *Evaluation Theory Development*, „Evaluation Comment" t. 2, s. 2–7.

Alkin M. C., Daillak R., White P. (1979) *Using Evaluations: Does Evaluation Make a Difference*, Sage, Beverly Hills.

Alkin M. C., Ellett F. S. (1979) *The Importance of Category Systems in Evaluation Theory: A Personal Viewpoint*, „CEDR Quarterly", Fall.

Anderson R. C. (1977) *The Notion of Schema and the Educational Enterprise*, w: R. C. Anderson, R. J. Spiro, W. E. Montague (red.), *Schooling and the Acquisition of Knowledge*, Erlbaum, Hillsdale N. J.

Apple M. W. (1974) *The Process and Ideology of Valuing in Educational Settings*, w: M. W. Apple, M. G. Subkoviak, H. S. Lufler (red.), *Educational Evaluation: Analysis and Responsibility*, McCutchan, Berkeley.

Apple M. W. (1979) *Ideology and Cerriculum*, Routledge and Kegan, London.

Atkin J. M. (1963) *Some Evaluation Problems in a Course Content Improvement Project*, „Journal of Research in Science Teaching" t. 1, s. 129–132.

Atkin J. M. (1967–1968) *Research Styles in Science Education*, „Journal of Research in Science Teaching" t. 5, s. 338–345.

Atkin J. M., House E. R. (1981) *The Federal Role in Curriculum Development, 1950–1980*, „Educational Evaluation and Policy Analysis" nr 5.

Ausubel D. P., Fitzgerald D. (1961) *Meaningful Learning and Retention: Interpersonal Cognitive Variables*, „Review of Educational Research" t. 31, s. 500–510.

Barnes R. E., Ginsburg A. L. (1979) *Relevance of the RMC Models for Title I Policy Concerns*, „Educational Evaluation and Policy Analysis", March–April.

Barry B. (1965) *Political Argument*, Routledge and Kegan, London.

Barry B. (1973) *The Liberal Theory of Justice*, Oxford University Press, London.

Becker H. S. (1958) *Problems of Interference and Proof in Participant Observation*, „American Sociological Review" t. 23, s. 652–660.

Becker H. S. (1967) *Whose Side Are We on?*, „Social Problems" nr 3.

Becker H. S., Geer B., Hughes E. C., Strauss A. L. (1961) *Boys in White: Student Culture in Medical School*, University of Chicago Press, Chicago.

Bellah R. N., Madsen R., Sullivan W. M., Swidler A., Tipton S. M. (1985) *Habits of the Heart: Individualism and Commitment in American Life*, University of California Press, Berkeley.

Berger P. L. (1974) „Policy and the Calculus of Meaning", w: *Pyramids of Sacrifice*, Basic Books, New York.

Berk R. A., Rossi P. H. (1990) *Thinking about Program Evaluation*, Sage, Newbury Park – London.

Birney B. A. (1988) *Brookfield Zoo's „Flying Walk" Exhibit: Formative Evaluation Aids in the Development of an Interactive Exhibit in an Informal Learning Setting*, „Environment and Behavior" t. 20, s. 416–434.

Bloom B. S. (1956) *Taxonomy of Educational Objectives*, McKay, New York.

Bloom B. S., Hastings J. T., Madaus G. (1971) *Handbook on Formative and Summative Evaluation of Student Learning*, McGraw-Hill, New York.

Blythe R. (1969) *Akenfield*, Penguin Books, London.

Bogdan R. C., Biklen S. K. (1982) *Qualitative Research for Education: An Introduction to Theory and Methods*, Allyn and Bacon, Boston.

Bohm D. (1985) *Unfolding Meaning: A Weekend Dialogue with David Bohm*, Routledge and Kegan, New York.

Boruch F. R., Cordray D. S. (1980) *An Appraisal of Educational Program Evaluations: Federal, State, and Local Agencies*, Northwestern University, Evanston Ill.

Brauner C. J. (1974) *The First Probe*, w pracy zbiorowej *Four Evaluation Examples: Anthropological, Economic, Narrative, and Portrayal*, „AERA Monograph Series on Curriculum Evaluation" nr 7, Rand McNally, Chicago.

Broadhead R. S., Fox K. J. (1990) *Takin' It to the Streets: AIDs Outreach as Ethnography*, „Journal of Contemporary Ethnography" nr 3.

Brown J. D. (1991) *Preprofessional Socialization and Identity Transformation: The Case of the Professional ex-*, „Journal of Contemporary Ethnography" nr 2.

Campbell D. T. (1969) *Reforms as Experiments*, „American Psychologist" t. 24, s. 409–429.

Campbell D. T. (1975) *Degrees of Freedom and Case Study*, „Comparative Political Studies" t. 8, s. 178–193.

Campbell D. T. (1977) *Keeping the Data Honest in the Experimenting Society*, w: H. W. Melton, D. H. G. Watson (red.), *Interdisciplinary Dimensions of*

Accounting for Social Goals and Social Organizations, University of Illinois, Urbana-Champaign, Grid Inc., Columbus Ohio.

Care N. S. (1978) *Participation and Policy*, „Ethics" t. 88, July.

Carr W. L., Kemmis S. (1986) *Becoming Critical: Education, Knowledge and Action Research*, Falmer, London.

Carter K. (1993) *The Place of Story in the Study of Teaching and Teacher Education*, „Educational Researcher" t. 22, s. 5-12.

Cazden C. B. (1988) *Classroom Discourse: The Language of Teaching and Learning*, Heinemann Educational Books, Portsmouth N. H.

Coles R. (1967) *Children of Crisis*, Little, Brown, Boston.

Coles R. (1989) *The Call of Stories: Teaching and the Moral Imagination*, Houghton Mifflin, Boston.

Cook T. D., Campbell D. T. (1976) *The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings*, w: M. D. Dunnette (red.), *Handbook of Industrial and Organizational Psychology*, Rand McNally, Chicago.

Cook T. D., Campbell D. T. (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Rand McNally, Chicago.

Cooley W. W., Lohnes P. R. (1976) *Evaluation Research in Education*, Irvington, New York.

Cronbach L. J. (1963) *Course Improvement Through Evaluation*, „Teacher College Record" t. 64, s. 672-683.

Cronbach L. J. (1975) *Beyond the Two Disciplines of Scientific Psychology*, „American Psychologist" nr 2.

Cronbach L. J. (1979) *Design of Evaluations*, Stanford Evaluation Consortium, Stanford.

Cronbach L. J. (1982) *Designing Evaluations of Educational and Social Programs*, Jossey-Bass, San Francisco.

Cronbach L. J., Ambron S. R., Dornbusch S. M., Hess R. D., Hornik R. C., Phillips D. C., Walker D. E., Weiner S. S. (1980) *Toward Reform of Program Evaluation: Aims, Methods and Institutional Arrangements*, Jossey-Bass, San Francisco.

Dawson J. (1977) „Validity in Qualitative Enquiry", tekst powielony, University of Illinois, Urbana-Champaign.

