

# Rozdział I.

## Ewaluacja – teoria i metodologia

Paweł Kawalec

---

### 1. Istota ewaluacji

W stosunkowo krótkim czasie ewaluacja wyrosła na jeden z zasadniczych czynników kształtujących ogólnoswiatową debatę nad najtrudniejszymi wyzwaniami współczesnych społeczeństw, jak migracje, ubóstwo, zapewnienie adekwatnej opieki zdrowotnej, degradacja środowiska czy zmiana klimatyczna<sup>1</sup>. Wpłynęły na to z pewnością zmiany polityczne, jak realizacja interwencji publicznych w dużej skali w USA<sup>2</sup>, polityki publiczne Unii Europejskiej czy transformacja w Europie Środkowej i Wschodniej. Innym czynnikiem były zmiany kulturowe, charakterystyczne dla odchodzenia od idei modernizmu<sup>3</sup>. Istotną rolę odegrały także procesy globalizacyjne i kształtowanie się programów międzynarodowej współpracy, czego przykładem może być działalność The Campbell Collaboration, prowadzona w takich obszarach, jak praktyka medyczna czy programy społeczne oraz edukacyjne.

W Polsce badania ewaluacyjne spopularyzowały się w okresie przeobrażeń społecznych jako element nowej strategii zarządzania reformami w takich dziedzinach, jak edukacja, polityka społeczna, ochrona zdrowia i środowiska, a także usługi, szkolenia czy programy restrukturalizacji zatrudnienia. Częściowo ta zmiana spowodowana była wymogami międzynarodowych programów pomocy organizacyjnej i finansowej jako czynnik o charakterze systemowym<sup>4</sup>.

---

<sup>1</sup> I.F. Shaw, J. Greene i M. Mark, red., *Handbook of evaluation: policies, programs and practices*, Sage Publications, London 2013; C. Perez et al., *Changing gear in R&I: Green growth for jobs and prosperity in the EU. Report of the European Commission expert group "R&I policy framework for green growth & jobs"*, European Commission, Strasbourg 2016.

<sup>2</sup> Syntetyczne wnioski z przeglądu największych programów publicznych w USA podaje: D.J. Besharov, K.J. Baehler i J.A. Klerman, red., *Improving public services: international experiences in using evaluation tools to measure program performance*, Oxford University Press, New York 2017.

<sup>3</sup> P. Dahler-Larsen, *From Programme Theory to Constructivism: On Tragic, Magic and Competing Programmes*, „Evaluation” 2001, t. 7, nr 3, s. 331–349.

<sup>4</sup> L. Korporowicz, *Badania ewaluacyjne*, [w:] *Zarządzanie badaniami naukowymi i pracami rozwojowymi w jednostkach naukowych*, red. P. Kawalec i P. Lipski, Wydawnictwo LBS, Lublin

Wskazuje się wiele powodów, dla których ewaluacja – zarówno jako działalność badawcza, jak też praktyczna – wzbudza współcześnie duże zainteresowanie<sup>5</sup>:

- wykazanie zasadności i skuteczności przedsięwzięć, co ułatwia uzyskanie społecznego przyzwolenia na ich finansowanie ze środków publicznych;
- zwiększenie przejrzystości i rozliczalności wydatkowania środków publicznych, a także prywatnych, na podstawie przejrzystych i wiarygodnych ocen uzyskanych korzyści w stosunku do nakładów;
- standaryzacja kryteriów i procedur oceny dzięki wymianie informacji między różnorodnymi podmiotami;
- wprowadzanie regulacji prawnych (w ramach państw oraz na poziomie podmiotów międzynarodowych), w myśl których ewaluacja jest obowiązkowym elementem procedury realizacji różnorodnych programów publicznych.

Początków ewaluacji jako systematycznej oceny interwencji publicznych upatruje się na przełomie XVIII i XIX wieku<sup>6</sup>. Podanie pełnego obrazu historycznego procesu kształtowania się ewaluacji jest bardzo trudnym zadaniem z uwagi na zróżnicowanie jej rozwoju w odrębnych dziedzinach (polityka społeczna, edukacja, zarządzanie itp.) oraz w poszczególnych obszarach geograficznych, gdyż mimo dominacji rozwiązań wypracowanych w USA i Wielkiej Brytanii powstały także odmienne podejścia do ewaluacji, np. w krajach Ameryki Południowej i Afryki, a także programów politycznych realizowanych w poszczególnych krajach (np. Johnsona „Great Society” w USA czy program Fabiana w UK po II wojnie światowej). Dlatego niekiedy używa się określenia „rozliczne historie ewaluacji” (*multiple histories of evaluation*)<sup>7</sup>. Mimo tego zróżnicowania obserwuje się dwa szerokie procesy<sup>8</sup>. Pierwszym jest przechodzenie w praktyce ewaluacji od działań przyczynkarskich i niesystematycznych działań ku standaryzacji, specjalizacji i instytucjonalizacji. Znajduje to swój wyraz w powstawaniu instytucji specjalizujących się w ewaluacji programów publicznych (m.in. Polskie Towarzystwo Ewaluacji, Polska Agencja Rozwoju Przedsiębiorczości, Centrum Ewaluacji i Analiz Polityk Publicznych UJ oraz American Evaluation Association, World Bank Evaluation Unit, The Joint Comitee on Educational Evaluation, Empowerment Evaluation Insitute itp.). Drugi proces to przechodze-

---

2008, t. 2, s. 65; E.F. Gates, *Learning from seasoned evaluators: Implications of systems approaches for evaluation practice*, „Evaluation” 2017, t. 23, nr 2, s. 152–171.