Delamont S. (1992) *Fieldwork in Educational Settings: Methods, Pitfalls and Perspectives*, Falmer, London.

Denzin N. K. (1971) *The Logic of Naturalistic Inquiry*, „Social Forces" nr 2.

Denzin N. K. (1978) *The Research Act*, McGraw-Hill, New York.

Dornbusch S. M., Scott W. R. (1975) *Evaluation and the Exercise of Authority*, Jossey-Bass, San Francisco.

Douglas J. D. (1976) *Investigative Social Research: Individual and Team Field Research*, Sage, Newbury Park Cal.

Eisner E. W. (1975) „Applying Educational Connoisseurship and Criticism to Education Settings", praca nie publikowana, fragmenty w: D. Hamilton i in. (red.), *Beyond the Numbers Game: A Reader in Educational Evaluation*, Macmillan, London 1977.

Eisner E. W. (1977) *On the Uses of Educational Connoisseurship and Educational Criticism for Evaluating Classroom Life*, „Teachers College Record" t. 78, s. 345–358.

Eisner E. W. (1979) *The Educational Imagination*, Macmillan, New York.

Eisner E. W. (1984) *The Art of Educational Evaluation: A Personal View*, Falmer, Lewes.

Eisner E. W. (red.) (1985) *Learning and Teaching the Ways of Knowing*, National Society for the Study of Education, University of Chicago Press, Chicago.

Elliott J. (1976) *Developing Hypotheses about Classrooms from Teachers' Practical Constructs*, North Dakota Study Group on Evaluation, University of North Dakota Press, Grand Forks.

Elliott J. (1977) *Democratic Evaluation as Social Criticism: Or Putting the Judgement Back into Evaluation*, w: N. Norris (red.), *Theory in Practice*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Elliott J. (1979) „Curriculum Evaluation and the Classroom", tekst powielony, Cambridge Institute of Evaluation.

Elliott J. (1980) *Implications of Classroom Research for Professional Development*, w: E. W. Hoyle, J. Megarry (red.), *World Yearbook of Education 1980: Professional Development of Teachers*, Kogan Page, London.

Elliott J. (1981a) *Educational Accountability and Evaluation of Teaching*, w: A. Lewy, D. Nevo (red.), *Evaluation Roles in Education*, Gordon and Breach, London.

Elliott J. (1981b) *The Cambridge Accountability Project*, „Cambridge Journal of Education" nr 2.

Elliott J. (1982) „Facilitating Action Research in Schools: Some Dilemmas", opracowanie przygotowane na konferencję zorganizowaną w lipcu przez Social Science Research Council na temat metod jakościowych i badań w dziedzinie edukacji, Whitelands College, London.

Elliott J. (1984) *Methodology and Ethics*, w: C. Adelman (red.), *The Politics and Ethics of Evaluation*, Croom Helm, London.

Elliott J., Bridges D., Ebbutt D., Gibson R., Nias J. (1981) *School Accountability*, Grant McIntyre, London.

Eraut M. (1976) *School-based Evaluation*, w: M. Raggett, M. Clarkson (red.), *Teaching the Eight to Thirteen*, vol. 2, Ward Lock Education, London.

Eraut M. (1978) *Accountability at School Level: Some Options and Their Implications*, w: T. Becher, S. Maclure (red.), *Accountability in Education*, National Foundation for Educational Research Publishing Company, Windsor.

Eraut M. (1984a) *Handling Value Issues*, w: C. Adelman (red.), *The Politics and Ethics of Evaluation*, Croom Helm, London.

Eraut M. (1984b) *Institution-based Curriculum Evaluation*, w: M. Skilbeck (red.), *Evaluating the Curriculum in the Eighties*, Hodder and Stoughton, London.

Erickson F. (1986) *Qualitative Methods in Research on Teaching*, w: M. C. Wittrock (red.), *Handbook of Research on Teaching*, Macmillan, New York.

Evaluation Research Society (1980) *Standards for Program Evaluation*.

Feagin J. R., Orum A. M., Sjoberg G. (1991) *A Case for the Case Study*, University of North Carolina Press, Chapel Hill.

Firestone W. A. (w druku) „Alternative Arguments for Generalizing from Data es Applied to Qualitative Research", „Educational Researcher".

Franke-Wikberg S. (1979) „The LONG Projekt: An Example of Educational Evaluation Given from Sweden", tekst powielony, Umea University.

Flick U. (1992) *Triangulation Revisited: Strategy of Validation or Alternative?*, „Journal for the Theory of Social Behaviour" t. 22, s. 175–198.

Freudenberg N. (1990) *Developing a New Agenda for the Evaluation of AIDS Education*, „Health Education Research" t. 5, s. 295–298.

Garfinkel H. (1967) *Studies in Ethnomethodology*, Prentice-Hall, New York.

Geertz C. (1973) „Thick Description: Toward an Interpretive Theory of Culture", w: *The Interpretation of Cultures*, Basic Books, New York.

Geertz C. (1983) *Local Knowledge: Further Essays Interpretive Anthropology*, Basic Books, New York.

Glaser B. G., Strauss A. L. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine, Chicago.

Glass G. V. (1969) „The Growth of Evaluation Methodology", tekst powielony, Laboratory of Educational Research, University of Colorado.

Glass G. V. (1972) *The Wisdom of Scientific Inquiry on Education*, „Journal of Research in Science Teaching" nr 1.

Goetz J. P., LeCompte M. D. (1984) *Ethnography and Qualitative Design in Educational Research*, Academic Press, New York.

Goode W. J., Hatt P. K. (1952) „The Case Study", w: *Methods of Social Research*, McGraw-Hill, New York.

Gouldner A. V. (1973) *For Sociology: Renewal and Critique in Sociology Today*, Allen Lane, London.

Guba E. G. (1969) *The Failure of Educational Evaluation*, „Educational Technology" t. 9, s. 29–38.

Guba E. G. (1978a) *Metaphor Adaptation Report: Investigative Journalism*, Northwest Regional Educational Laboratory, Portland Or.

Guba E. G. (1978b) *Toward a Methodology of Naturalistic Inquiry in Educational Evaluation*, „CSE Monograph Series in Evaluation" nr 8, Center for the Study of Evaluation, Los Angeles.

Guba E. G., Lincoln Y. S. (1981) *Effective Evaluation*, Jossey-Bass, San Francisco.

Guba E. G., Lincoln Y. S. (1985) *Naturalistic Inquiry*, Sage, Beverly Hills – London.

Guttentag M. (1973) *Subjectivity and Its Use in Evaluation Research*, „Evaluation" nr 2.

Guttentag M., Snapper K. (1974) *Plans, Evaluations, and Decisions*, „Evaluation" nr 2.

Hamilton D. (1976a) *Classroom Research: A Cautionary Tale*, w: J. Wolfson (red.), *Personality and Learning*, Open University Press, Hodder and Stoughton, London.

Hamilton D. (1976b) *Classroom Research: A Critique and a New Approach*, w: M. Strubbs, S. Delamont (red.), *Explorations in Classroom Observation*, Wiley, London.

Hamilton D. (1976c) *Curriculum Evaluation*, Open Books, London.

Hamilton D. (1978) *Making Sense of Curriculum Evaluation: Continuities and Discontinuities in an Educational Idea*, w: L. Shulman (red.), *Review of Research in Education*, vol. 5, Peacoc, Itasca Ill.

Hamilton D. (1980a) *Educational Research and the Shadows of Francis Galton and Ronald Fisher*, w: W. B. Dockrell, D. Hamilton (red.), *Rethinking Educational Research*, Hodder and Stoughton, London.

Hamilton D. (1980b) *Educational Research and the Shadow of John Stuart Mill*, w: D. Hamilton, J. V. Smith (red.), *The Meritocratic Intellect: Studies in the History of Educational Research*, Aberdeen University Press, Aberdeen.

Hamilton D. (1980c) *Some Contrasting Assumption about Case Study Research and Survey Analysis*, w: H. Simons (red.), *Towards a Science of the Singular: Essays about Case Study in Educational Research and Evaluation*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Hamilton D., Jenkins D., King C., MacDonald B., Parlett M. (red.) (1977) *Beyond the Numbers Game: A Reader in Educational Evaluation*, Macmillan, London.

Harlen W. (1973) *Science 5/13 Project*, w: D. Tawney (red.), *Evaluation in Curriculum Development: Twelve Case Studies*, Macmillan, London.

Harlen W. (1975) *Science 5/13: A Formative Evaluation*, Schools Council Research Series, Macmillan, London.

Harlen W. (1976) *Evaluation in the Context of Helping Children Learn Science*, „Education" nr 2.

Harlen W. (1977a) *Match and Mismatch: Raising Questions*, Oliver and Boyd, Edinburgh.

Harlen W. (1977b) *Match and Mismatch: Finding Answers*, Oliver and Boyd, Edinburgh.

Harlen W. (1979a) „School-based Evaluation. Pupil Assesment", w: *School-based Evaluation*, „Classroom Action Research Network Bulletin" nr 3, Cambridge Institute of Education, Cambridge.

Harlen W. (1979b) *Accountability That Is of Benefit to Schools*, „Journal of Curriculum Studies" nr 4.

Harlen W., Elliott J. (1982) *A Checklist for Planning or Reviewing an Evaluation*, w: R. McCormick (red.) *Calling Education to Account*, Heinemann, London.

Harrington P., Sanders J. R. (1979) „Guidelines for Goal-free Evaluation", tekst powielony, Evaluation Center, Western Michigan University.

Herriott R. E., Firestone W. A. (1983) *Multisite Qualitative Policy Research: Optimizing Description and Generalizability*, „Educational Researcher" nr 2.

House E. R. (1973a) *School Evaluation: The Politics and Process*, McCutchan, Berkeley.

House E. R. (1973b) *The Conscience of Educational Evaluation*, w: E. R. House (red.), *School Evaluation: The Politics and Process*, McCutchan, Berkeley.

House E. R. (1973c) „The Price of Productivity: Who Pays?", tekst powielony, University ef Illinois, Urbana-Champaign.

House E. R. (1974) *The Politics of Educational Innovation*, McCutchan, Berkeley.

House E. R. (1976) *Justice in Evaluation*, w: G. V. Glass (red.), *Evaluation Studies Annual Review*, vol. 1, Sage, Beverly Hills.

House E. R. (1977) *The Logic of Evaluative Argument*, „CSE Monograph Series in Evaluation" nr 7, Center for the Study of Evaluation, Los Angeles.

House E. R. (1978a) *Assumption Underlying Evaluation Models*, „Educational Researcher" t. 7, s. 4–12.

House E. R. (1978b) *Evaluation as Scientific Management in United States School Reform*, „Comparative Education Review" nr 3.

House E. R. (1979a) *Coherence and Credibility: The Aesthetics of Evaluation*, „Educational Evaluation and Policy Analysis" nr 5.

House E. R. (1979b) *The Objectivity, Fairness and Justice of Federal Evaluation Policy as Reflected in the Follow Through Evaluation*, „Educational Evaluation and Policy Analysis" nr 1.

House E. R. (1980) *Evaluating with Validity*, Sage, Beverly Hills.

House E. R. (1986) *Participatory Evaluation: The Stakeholder Approach*, w: E. R. House (red.), *New Directions in Educational Evaluation*, Falmer, Lewes.

House E. R., Gjerde C. L. (1973) *PLATO Comes to the Community College*, Center for Instructional Research and Curriculum Evaluation, University of Illinois, Urbana-Champaign.

House E. R., Glass G. V., McLean L. D., Walker D. (1978) *No Simple Answer: Critique of the Follow Through Evaluation*, „Harvard Educational Review", May.

House E. R., Hogben D. (1974) *A Goal-free Evaluation for „Me and My Environment"*, „Formative Evaluation Report" nr 3, Biological Sciences Curriculum Study.

House E. R., Rivers C. W., Stufflebeam D. L. (1974) *An Assessment of the Michigan Accountability System*, „Phi Delta Kappa", June.

House E. R., Steele J. M., Kerins T. (1971) *The Gifted Classroom*, Center for Instructional Research and Curriculum Evaluation, University of Illinois, Urbana-Champaign.

Huberman A. M., Miles M. B. (1984) *Innovation up Close: How School Improvement Works*, Plenum, New York.

Jenkins D. (1980) *An Adversary's Account of SAFARI's Ethics of Case Study*, w: H. Simons (red.), *Towards a Science of the Singular: Essays about Case Study in Educational Research and Evaluation*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Jenkins D., Kemmis S., MacDonald B., Verma G. (1977) „Racism and Educational Evaluation", tekst powielony, Centre for Applied Research in Education, University of East Anglia, Norwich.

Jenkins D., O'Toole B. (1978) *Curriculum Evaluation, Literary Criticism, and the Paracurriculum*, w: G. Willis (red.), *Qualitative Evaluation: Concepts and Cases in Curriculum Criticism*, McCutchan, Berkeley.

Jenkins D., Simons H., Walker R. (1981) *Thou Nature Art My Goddess: Naturalistic Inquiry in Educational Evaluation*, „Cambridge Journal of Education" nr 3.

Joint Committee on Standards for Educational Evaluation (1981) *Standards for Evaluations of Educational Programs, Projects, and Materials*, McGraw-Hill, New York.

Jones R. A. (1985) *Research Methods in the Social and Behavioral Sciences*, Sinauer, Sunderland Mass.

Judd C. M., Smith E. R., Kidder L. H. (1991) *Research Methods in Social Relations*, Holt, Rinehart and Winston, New York.

Kallos D. (1978) *Notes on Schooling, Curriculum, and Teaching*, w: G. Willis (red.), *Qualitative Evaluation: Concepts and Cases in Curriculum Criticism*, McCutchan, Berkeley.

Karier C., Violas P., Spring J. (1973) *Roots of Crisis: American Education in the Twentieth Century*, Rand McNally, Chicago.

Katzer J., Cook K. H., Crouch W. W. (1978) *Evaluating Information: A Guide for Users of Social Science Research*, Addison-Wesley, Reading Mass.