<sup>5</sup> B. Szatur-Jaworska, red., *Ewaluacja w służbach społecznych*, Mazowieckie Centrum Polityki Społecznej, Warszawa 2010; B. Bjørnholt i F. Larsen, *The politics of performance measurement: 'Evaluation use as mediator for politics'*, „Evaluation” 2014, t. 20, nr 4, s. 400–411; D.J. Besharov, K.J. Baehler i J.A. Klerman, red., *Improving public services: international experiences in using evaluation tools to measure program performance*, op. cit.

<sup>6</sup> M. Theiss, *Historyczne źródła i czynniki rozwoju ewaluacji*, [w:] *Ewaluacja w służbach społecznych*, red. B. Szatur-Jaworska, op. cit., s. 15–22.

<sup>7</sup> I.F. Shaw, J. Greene i M. Mark, red., *Handbook of evaluation: policies, programs and practices*, op. cit., s. 9.

<sup>8</sup> M. Theiss, *Historyczne źródła i czynniki rozwoju ewaluacji*, op. cit.

nie w naukach społecznych od pomiaru stopnia realizacji założonych celów do opartej na dowodach oceny korzyści społecznych danej interwencji, jej zamierzonych i niezamierzonych konsekwencji, nowej wiedzy, którą ona generuje, oraz tego, na ile angażuje ona interesariuszy w proces zmiany.

W dziejach ewaluacji wyodrębnia się siedem okresów: do roku 1900 (początkowy), 1900–1930 (kształtowania się i testowania), 1930–1945 (Tylerowski), 1946–1957 („niewinności”), 1958–1972 (rozwoju), 1973–2004 (profesjonalizacji) oraz po 2005 (globalizacji i ekspansji)<sup>9</sup>. Za początek historii ewaluacji uznaje się niekiedy rok 1792 – zastosowanie przez W. Farisha metod ilościowych w ocenie egzaminów<sup>10</sup>. Pierwsze systematyczne zastosowania ewaluacji pojawiły się w programach interwencji publicznej jako odpowiedź na radykalne zmiany, jakie w XIX wieku dokonały się pod wpływem rewolucji przemysłowej w takich obszarach, jak zdrowie publiczne czy edukacja. W okresie początkowym te działania miały charakter doraźny, jednorazowy, dość przypadkowy i powierzane były dedykowanym komisjom, jak the Royal Commission of Inquiry into Primary Education in Ireland. Jednym z pierwszych systematycznie realizowanych programów ewaluacyjnych były zainicjowane w 1815 roku w USA działania Army Ordnance Department, których celem było ujednoczenie produkcji rzemieślniczej na potrzeby armii amerykańskiej. To działanie można traktować jako prekursora zarządzania naukowego F. Taylora i jego programu ewaluacji pracowników. Taylor na trwałe powiązał ewaluację ze standaryzacją oraz efektywnością. Stworzona z kolei przez R. Tylera kategoria „ewaluacji edukacyjnej” stała się – głównie za pośrednictwem the Progressive Education Movement oraz the Eight Year Study – podstawą oceny osiągnięcia przez uczniów zamierzonych efektów w programach edukacyjnych<sup>11</sup>. Początkowe zorientowanie ewaluacji na uzyskiwanie mierzalnych rezultatów i procedury pomiarowe często sprowadzało się do uchwycenia tylko zmiany zachowań jednostek, marginalizując ponadindywidualne efekty działania programu, a także wymiar instytucjonalny, organizacyjny i społeczny. W związku z tym pod koniec lat 60. dyskutowany był problem „redukcjonizmu behawioralnego”, eliminującego z badań inaczej określone cele, np. sferę doświadczeń emocjonalnych, zmiany systemów wartości

---

<sup>9</sup> G.F. Madaus i D.L. Stufflebeam, *Program evaluation: a historical overview*, [w:] *Evaluation models: viewpoints on educational and human services evaluation*, red. D.L. Stufflebeam, G.F. Madaus, i T. Kellaghan, Kluwer Academic Publishers, Boston 2000, 2. wyd., s. 3–18; D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, Jossey-Bass & Pfeiffer Imprints, Wiley, San Francisco 2014, 2. wyd., s. 30–40; R.L. Miller et al., *The Oral History of Evaluation: The Professional Development of Robert Stake*, „American Journal of Evaluation” 2016, t. 37, nr 2, s. 287–294.