Kelly E. F. (1975) *Curriculum Evaluation and Literary Criticism: Comments on the Analogy*, „Curriculum Theory Network" t. 5, s. 98–106.

Kelly E. F. (1978) *Curriculum Criticism and Literary Criticism: Comments on the Anthology*, w: G. Willis (red.), *Qualitative Evaluation: Concepts and Cases in Curriculum Criticism*, McCutchan, Berkeley.

Kemmis S. (1980) *The Imagination of the Case and the Invention of the Study*, w: H. Simons (red.), *Towards a Science of the Singular: Essays about Case Study in Educational Research and Evaluation*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Kennedy M. M. (1979) *Generalizing from Single Case Studies*, „Evaluation Quarterly" t. 3, s. 661–678.

Kogan M., Packwood T. (1974) *Advisory Councils and Committees in Education*, Routledge and Kegan, London.

Koscoff J., Fink A. (1982) *Evaluation Basics: A Practitioner's Manual*, Sage, Newbury Park – London.

Kozol J. (1991) *Savage Inequalities*, Harper, New York.

Kushler M. B. (1989) *Use of Evaluation to Improve Energy Conservation Programmes: A Review and Case Study*, „Journal of Social Issues" t. 45, s. 153–168.

Lang K., Lang G. E. (1960) *Decisions for Christ: Billy Graham in New York*, w: M. Stein, A. J. Vidich, D. M. White (red.), *Identity and Anxiety*, Free Press, New York.

Lewis O. (1966) *La Vida*, Random House, New York.

Lipsitz G. (1988) *Life in the Struggle: Ivory Perry and the Culture of Opposition*, Temple University Press, Philadelphia.

Locke L. F., Spirduso W. W., Silverman S. J. (1987) *Proposals that Work: A Guide for Planing Dissertations and Grant Proposals*, Sage, Newbury Park Cal.

Lofland J. (1971) *Analyzing Social Settings*, Wadsworth, Belmont Cal.

Lofland J., Lofland L. H. (1984) *Analyzing Social Settings: A Guide to Qualitative Observational Research*, Wadsworth, Belmont Cal.

Lord G. F., Price A. C. (1992) *Growth Ideology in a Period of Decline: Deindustrialization and Restructuring Flint Style*, „Social Problems" nr 2.

Lundgren U. P. (1977) *Model Analysis of Pedagogical Processes*, Stockholm Institute of Education, Stockholm.

Lynd R. S., Lynd H. M. (1929) *Middletown: A Study in American Culture*, Harcourt, Brace, New York.

MacDonald B. (1970) „The Evaluation of the Humanities Curriculum Project: A Holistic Approach", referat przedstawiony na dorocznej sesji American Educational Research Association w Nowym Jorku.

MacDonald B. (1973) *Briefing Decision-Makers*, w: E. R. House (red.), *School Evaluation: The Politics and Process*, McCutchan, Berkeley.

MacDonald B. (1974) *Evaluation and the Control of Education*, w: B. MacDonald, R. Walker (red.), *SAFARI I: Innovation, Evaluation, Research and the Problem of Control*, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B. (1977) *The Portrayal of Persons as Evaluation Data*, w: N. Norris (red.), *SAFARI 2: Theory in Practice*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Mac Donald B. (1978a) *Accountability, Standards and the Process of Schooling*, w: T. Becher, S. Maclure (red.), *Accountability in Education*, National Foundation for Educational Research Publishing Company, Windsor.

MacDonald B. (1978b) *The Experience of Innovation*, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B. (1979) *Hard Times — Accountability in England*, „Educational Analysis" t. 1, Summer.

MacDonald B. (1980a) „Barriers to Democratic Evaluation", tekst powielony, przygotowany na Uniwersytecie Wschodniej Anglii dla Uniwersytetu Stanu Illinois.

MacDonald B. (1980b) *Letters from a Headmaster*, w: H. Simons (red.), *Towards a Science of the Singular: Essays about Case Study in Educational Research and Evaluation*, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B. (1981) *Interviewing in Case Study Evaluation*, „Phi Delta Kappa CEDR Quarterly" nr 4.

MacDonald B. (1982a) „Educational Evaluation in the Contemporary World", referat wprowadzający na międzynarodowym sympozjum „De didactica general y didacticas especiales" w La Manga del Mar Menor w Hiszpanii, 27 września – 2 października.

MacDonald B. (1982b) *Mandarins and Lemons — The Executive Investment in Program Evaluation*, „Phi Delta Kappa CEDR Quarterly" nr 2.

MacDonald B. (1984) „Democratic Evaluation in Practice", referat przedstawiony w czasie obrad Second National Evaluation Conference w Melbourne w Australii, 26–27 lipca.

MacDonald B., Adelman C., Kushner S., Walker R. (1982) *Bread and Dreams: A Case Study of Bilingual Schooling in the USA*, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B., Argent M. J., Elliott J., May N., Naylor J. T., Norris N. (1987) *Police Probationer Training in England and Wales*, Home Office, London.

MacDonald B., Jenkins D. (1980) „Understanding Computer Assisted Learning. Final Report of the Evaluation of the National Development Programme in Computer Assisted Learning", tekst powielony, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B., Jenkins D., Kemmis S., Tawney D. (1975) *The Programme at Two. An Evaluation Report on the National Development Programme in Computer Assisted Learning*, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B., Jenkins D., Kemmis S., Verma G. (1979) *Racism and Educational Evaluation*, w: G. Verma, C. Bagley (red.), *Race, Education and Identity*, Macmillan, London.

MacDonald B., Jenkins D., Kushner S., Logan T., Norris N. (1981) *Evaluators at Work: Report of a Workshop*, IFAPLAN, Köln, z upoważnienia Komisji Wspólnoty Europejskiej.

MacDonald B., Norris N. (1981) „Looking up for a Change — Political Horizons in Policy Evaluation", tekst powielony, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B., Parlett M. (1973) *Rethinking Evaluation: Notes from the Cambridge Conference*, „Cambridge Journal of Education" nr 2.

MacDonald B., Sanger J. (1982) *Just for the Record? Notes Towards a Theory of Interviewing in Evaluation*, w: E. R. House (red.), *Evaluation Studies Review Annual*, vol. 7, Sage, Beverly Hills.

MacDonald B., Stake R. E. (1974) „Confidentiality: Procedures and Principles of the UNCAL Evaluation with Respect to Information about Projects in the National Development Programme in Computer Assisted Learning", tekst powielony, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B., Walker R. (red.) (1974) *SAFARI I: Innovation, Evaluation, Research and the Problem of Control*, Centre for Applied Research in Education, University of East Anglia, Norwich.

MacDonald B., Walker R. (1975) *Case Study and the Social Philosophy of Educational Research*, „Cambridge Journal of Education" nr 1.

Macpherson C. B. (1966) *The Real World of Democracy*, Oxford University Press, Oxford.

Mager R. F. (1962) *Preparing Objectives for Programmed Instruction*, Fearon, San Francisco.

Malinowski B. (1987) *Argonauci Zachodniego Pacyfiku*, w: *Dzieła* t. 3, tłum. B. Olszewska-Dyoniziak, S. Szymkiewicz, PWN, Warszawa.

Marshall C., Rossman G. (1989) *Designing Qualitative Research*, Sage, Newbury Park Cal.

Maxwell J. A. (1992) *Understanding and Validity in Qualitative Research*, „Harward Educational Review" t. 63, s. 279–300.

McCutcheon G. (1978) *Of Solar Systems, Responsibility and Basics: An Educational Criticism of Mr. Clement's Fourth Grade*, w: G. Willis (red.), *Qualitative Evaluation: Concepts and Cases in Curriculum Criticism*, McCutchan, Berkeley.

McKillip J. (1987) *Need Analysis: Tools for the Human Sciences and Education*, Sage, Newbury Park – London.

McLaughlin M. W. (1975) *Evaluation and Reform*, Ballinger, Cambridge Mass.

Miles M. B., Huberman A. M. (1984) *Qualitative Data Analysis: A Sourcebook of New Methods*, Sage, Beverly Hills.

Morris E. (1979) *The Rise of Theodore Roosevelt*, Coward, McCann and Geognegan, New York.

Mullen E., Dumpson J. i in. (1972) *Evaluation for Social Interventions*, Jossey-Bass, San Francisco.

Nash R. (1973) *Classrooms Observed: The Teacher's Perception and the Pupil's Performance*, Routledge and Kegan, London.

Nevo D. (1981) *The Evaluation of a Multi-Dimensional Project*, w: A. Lewy, S. Kugelmass, G. Ben-Shakar, N. Blass, R. F. Boruch, D. J. Davis, B. Nevo, D. Nevo, P. Tamir, I. Zak; *Decision Oriented Evaluation in Education: The Case of Israel*, International Science Services, Philadelphia.

Nilsson N., Hogben D. (1983) *Metaevaluation*, w: E. R. House (red.), *Philosophy of Evaluation*, Evaluation Research Society, Jossey-Bass, London – San Francisco.

Nisbet R. (1976) *Sociology as an Art Form*, Heinemann, London.

Norris N. (1984) „The Context and Tradition of Contemporary Educational Evaluation", tekst powielony, Centre for Applied Research in Education, University of East Anglia, Norwich.

Norris N. (1990) *Understanding Educational Evaluation*, Kogan Page, London.

Owens T. R. (1973) *Educational Evaluation by Adversary Proceeding*, w: E. R. House (red.), *School Evaluation: The Politics and Process*, McCutchan, Berkeley.

Parlett M. (1975) „A Descriptive Analysis of a Liberal Arts College: Methods and Interpretation in a Commissioned Case Study", opracowanie przygotowane na sympozjum „The Applied Anthropology of School Organization: An International Comparison of Descriptive Methodologies and Utilitarian Objectives", przedstawione również na dorocznej sesji Society for Applied Anthropology w Amsterdamie.

Parlett M., Dearden G. (1977) *Introduction to Illuminative Evaluation: Studies in Higher Education*, Pacific Soundings Press, Cardiff-by-the Sea Cal.

Parlett M., Hamilton D. (1977) *Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes*, w: D. Hamilton i in. (red.), *Beyond the Numbers Game: A Reader in Educational Evaluation*, Macmillan, London.

Patton M. Q. (1978) *Utilization Focused Evaluation*, Sage, Beverly Hills.

Patton M. Q. (1980) *Qualitative Evaluation Methods*, Sage, Beverly Hills.

Patton M. Q. (1981) *Creative Evaluation*, Sage, Newbury Park – London.

Patton M. Q. (1982) *Practical Evaluation*, Sage, Newbury Park – London.

Patton M. Q. (1990) *Qualitative Evaluation and Research Methods*, Sage, Newbury Park Cal.

Paulston R. G. (w druku) „Paradigm Conflict in the Assessment of Educational Reform: Evaluation as Social Critique", „Educational Evaluation and Policy Analysis".

Peshkin A. (1986) *God's Choice*, University of Chicago Press, Chicago.

Petrie H. G. (1979) *COPE Begins Second Evaluation Cycle*, „Illini Week", October 12.

Pick C., Walker R. (1976) *Other Rooms, Other Voices*, Ford Safari Project, Centre for Applied Research in Education, University of East Anglia, Norwich.

Polanyi M. (1962) *Personal Knowledge: Towards a Post-critical Philosophy*, University of Chicago Press, Chicago.

Popham W. J. (1975) *Educational Evaluation*, Prentice-Hall, Englewood Clifs N. J.

Provus M. M. (1971) *Discrepancy Evaluation*, McCutchan, Berkeley.

Quinn W., Quinn N. (1992) *Buddy Evaluation*, North Central Regional Educational Laboratory, Oakbrook Ill.

Ragin C. C. (1987) *The Comparative Method*, University of California Press, Berkeley.

Rawls J. A. (1994) *Teoria sprawiedliwości*, tłum. M. Panufnik, J. Pasek, A. Romaniuk, PWN, Warszawa.

Rippey R. M. (red.) (1973) *Studies in Transactional Evaluation*, McCutchan, Berkeley.

Rivlin A. M. (1971) *Systematic Thinking for Social Action*, Brookings Institution, Washington.

Robson C., Sebba J., Mittler P., Davies G. (1988) *Inservice Training and Special Educational Needs: Running Short School-focused Courses*, Manchester University Press, Manchester.

Ross H. L. (1973) *Law, Science and Accidents: The British Road Safety Act of 1967*, „Journal of Legal Issues" t. 2, s. 1–75.

Rossi P. H., Freeman H. E., Wright S. R. (1979) *Evaluation: A Systematic Approach*, Sage, Beverly Hills.

Rubin H. J. (1983) *Applied Social Research*, Merrill, Columbus Ohio.

Rubin H. J. (1994) *There Aren't Going to Be Any Bakeries Here if There Is No Money to Afford Jellyrolls: The Organic Theory of Community Based Development*, „Social Problems" nr 3.

Rubin I. S. (1977) *Universities in Stress: Decision Making Under Conditions of Reduced Resources*, „Social Science Quarterly" t. 58, s. 242–254.

Rubin I. S. (1992) *Budget Reform and Political Reform: Conclusions from Six Cities*, „Public Administration Review" t. 52, s. 454–466.

Rumelhart D. E., Ortony A. (1977) *The Representation of Knowledge in Memory*, w: R. C. Anderson, R. J. Spiro, W. E. Montague (red.), *Schooling and the Aquisition of Knowledge*, Erlbaum, Hillsdale N. J.

Satcher D., Koscoff J., Fink A. (1980) *Results of a Needs Assessment Strategy in Developing a Family Practice Program in an Inner-City Community*, „Journal of Family Practice" t. 10, s. 871–879.

Saunders S. (1990) *An Evaluation of a Residential PSE Course for an Academic A-Level Class*, „Pastoral Care in Evaluation" nr 8.

Schön D. (1983) *The Reflective Practitioner: How Professionals Think in Action*, Basic Books, New York.