<sup>10</sup> K. Hoskin, *The Examination, Disciplinary Power and Rational Schooling*, „History of Education” 1979, t. 8, nr 2, s. 144. Hoskin uznaje wręcz to odkrycie za przełom w modelowaniu matematycznym, twierdząc, że Farish jako pierwszy w dziejach nauki umożliwił naukowe podejście do badania jednostek.

<sup>11</sup> G.F. Madaus i D.L. Stufflebeam, *Educational Evaluation: Classic Works of Ralph W. Tyler*, Kluwer Academic Publishers, Boston 1989.

itd. Mniej znany jest w literaturze fakt, że w tym samym okresie istotne badania prowadził F. Znaniecki jako dyrektor programu badawczego realizowanego w latach 1930–1933 na Uniwersytecie Columbia. Raport z tych badań zupełnie inaczej ukierunkowywał ewaluację, wychodząc z całkiem innych założeń metodologicznych (badanie społecznego doświadczenia, postulat „współczynnika humanistycznego”)<sup>12</sup> opartych na interpretatywnym sposobie badania „faktów kulturowych” (np. adekwatności kształcenia kompetencji społecznych i kulturowych w kontekście wyzwań modernizującego się społeczeństwa amerykańskiego)<sup>13</sup>. Znaniecki znacząco rozszerzył zakres kryteriów ewaluacyjnych oraz uwzględniał zmiany wartości społecznych jako podstawy procesów i procedur wartościowania<sup>14</sup>.

The National Defense Education Act z 1958 roku (rok po starcie rosyjskiego Sputnika) był przełomem w angażowaniu środków publicznych na poziomie centralnym w USA i zainicjował szereg szeroko zakrojonych programów. Na zmianę klasycznych modeli ewaluacji wpłynęły prace L.J. Cronbacha<sup>15</sup>, M. Scrivena<sup>16</sup> oraz R. Stake’a<sup>17</sup>. Ich efektem było rozszerzenie zakresu danych w badaniach ewaluacyjnych z uwzględnieniem danych jakościowych, odejście od badania wyłącznie rezultatów analizowanych programów i zwrócenie uwagi na proces realizacji programu, jego złożoność i kontekst, wprowadzenie analizy o charakterze całościowym i wieloaspektowym w celu zapewnienia adekwatności ewaluacji w stosunku do potrzeb społecznych.

Początek lat 70. to okres wyraźnego kształtowania się ewaluacji jako odrębnej dyscypliny, czego wyrazem jest powstawanie czasopism naukowych (m.in. *Educational Evaluation and Policy Analysis*, *Evaluation Review*, *Evaluation and Program Planning*), stowarzyszeń i profesjonalnych organizacji (m.in. the Evaluation Network, the Evaluation Research Society), a także programów uniwersyteckich (m.in. University of Illinois, Stanford University, Boston College, UCLA, the University of Minnesota)<sup>18</sup>. Przełomu w tym, jakie ewaluacja pełni funkcje i role społeczne, dokonała konferencja zorganizowana w 1972 roku przez Churchill College w Cambridge, w której uczestniczyli ewaluatorzy oraz przedstawiciele

---

<sup>12</sup> Por. też studium przeglądowe: E. Reale et al., *A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research*, „Research Evaluation” 2017; M. Ochsner, S.E. Hug i H.D. Daniel, red., *Research Assessment in the Humanities*, Springer International Publishing, Cham 2016.

<sup>13</sup> F. Znaniecki, *Education and social change*, red. E. Hałas i P. Lang, Frankfurt am Main 1998.

<sup>14</sup> L. Korporowicz, *Badania ewaluacyjne*, op. cit.

<sup>15</sup> L.J. Cronbach, *Course Improvement Through Evaluation*, „Teachers College Record” 1963, t. 64, nr 8, s. 672–683.

<sup>16</sup> M. Scriven, *The Methodology of Evaluation*, [w:] *Perspectives of curriculum evaluation*, red. R.W. Tyler, R.M. Gagne i M. Scriven, Rand-McNally, Chicago 1967, s. 39–83.

<sup>17</sup> R.E. Stake, *The countenance of educational evaluation*, Teachers College Record 1967, t. 68, s. 523–540.

<sup>18</sup> Pełny przegląd chronologiczny podaje m.in. D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 5.

wielu urzędów i fundacji<sup>19</sup>. Wyraz dała mu prezentowana w czasie konferencji praca M. Parletta i D. Hamiltona *Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes*<sup>20</sup>, która postulowała podniesienie roli społecznej ewaluatorów, podejmowanie dialogu z różnymi audytoriami i uczestnikami procesu ewaluacyjnego, „oświetlanie” złożoności ludzkich doświadczeń towarzyszących realizacji programu, rozumiejącej analizy organizacyjnych, psychologicznych i kulturowych aspektów ewaluowanych działań i samej ewaluacji oraz przejrzystej komunikatywności ze wszystkimi interesariuszami. Zainicjowano ewaluację dialogiczną jako alternatywę dla obiektywistycznie i pomiarowo zorientowanej ewaluacji z lat 30., która później rozwijana była przez takich teoretyków ewaluacji, jak R. Stake, E.G. Guba i Y.S. Lincoln.