Scriven M. (1967) *The Methodology of Evaluation*, w pracy zbiorowej *Perspectives of Curriculum Evaluation*, „AERA Monograph Series on Curriculum Evaluation" nr 1, Rand McNally, Chicago.

Scriven M. (1972a) *Pros and Cons about Goal-free Evaluation*, „Evaluation Comment" nr 4.

Scriven M. (1972b) „The Pathway Comparison Model of Evaluation", tekst powielony, University of California, Berkeley.

Scriven M. (1973) *Goal-free Evaluation*, w: E. R. House (red.), *School Evaluation: The Politics and Process*, McCutchan, Berkeley.

Scriven M. (1974) *Evaluation Perspectives and Procedures*, w: W. J. Popham (red.), *Evaluation in Education: Current Applications*, McCutchan, Berkeley.

Scriven M. (1976a) „Bias Control Systems in Evaluation", referat przedstawiony na dorocznej sesji American Educational Research Association w San Francisco.

Scriven M. (1976b) *Evaluation Bias and Its Control*, w: G. V. Glass (red.), *Evaluation Studies Review Annual*, vol. 1, Sage, Beverly Hills.

Scriven M. (1976c) *Maximizing the Process of Causal Investigations: The Modus Operandi Method*, w: G. V. Glass (red.), *Evaluation Studies Review Annual*, vol. 1, Sage, Beverly Hills.

Sharp R., Green A. (1975) *Education and Social Control: A Study in Progressive Primary Education*, Routledge and Kegan, London.

Sheehan N. (1988) *A Bright and Shining Lie: John Vann and America in Vietnam*, Random House, New York.

Simons H. (1971) *Innovation and the Case Study of Schools*, „Cambridge Journal of Education" nr 3.

Simons H. (1977) *Building a Social Contract: Negotiation and Participation in Condensed Field Research*, w: N. Norris (red.), *Theory in Practice*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Simons H. (red.) (1980) *Towards a Science of the Singular: Essays about Case Study in Educational Research and Evaluation*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Smith L. M. (1978) *An Evolving Logic of Participant Observation, Educational Ethnography and Other Case Studies*, w: L. Shuman (red.), *Review of Research in Education*, vol. 6, Peacoc, Itasca Ill.

Smith L. M., Keith P. M. (1971) *Anatomy of Educational Innovation*, Wiley, New York.

Smith L. M., Pohland P. A. (1974) *Education, Technology, and the Rural Highlands*, w pracy zbiorowej *Four Evaluation Examples: Anthropological, Economic, Narrative, and Portrayal*, „AERA Monograph Series on Curriculum Evaluation" nr 7, Rand McNally, Chicago.

Spiro R. J., Vispoel W. P., Schmitz J. G., Samarapungavan A., Boerger A. E. (1987) *Knowledge Acquisition for Application: Cognitive Flexibility and Transfer in Complex Content Domains*, w: B. C. Britton (red.), *Executive Control Processes*, Erlbaum, Hillsdale N. J.

Stake R. E. (1967a) *The Countenance of Educational Evaluation*, „Teacher College Record" t. 68, s. 523-540.

Stake R. E. (1967b) *Toward a Technology for the Evaluation of Educational Programs*, w pracy zbiorowej *Perspectives of Curriculum Evaluation*, „AERA Monograph Series on Curriculum Evaluation" nr 1, Rand McNally, Chicago.

Stake R. E. (1972) „Responsive Evaluation", tekst powielony, University of Illinois, Urbana-Champaign.

Stake R. E. (1975) *Evaluating the Arts in Education: A Responsive Approach*, Merrill, Columbus Ohio.

Stake R. E. (1976) *Evaluation Educational Programmes: The Need and the Response*, OECD Publications Center, Washington.

Stake R. E. (1977) *An Approach to the Evaluation of Instructional Programs (Program Portrayal vs. Analysis)*, w: D. Hamilton i in. (red.) *Beyond the Numbers Game: A Reader in Educational Evaluation*, Macmillan, London.

Stake R. E. (1980) *The Case Study Method in Social Inquiry*, w: H. Simons (red.), *Towards a Science of the Singular: Essays about Case Study in Educational Research and Evaluation*, Centre for Applied Research in Education, University of East Anglia, Norwich.

Stake R. E. (1981) *Setting Standards for Educational Evaluators*, „Evaluation News" nr 2.

Stake R. E. (1988) *Case Study Methods in Educational Research: Seeking Sweet Water*, w: R. M. Jaeger (red.), *Complementary Methods for Research in Education*, American Educational Research Association, Washington.

Stake R. E., Easley J. (red.) (1978) *Case Studies in Science Education, a Project for the National Science Foundation Conducted by the Centre for Instructional Research and Curriculum Evaluation and the Committee on Culture and Cognition*, CIRCE, University of Illinois, Urbana-Champaign.

Stake R. E., Gjerde C. (1974) *An Evaluation of TCITY, the Twin-City Institute for Talented Youth*, w pracy zbiorowej *Four Evaluation Examples: Anthropological, Economic, Narrative and Portrayal*, „AERA Monograph Series on Curriculum Evaluation" nr 7, Rand McNally, Chicago.

Stake R. E., Trumbull D. J. (1982) *Naturalistic Generalizations*, „Review Journal of Philosophy and Social Science" t. 7, s. 1–12.

Stanley J. C. (1972) *Controlled Field Experiments as a Model for Evaluation*, w: P. H. Rossi, W. Williams (red.), *Evaluating Social Programs*, Seminar Press, New York.

Stenhouse L. (1975) *An Introduction to Curriculum Research and Development*, Heinemann, London.

Stenhouse L. (1984) *Library Access, Library Use and User Education in Academic Sixth Forms: An Autobiographical Account*, w: R. G. Burgess (red.), *The Research Process in Educational Settings: Ten Case Studies*, Falmer, London.

Stenzel N. (1979) *Committee Hearings as an Evaluation Format*, Northwest Regional Educational Laboratory, Portland Or.

Stouffer S. A. (1941) *Notes on the Case-Study and the Unique Case*, „Sociometry" t. 4, s. 349–357.

Strauss A. L., Corbin J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage, Newbury Park Cal.

Stufflebeam D. L. (1969) *Evaluation as Enlightenment for Decision Making*, w: W. H. Beatty (red.), *Improving Educational Assessment and an Inventory for Measures of Affective Behavior*, National Education Association, Washington

Stufflebeam D. L. (1972) *The Relevance of the CIPP Evaluation Model for Educational Accountability*, „SRIS Quarterly" t. 5, s. 3–6.

Stufflebeam D. L. (1973) *An Introduction to the PDK Book*, w: B. Worthen, J. R. Sanders (red.), *Educational Evaluation: Theory and Practice*, Charles A. Jones, Worthington Ohio.

Stufflebeam D. L. (1974) *Meta-Evaluation*, „Occasional Paper" nr 3, Western Michigan University, Kalamazoo.

Stufflebeam D. L., Foley W. J., Gephart W. J., Guba E. G., Hammon R. L., Merriman H. O., Provus M. M. (1971) *Educational Evaluation and Decision--making*, Peacock, Itasca Ill.

Stufflebeam D. L., Webster W. J. (1980) *An Analysis of Alternative Approaches to Education*, „Educational Evaluation nad Policy Analysis" nr 3.

Suchman E. A. (1967) *Evaluative Research: Principles in Public Service and Action Programs*, Russell Sage Foundation, New York.

Tallmadge G. K. (1977) *Joint Dissemination Review Panel Ideabook*, US Government Printing Office, Washington.

Taylor P. W. (1961) *Normative Discourse*, Prentice-Hall, Englewood Cliffs N.J.

Tobin J. (1989) *Preschool in Three Cultures*, Yale University Press, New Haven.

Tyler R. W. (1950) *Basic Principles of Curriculum and Instruction*, University of Chicago Press, Chicago.

Tyler R. W., Lapan S., Moore J., Rivers C. W., Skibo D. (1978) *Impact of Minimum Competence Testing in Florida*, „Today's Education", September–October.

US General Accounting Office (1978) *Assessing Social Program Impact Evaluations: A Checklist Approach*, US General Accounting Office, Washington.

Vallance E. (1978) *Scanning Horizons and Looking at Weeks: A Critical Description of „The Great Plains Experience"*, w: G. Willis (red.), *Qualitative Evaluation: Concepts and Cases in Curriculum Criticism*, McCutchan, Berkeley.

Van Maanen J. (1988) *Tales of the Field: On Writing Ethnography*, University of Chicago Press, Chicago.

Walker R. (1982) *The Use of Case Studies in Applied Research and Evaluation*, w: A. Hartnett (red.), *The Social Sciences in Educational Studies*, Heinemann, London.

Walker R., MacDonald B. (1975) *The Press We Deserve*, „Time Educational Supplement" nr 6.

Wax R. H. (1971) *Doing Fieldwork: Warning and Advice*, University of Chicago Press, Chicago.

Webb E. J., Campbell D. T., Schwartz R., Sechrest L. (1966) *Unobtrusive Measures: Nonreactive Research in the Social Sciences*, Rand McNally, Chicago.

Weiss C. H. (red.) (1972) *Evaluating Action Programs: Readings in Social Action and Evaluation*, Allyn and Bacon, Boston.

Weiss C. H. (1975) *Evaluation Research in the Political Context*, w: E. L. Struening, M. Guttentag (red.), *Handbook for Evaluation Research*, vol. 1, Sage, Beverly Hills.

Weiss C. H. (red.) (1977) *Using Social Research in Public Policy Making*, Lexington Books, Lexington Mass.

Weiss C. H. (1986) *Towards the Future of Stakeholder Approaches in Evaluation*, w: E. R. House (red.), *New Directions in Educational Evaluation*, Falmer, Lewes.

Welch W. (1978) „Goal-free Formative Evaluation — an Example", referat przedstawiony na sesji American Education Research Association w marcu w Toronto.

Wolcott H. F. (1977) *Teachers vs. Technocrats*, Center for Educational Policy and Management, University of Oregon, Eugene Or.

Wolf R. L. (1975) *Trial by Jury: The Process*, „Phi Delta Kappa", November.

Wolf R. L. (1979) *The Use of Judical Evaluation Methods in the Formation of Educational Policy*, „Educational Evaluation and Policy Analysis" t. 1, s. 19–28.

Wolf R. L., Tymitz B. (1977) „Whatever Happened to the Giant Wombat: An Investigation of the Impact of the Ice Age Mammals and Emergency of Man Exhibit", tekst powielony, National Museum of Natural History, Smithsonian Institute, Washington.

Worthen B. R. (1975) *Competencies for Educational Research and Evaluation*, „Educational Researcher" nr 1.

Worthen B. R., Sanders J. R. (1973) *Educational Evaluation: Theory and Practice*, Wadsworth, Belmont Cal.

Wright G. H. von (1971) *Explanation and Understanding*, Routledge and Kegan, London.

Yin R. K. (1989) *Case Study Research: Design and Methods*, Sage, Newbury Park Cal.

Yin R. K. (1992) „Evaluation: A Singular Craft", referat przedstawiony na dorocznej sesji American Evaluation Association w listopadzie w Seattle.

Słownik ważniejszych pojęć

(opracował Leszek Korporowicz)

Analiza systemowa w ewaluacji *(system analysis).* — Jest to próba określenia skuteczności, wydajności czy produktywności ewaluowanego programu przez ustalenie zależności między „wkładem" w konkretne działanie a jego końcowymi rezultatami. Analiza ta łączy się na ogół z odpowiedzią na pytanie, czy spowodowana przez program zmiana zgodna jest ze swoim projektem i czy następuje w pożądanym kierunku. Ewaluacja systemowa rozpoznawać więc musi zasadnicze funkcje programu oraz dysponować zestawem jasno określonych kryteriów.

Ewaluacja *(evaluation).* — Istnieje wiele bardzo różnych definicji ewaluacji w większym lub mniejszym stopniu akceptujących jej oceniający charakter. Dla celów roboczych przyjąć jednak można definicję jednego z międzynarodowych komitetów ewaluacyjnych (Joint Committee on Standards for Educational Evaluation, 1981), który określił ewaluację jako „systematyczne badanie wartości lub zalet jakiegoś obiektu". Najczęściej obiektem tym jest konkretne działanie lub program społeczny, a termin „badanie" odzwierciedlać może bardziej analityczno-opisowe niż oceniające przeznaczenie ewaluacji, której celem jest na ogół usprawnienie lub lepsze zrozumienie ewaluaowanych działań.

Ewaluacja demokratyczna *(democratic evaluation).* — Nastawiona jest na uwzględnienie różnych interesów i punktów widzenia uczestniczących w niej stron zgodnie z ich własną definicją i interpretacją. Ewaluator jest w tym przypadku nie tyle osobą formułującą oceny, ile pośrednikiem w wymianie informacji, przekonań i doświadczeń, i to w sposób, który nie pozbawia nikogo kontroli nad ich użyciem i interpretacją.

Ewaluacja dialogiczna *(responsive evaluation).* — Dokonuje się przez ciągłą wymianę informacji, obserwacji i spostrzeżeń pomiędzy ewaluatorem i ewaluowanymi, odpowiada też na zapotrzebowanie informacyjne odbiorców.

Utrzymanie takiej interakcji ma na celu ustawiczną korektę jednostronnego widzenia programu, respektowanie reakcji jego uczestników oraz weryfikację jakości gromadzonych danych.

Ewaluacja formatywna *(formative evaluation).* — Realizowana jest w trakcie trwania ewaluowanego programu i wspomagać ma jego realizację poprzez bieżącą analizę danych w celu ewentualnej korekty obserwowanych działań.

Ewaluacja konkluzywna *(summative evaluation).* — Koncentruje się na analizie rezultatów lub skutków programu poprzez wykrycie wszystkich znajdujących się w polu analizy konsekwencji o planowanym i nie planowanym charakterze. Wnioski z tych analiz w postaci konkluzji uzupełnione są często rekomendacjami dotyczącymi przyszłych programów lub decyzji.