Stopniowo coraz większego znaczenia nabierała metaewaluacja<sup>21</sup> oraz – kształtujące się od początku lat 80. – standaryzacje ewaluacji, głównie programów edukacyjnych. Dynamicznie rozwijała się również metodologia ewaluacji w odpowiedzi na liczne wyzwania i coraz szerszy zakres zastosowań ewaluacji. Nowe zarządzanie publiczne zainicjowało szereg zmian w ewaluacji, o których mowa w podpunkcie 1.5, lecz również ograniczyło środki na ewaluacje szeroko zakrojonych programów publicznych. Ważną datą jest rok 1995, gdy American Evaluation Association zorganizowało konferencję ukierunkowaną na międzynarodową współpracę w dziedzinie ewaluacji, której efektem było m.in. powstanie ponad 20 towarzystw ewaluacyjnych na świecie.

## 1.1. Pojęcie i typy ewaluacji

W potocznym sensie ewaluację najogólniej można charakteryzować jako pewien rodzaj oceniania. Na tym tle istotne charakterystyki oceniania ewaluacyjnego podaje tabela 1.1.

Ewaluacja jednak w sensie ścisłym jest szerszym procesem niż samo ocenianie<sup>22</sup>. Ponadto, znane są również podejścia, w których element oceny nie jest w ogóle elementem procesu ewaluacji. Ewaluacja oznacza również samą dyscyplinę badawczą<sup>23</sup>. D. Stufflebeam i in. zróżnicowanie pojęć ewaluacji upatrują w złożonej genezie tej kategorii, która we wczesnych okresach odnoszona była do oceny osiągnięć na tle jasno zdefiniowanych celów, w innym kontekście

<sup>19</sup> L. Korporowicz, *Badania ewaluacyjne*, op. cit.

<sup>20</sup> M. Parlett i D. Hamilton, *Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes*, Centre for Research in Educational Sciences, University of Edinburgh, Edinburgh 1972.

<sup>21</sup> C.L.S. Coryn et al., *A Decade of Research on Evaluation: A Systematic Review of Research on Evaluation Published Between 2005 and 2014*, „American Journal of Evaluation” 2017, s. 1–19.

<sup>22</sup> K. Olejniczak, *Mechanizmy wykorzystania ewaluacji: studium ewaluacji średniookresowych INTERREG III*, Wydawnictwo Naukowe SCHOLAR, Warszawa 2008.

<sup>23</sup> M. Scriven, *Evaluation as a discipline*, „Studies in Educational Evaluation” 1994, t. 20, nr 1, s. 147–166.

natomiast używana była w odniesieniu do przeprowadzania testów lub eksperymentów, jak w przypadku rolnictwa czy psychologii eksperymentalnej<sup>24</sup>. Mimo zróżnicowania rozumienia celów ewaluacji, jej modeli, orientacji teoretycznych czy zastosowań można wskazać wspólny rdzeń pojęcia ewaluacji<sup>25</sup>.

Tabela 1.1

### Ocenianie potoczne a ocenianie ewaluacyjne

Wymiary porównania	Ocenianie nieewaluacyjne	Ocenianie ewaluacyjne
Kto dokonuje oceny?	Każdy, amator	Profesjonalista
Jakie są kryteria oceny?	Nieustalone, niezdefiniowane, przypadkowe, niejasne, zmienne	Ustalone, zdefiniowane, określone przed przystąpieniem do procesu zbierania informacji i oceniania
Informacje o przedmiocie oceny	Wrywkowe, zasłyszane, subiektywne	Rzetelne, obiektywne, zebrane metodycznie i według kryteriów oceny
W jaki sposób dokonuje się oceny?	Intuicyjny, momentalny, przypadkowy	Metodyczny, przemyślany, czasochłonny, według ustalonych procedur porównywania informacji o przedmiocie oceny z kryteriami oceny i wyciągania wniosków z takich porównań według ustalonych zasad

Źródło: opracowanie na podstawie: R. Szarfenberg, *Zagadnienia terminologiczne*, [w:] *Ewaluacja w służbach społecznych*, red. B. Szatur-Jaworska, op. cit., s. 24–37.

Jedno z najwcześniejszych i najlepiej znanych pojęć ewaluacji odnosi się do określenia, czy osiągnięte zostały zamierzone cele<sup>26</sup>. Ma ono jednak liczne ograniczenia, a nawet może się okazać kontrproduktywne, jak np. przy obiektywnej ewaluacji produktów czy usług konsumenckich. Jeden z problemów tego pojęcia wynika z faktu, że cele mogą być błędnie sformułowane, dysfunkcyjne, nieistotne, nieukierunkowane na potrzeby beneficjentów lub motywowane korzyścią finansową. W tym pojęciu ewaluacja ma także ograniczone możliwości wpływania na zmiany w samym programie, gdyż skoncentrowana jest przede wszystkim na rezultatach zakończonego już programu, z pominięciem na ogół rezultatów niezamierzonych.

Oficjalne rozporządzenia Unii Europejskiej z zakresu polityki budżetowej i polityki regionalnej na ogół nie formułują definicji ewaluacji wprost, lecz ograniczają się do opisu jej celów i dopuszczalnych odmian oraz podają cechy, które

<sup>24</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 6; H. White, *An introduction to the use of randomised control trials to evaluate development interventions*, „Journal of Development Effectiveness” 2013, t. 5, nr 1, s. 30–49.

<sup>25</sup> L. Korporowicz, *Badania ewaluacyjne*, op. cit., s. 68.

<sup>26</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 6.

powinna spełniać, jak analityczny i systematyczny charakter badania, rzetelność (*reliability*), użyteczność w zastosowaniach praktycznych oraz zorientowanie na zrównoważony rozwój<sup>27</sup>.

Definicje ewaluacji na ogół odnoszą się do czterech zasadniczych obszarów: systematyczne badanie, kryteria uzasadnione na podstawie jawnie przyjętego i uprawnionego systemu wartości, ewaluanda (polityki, programy, projekty, interwencje) oraz zamierzony cel lub wykorzystanie (np. rozliczalność, uczenie się, rozwój)<sup>28</sup>.

Stufflebeam i in. posługują się zmodyfikowaną wersją definicji wprowadzonej w 1994 roku przez the Joint Committee on Standards for Educational Evaluation: „systematyczne badanie jakości lub zalet jakiegoś obiektu”<sup>29</sup>. Przewagą tej definicji ewaluacji na ogólnym poziomie jest jej zwięzłość oraz zgodność z potocznymi słownikowymi definicjami ewaluacji.

Zgodnie z tą definicją ewaluacja angażuje sądy wartościujące, odnoszące się do „jakości” (*merit*) oraz „zalet” (*worth*). Różnicę między nimi wyartykułował Scriven<sup>30</sup>, odnosząc jakość do wartości, jaką wsobnie posiada dane ewaluandum, natomiast zalety traktując jako wartości relatywne z uwagi na odniesienie ewaluandum do kontekstu danej kultury, potrzeb, kosztów czy innych okoliczności. Program, który prezentuje wysoką jakość, może nie posiadać istotnych zalet z punktu widzenia potrzeb lub zaangażowanych kosztów (finansowych i niefinansowych). Tabela 1.2 syntetycznie obrazuje różnice obu pojęć<sup>31</sup>.

Również L. Korporowicz posługuje się zmodyfikowaną wersją definicji przyjętej przez amerykański Joint Committee on Standards for Educational Evaluation, przyjmując, iż ewaluacja to systematyczne badanie wartości lub cech konkretnego programu, działania lub obiektu z punktu widzenia przyjętych w tym celu kryteriów w celu jego usprawnienia, rozwoju lub lepszego rozumienia. Podkreśla on przy tym, że kategoria wartości jest rdzeniem pojęcia ewaluacji, która często bywa mylona z innymi kategoriami, jak „szacowanie”, „mierzenie”, „ocena”, czy bliskimi ewaluacji działaniami *quasi*-badawczymi, jak monitorowanie, akredytacja czy audyt.

---

<sup>27</sup> K. Olejniczak, *Mechanizmy wykorzystania ewaluacji: studium ewaluacji średniookresowych INTERREG III*, op. cit., s. 20.

<sup>28</sup> P. Dahler-Larsen, *The Changing Role of Evaluation in a Changing Society*, [w:] *The Future of Evaluation*, red. R. Stockmann i W. Meyer, Palgrave Macmillan UK, London 2016, s. 25.

<sup>29</sup> Joint Committee on Standards for Educational Evaluation, red., *The program evaluation standards: how to assess evaluations of educational programs*, Sage Publications, Thousand Oaks 1994, 2. wyd., s. 3; Zmodyfikowane tłumaczenie za: L. Korporowicz, *Badania ewaluacyjne*, op. cit., s. 66. L. Korporowicz tłumaczy termin „merit” jako wartość, tu przyjęto częściej stosowane tłumaczenie jako „jakość”.

<sup>30</sup> M. Scriven, *Evaluation thesaurus*, Sage Publications, Newbury Park, CA 1991, 4. wyd.

<sup>31</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 9.

Tabela 1.2

## Charakterystyka jakości oraz zalet ewaluandum

Jakość	Zaleta
Można ocenić dla dowolnego przedmiotu.	Oceniana tylko dla przedmiotów, które wykazują akceptowalny poziom jakości.
Stanowi wsobną wartość przedmiotu.	Stanowi relacyjną wartość przedmiotu.
Związana z przyjętymi standardami jakości dla danego rodzaju przedmiotu.	Związana z przyjętymi standardami jakości oraz danymi o odnośnych potrzebach.
Dotyczy odniesienia jakości przedmiotu do standardów oraz konkurencyjnych przedmiotów tego samego rodzaju.	Wiąże się z oceną jakości przedmiotu oraz znaczenia i wartości dla określonej grupy konsumentów.
Może być oceniona przez porównanie przedmiotu ze standardami lub przedmiotami konkurencyjnymi.	Ocena może mieć charakter porównawczy lub nie porównawczy.