Ewaluacja „naturalistyczna" *(naturalistic evaluation).* — Pojęcie to funkcjonuje od lat siedemdziesiątych na określenie wszystkich alternatywnych koncepcji studiów ewaluacyjnych, których wspólnym mianownikiem jest integralne, sytuacyjne i kontekstualne ujęcie przedmiotu ewaluacji, preferowanie metod jakościowych nastawionych na dotarcie do doświadczeń uczestników programu w ich „naturalnej", potocznej postaci, a więc do tego, jak oni sami je postrzegają i przeżywają.

Ewaluacja wyjaśniająca *(illuminative evaluation).* — Jej celem jest „oświetlenie" doświadczeń uczestników ewaluowanego programu. Pozwolić to ma na opisanie, zrozumienie i wyjaśnienie dynamiki jego działania, wpływu, jaki ma na postrzeganie wprowadzanych innowacji oraz ról uczestników programu, ocenę zalet i wad programu z punktu widzenia potrzeb i systemu wartości jego uczestników.

Podejście *goal-free.* — Ma ono na celu wyeliminowanie wpływu, jaki na ewaluatora może mieć wcześniejsza znajomość celów ewaluowanego programu i w konsekwencji ujednostronnienie oraz sprofilowanie jego uwagi. Zabieg ten pozwolić ma na bezstronne i bezzałożeniowe zainteresowanie wszystkimi, w tym również ubocznymi aspektami obserwowanych działań.

Podejście typu „inspekcja kwalifikacji" *(professional review approach).* — Wyróżniającą cechą tego typu ujęcia jest to, że standardy i kryteria ewaluacyjne ustanawiają osoby będące specjalistami w dziedzinie i działaniach, które poddawane są ewaluacji. Najczęściej stanowią też oni podstawowy trzon zespołu ewaluatorów.

Podejście typu „krytyka artystyczna" *(art-criticism approach).* — Ujęcie to adaptuje sposoby oceny dzieł sztuki. Jego zwolennicy wychodzą z założenia, że ewaluacja, podobnie jak krytyka artystyczna, nie jest dążeniem do

negatywnej oceny przedmiotu, ale prowadzi do oświetlenia jego specyficznych jakości i wartości, do ujawnienia tego, co w konkretnym programie wyjątkowe i znaczące, w odróżnieniu od tego, co nietwórcze i trywialne.

Podejście *quasi*-sądowe *(quasi-legal approach)*. — Polega ono na stosowaniu przez starannie wyselekcjonowane komisje procedur przypominających procedury sądowe, jak przesłuchanie stron, gromadzenie i weryfikacja świadectw. W „krzyżowym ogniu pytań", w toku przedstawiania i rozważania argumentów „za" i „przeciw", dąży się do konkretnych konkluzji, które mogłyby być podstawą formalnej oceny programu, działań lub kompetencji.

Podejście zoperacjonalizowanych zadań *(behavioral objectives approach)*. — Istotą tego ujęcia jest dążenie do określenia stopnia realizacji precyzyjnie wyznaczonych celów programu przejawiających się w konkretnych zachowaniach lub kompetencjach. Podejście to często posługuje się różnorodnymi zestawami testów w celu mierzalnego określenia swoich ocen.

Podejście zorientowane na podjęcie decyzji *(decision-making approach)*. — Podejście to nastawione jest na organizację procesu realizacji ewaluacji, jej struktury i formy w sposób mający ułatwić podjęcie decyzji. Tym, co wymaga odpowiedzi, to pytania zadawane przez decydentów; ewaluacja ma dostarczyć potrzebnych im informacji i ocenić alternatywne drogi działania.

Samoewaluacja *(self-evaluation)*. — Istotny element procesów demokratyzowania ewaluacji, prowadzona jest bowiem nie przez ekspertów, ale przez uczestników lub realizatorów ewaluowanych programów w dążeniu do podniesienia ich własnej samoświadomości i refleksyjności.

Studium przypadku *(case study)*. — Jest raczej rodzajem podejścia niż jedną z klasycznych metod, gdyż odwoływać się może zarówno do ilościowych, jak i jakościowych technik badawczych. Chociaż istnieją różne jego typy, zmierzające do odmiennych poziomów generalizacji, zasadniczym celem studium przypadku nie jest tworzenie teoretycznych uogólnień, ale zainteresowanie nim samym w wielorakości i złożoności jego uwarunkowań, jego specyfiką i niepowtarzalnością.

Triangulacja *(triangulation)*. — Zabieg metodologiczny polegający na "oświetlaniu" przedmiotu badań z możliwie różnych stron i uwzględnianiu odmiennych punktów widzenia w celu ich wzajemnej weryfikacji, dopełnienia, relatywizacji, ale i dookreślenia. Metoda często stosowana w studium przypadku.

Indeks

(opracował Leszek Korporowicz)

agenda zewnętrzna 110
analiza dokumentów 116, 168
analiza systemowa w ewaluacji 96–102, 235
autoteliczne studium przypadku 123, 136, 138, 141

badania interpretacyjne 130
badanie eksploracyjne 127
badania jakościowe 202, 203, 207, 288, 290
badania terenowe 121, 163, 167, 172, 176, 183, 184, 190, 193, 199, 244, 252

dane jakościowe 163, 167
definicje ewaluacji 52, 53, 150
demokracja liberalna 225, 226, 229, 241
dobór próby 141
dobór przypadku 138
doświadczenie osobiste 90, 134, 181, 203
doświadczenie uczniów 117
dyskursywne i niedyskursywne sposoby prezentacji 107

epistemologia konstruktywistyczna 136
etnograficzny etos badań 130
ewaluacja autokratyczna 128, 135, 138
ewaluacja biurokratyczna 128, 135
ewaluacja demokratyczna 32, 38, 45, 76, 82, 86, 91, 116, 235, 237, 242
ewaluacja dialogiczna 38, 40, 109, 115, 116, 139, 153
ewaluacja formatywna 53, 110, 154, 156, 197
ewaluacja jakościowa 163, 165
ewaluacja kadr 110
ewaluacja konkluzywna 53, 156
ewaluacja liberalno-demokratyczna 227
ewaluacja „naturalistyczna" 38, 39, 40, 42, 81, 86, 117
ewaluacja wyjaśniająca 38, 40, 75, 78, 117
ewaluator wewnętrzny 60
ewaluator zewnętrzny 60

funkcje ewaluacji 22, 35, 52, 53, 57, 247

generalizacja 41 132, 134, 138, 143, 144
grupa odniesienia 144, 228, 230, 231, 237, 238

holistyczne szacowanie programu 111

ideały i wartości społeczne 57, 62
identyfikacja przypadku 121
interes publiczny 232, 233, 234, 235, 237
interpretacja 137, 138
instruktażowe studium przypadku 125
instrumentalne studium przypadku 123, 136, 138
introspekcja 167, 168

język krytyki 107

klient 22, 23, 24, 36, 144, 149, 154, 184, 191