Źródło: opracowanie na podstawie D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 9.

Wyjściową definicję z 1994 roku Stufflebeam i in. uzupełniają o dodatkowych pięć wartości, które mogą mieć zastosowanie w zależności od kontekstu ewaluacji<sup>32</sup>:

- uczciwość (*probity*) – wymaga sprawdzenia podczas ewaluacji zachowywania przez program standardów moralnych i dotyczy ona również metaewaluacji;
- wykonalność – stosuje się do wszystkich programów niezależnie od pozostałych wartości, program powinien spełniać to kryterium, a ewaluacja może wskazać istotne modyfikacje lub rekomendować jego zakończenie;
- bezpieczeństwo – stosuje się do programów, produktów i usług we wszystkich dziedzinach, gdzie może powstać zagrożenie dla konsumentów;
- istotność (*significance*) – szczególnie ważne dla programów o dalekosiężnych konsekwencjach w celu określenia adekwatnego planu upowszechnienia wyników ewaluacji;
- równość (*equity*) – dotyczy równości szans, która ma szczególne znaczenie w przypadku usług publicznych świadczonych dla grup ludzi.

W literaturze przedmiotu podkreśla się kilka aspektów dotyczących pojęcia ewaluacji. Ma ona charakter analityczny i usystematyzowany. Ocena wartości programów publicznych wymaga zrozumienia i wyjaśnienia mechanizmów społecznych, dlatego ewaluacja wykorzystuje bogatą paletę podejść i metod stosowanych w naukach społecznych<sup>33</sup>, lecz również sama wymaga systematycznej

<sup>32</sup> Ibidem, s. 11–14.

<sup>33</sup> E. Stern et al., *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*, „DFID: Department for International Development” 2012.



refleksji. Ponadto, ewaluacja wykracza poza badania naukowe, gdyż zawiera istotny element wartościujący jakość i zalety danej interwencji, opierający się na przyjętych kryteriach i standardach<sup>34</sup>, które bywają odmiennie definiowane i interpretowane w poszczególnych badaniach.

OECD proponuje rozbudowaną definicję ewaluacji<sup>35</sup>. Zawiera ona cechy ewaluacji (usystematyzowanie, obiektywizm i wiarygodność), jak i jej cele (oszacowanie jakości, zalet, uczenie się), odniesienie do zróżnicowania przedmiotów ewaluacji (programy lub projekty, procesy, efekty) oraz do kryteriów oceny i sposobów wykorzystania wniosków.

Bank Światowy definiuje ewaluację jako: „proces określania jakości lub znaczenia działań rozwojowych, polityki lub programu, w celu określenia istotności celów, skuteczności projektu i jego wdrożenia, efektywności lub wykorzystania zasobów oraz trwałości rezultatów. Ewaluacja powinna (umożliwiać) włączenie wniosków do procesu podejmowania decyzji zarówno partnera, jak i zleceniodawcy”<sup>36</sup>. Ta definicja, podobnie jak OECD, wyraźnie podkreśla funkcję rozliczeniowo-podsumowującą oraz funkcję uczenia się z ewaluacji<sup>37</sup>.

W dokumentach Unii Europejskiej ewaluacja jest definiowana krótko i lakonicznie jako „ocena wartości interwencji publicznej w odniesieniu do konkretnych kryteriów i standardów”<sup>38</sup>, z uwzględnieniem oceny działań publicznych na podstawie ich rezultatów, wpływu i adekwatności w stosunku do potrzeb, które mają zaspokoić<sup>39</sup>. Zmodyfikowaną wersję tej definicji proponują Olejniczak i in.: „ewaluacja to systematyczne badanie, prowadzone z użyciem zróżnicowanych metod, złożone ze zbierania danych, analizy, oceny oraz informowania o wynikach. Jego celem jest oszacowanie (w odniesieniu do jasno sformułowa-

---

<sup>34</sup> Joint Committee on Standards for Educational Evaluation, *The program evaluation standards*, op. cit.; *Standardy ewaluacji*, Polskie Towarzystwo Ewaluacyjne, Warszawa 2008; R.L. Miller, *Developing Standards for Empirical Examinations of Evaluation Theory*, „American Journal of Evaluation” 2010, t. 31, nr 3, s. 390–399.

<sup>35</sup> OECD, *Glossary of Key Terms in Evaluation and Results-Based Management*, OECD-DAC, Paris 2002.

<sup>36</sup> Strona Independent Evaluation Group Banku Światowego, <http://ieg.worldbankgroup.org/what-monitoring-and-evaluation>.

<sup>37</sup> K. Olejniczak, M. Kozak i B. Lendzion, red., *Teoria i praktyka ewaluacji interwencji publicznych*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008, s. 21; S. Borrás i S. Højlund, *Evaluation and policy learning: The learners' perspective: Evaluation and policy learning*, „European Journal of Political Research” 2015, t. 54, nr 1, s. 99–120.

<sup>38</sup> European Commission, *Evaluation design and management*, Office for Official Publications of the European Communities, Luxembourg 1999.

<sup>39</sup> European Commission, *Evaluation matters: The Evaluation Policy for European Union Development Co-operation*, Publications Office of the European Union, Luxembourg 2014; Zmiany w sposobie rozumienia ewaluacji w kolejnych dokumentach Komisji Europejskiej omawia: K. Olejniczak, M. Kozak i B. Lendzion, red., *Teoria i praktyka ewaluacji interwencji publicznych*, op. cit., s. 21 i nast.

nych kryteriów) jakości i wartości procesu oraz efektów wdrażania interwencji publicznych<sup>40</sup>.

Wielość definicji ewaluacji wynika m.in. ze zróżnicowanych podstaw teoretycznych ewaluacji<sup>41</sup>. Jedną z nich jest pozytywistyczne założenie, że dzięki systematycznej, ustrukturyzowanej i powtarzalnej obserwacji naukowej możliwe jest odkrycie wiedzy obiektywnej. Opozycyjną perspektywą jest konstruktywizm społeczny, odrzucający założenie o istnieniu jednej obiektywnej rzeczywistości społecznej i przyjmujący tezę o współlistnieniu wielości rzeczywistości, które są wyznaczone przez perspektywę obserwatora i aktora oraz przyjętych przez nich interpretacji. Rzeczywistość społeczną traktuje się jako konstrukt aktorów zaangażowanych w dany proces. Dla zrozumienia sytuacji zasadnicze znaczenie ma więc uchwycenie i zinterpretowanie kontekstu i procesów postrzeganych z różnych perspektyw poszczególnych aktorów<sup>42</sup>. Paradygmaty teoretyczne ewaluacji mają konsekwencje dla przyjmowanych podejść do ewaluacji, doboru metod, a także wyodrębnienia typów ewaluacji i określenia jej funkcji. W bardziej praktycznie zorientowanych podejściach często występuje pragmatycznie uzasadniany eklektyzm elementów dobranych z odmiennych czy przeciwstawnych tradycji badawczych w ewaluacji.

Istnieje szereg kategorii i praktyk zbliżonych do ewaluacji, jednak z uwagi na ich odrębność warto je jasno odróżniać, gdyż wymagają one innych umiejętności, innego zaplanowania w czasie oraz sposobu organizacji<sup>43</sup>. Takimi kategoriami są m.in. audyt i monitoring, które od ewaluacji różni perspektywa spojrzenia na polityki publiczne. Audyt weryfikuje legalność działań oraz regularność wykorzystania zasobów. Monitoring sprawdza sposób zarządzania interwencją oraz generuje regularne analizy dotyczące postępu w uzyskiwaniu efektów. Ewaluacja natomiast ocenia wdrożenie programu na podstawie efektów, wyników oraz jego wpływu społecznego<sup>44</sup>. Te kategorie różnią także stosowane kryteria. Audyt opiera się na standardowych kryteriach, jak budżet, regulacje prawne czy standardy zawodowe. Monitoring odnosi się do kryteriów oceny celów operacyjnych, które mają zostać osiągnięte. Ewaluacja natomiast rozpoczyna się od ustalenia własnych kryteriów, które formułuje się na podstawie celów, jakie ma dana polityka publiczna. W przeciwieństwie do audytu i monitoringu ma ona istotny komponent teoretyczny oraz badawczy. Obrazowo to zróżnicowanie przedstawia rysunek 1.1. oraz zestawienie w tabeli 1.3.

<sup>40</sup> K. Olejniczak, M. Kozak i B. Lenzion, red., *Teoria i praktyka ewaluacji interwencji publicznych*, op. cit., s. 22.

<sup>41</sup> K. Olejniczak, *Mechanizmy wykorzystania ewaluacji: studium ewaluacji średniookresowych INTERREG III*, op. cit., s. 21.

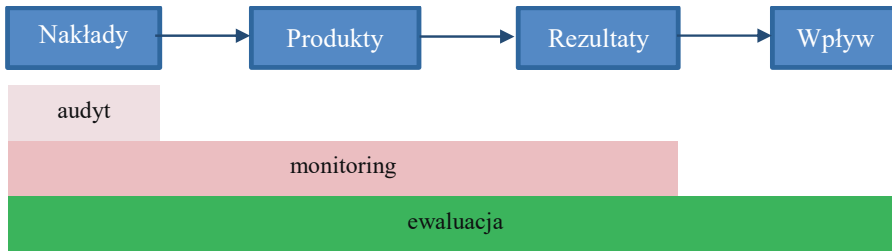
<sup>42</sup> B. Kotarba, *Znaczenie kontekstu sytuacyjnego w lokalnej polityce publicznej*, „Wrocławskie Studia Politologiczne” 2016, t. 21, s. 119–130.

<sup>43</sup> European Commission, *Evaluation design and management*, op. cit.

<sup>44</sup> G.E. Derrick i G.N. Samuel, *The Evaluation Scale: Exploring Decisions About Societal Impact in Peer Review Panels*, „Minerva” 2016, t. 54, nr 1, s. 75–97.

Rysunek 1.1

## Ewaluacja a działania pokrewne: audyt i monitoring



Źródło: opracowanie na podstawie: European Commission, *Evaluation design and management*, Office for Official Publications of the European Communities, Luxembourg 1999.

Tabela 1.3

## Różnice między ewaluacją, badaniem naukowym, monitoringiem i audytem

	Cel	Zakres	Metody
Ewaluacja	Oszacowanie rzeczywistych efektów programu. Ocena zasadności programu w danej formie.	Szeroki i/lub wąski. Szersze spojrzenie na program oraz sposób jego prowadzenia. Zidentyfikowanie i ocena: – celów danego przedsięwzięcia (w tym założonych relacji przyczynowo-skutkowych); – sposobów realizacji (ewentualnych słabości i pól potencjalnych usprawnień); – długookresowych zamierzonych i niezamierzonych efektów, wybiegających poza czas i miejsce wdrażania programu.	Zwykle multidyscyplinarne (ekonomia, socjologia itp.). Nie zawsze oparte na ścisłej metodologii naukowej.
Badanie naukowe	Poszerzenie wiedzy teoretycznej. Identyfikacja i weryfikacja teorii wyjaśniających działanie programu.	Szeroki. Szerokie spojrzenie na program jako element większego systemu.	Zakorzenione zwykle w jednej dyscyplinie. Ścisłe przestrzegające metodologii badań naukowych.
Monitoring	Zweryfikowanie, czy program postępuje zgodnie z planem.	Wąski. Rejestrowanie postępów z prowadzonych operacji. Bieżące weryfikowanie tempa i kierunku, w którym zmierza program.	Wąski. Rejestrowanie postępów z prowadzonych operacji. Bieżące weryfikowanie tempa i kierunku, w którym zmierza program.

Tabela 1.3 – cd.

	Cel	Zakres	Metody
Audyt	Sprawdzenie, czy działania wykonywane są zgodnie z procedurami.	Wąski. Poprawność procedur. Poprawność rozliczeń. Zgodność z zapisami prawnymi i zasadami kontraktu.	Prawnie określone zasady, procedury i metody prowadzenia kontroli.

Źródło: opracowanie na podstawie: K. Olejniczak, *Mechanizmy wykorzystania ewaluacji: studium ewaluacji średniookresowych INTERREG III*, op. cit., Warszawa 2008.

Kryteria ewaluacji zalicza się na ogół do jednej z pięciu kategorii:

- trafność (*relevance*) – dotyczy zgodności między celami przedsięwzięcia a potrzebami interesariuszy;
- efektywność (*efficiency*) – określa stosunek uzyskanej wartości, korzyści do nakładów;
- skuteczność (*effectiveness*) – określa stopień osiągnięcia założonych celów;
- użyteczność (*utility*) – określa wieloaspektowe (społeczne, gospodarcze, środowiskowe itp.) konsekwencje interwencji;
- trwałość (*sustainability*) – określa wpływ na tworzenie warunków zrównoważonego rozwoju.

Związek z elementami programu czy interwencji publicznej przedstawia rysunek 1.2.

Ocena wpływu (*impact assessment*) ma zbliżone znaczenie do ewaluacji pod względem celu i samego procesu. Ocena wpływu koncentruje się na wyniku polityki, łącznie z jej długookresowym wpływem. Coraz częściej zawęża się jej rozumienie do ilościowej determinacji – najczęściej z wykorzystaniem analizy kontrfaktycznej – ilości i rodzajów rezultatów, wyników i wpływu<sup>45</sup>. Ewaluacja natomiast jest bardziej całościowym działaniem, które obejmuje również ocenę celów interwencji oraz procesu prowadzącego do uzyskania danych rezultatów. Sednem ewaluacji jest określenie zależności przyczynowych, które łączą wpływ na wyniki i efekty (np. wzrost gospodarczy, poprawa stanu zdrowia czy środowiska) z nakładami (np. inwestycje w B+R)<sup>46</sup>. W tym celu ewaluacje

<sup>45</sup> E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies i B. Befani, *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*, „DFID: Department for International Development” 2012, s. 24.

<sup>46</sup> E. Gates i L. Dyson, *Implications of the Changing Conversation About Causality for Evaluators*, „American Journal of Evaluation” 2017, t. 38, nr 1, s. 29–46; B. Bozeman i A.N. Link, *Toward an assessment of impacts from US technology and innovation policies*, „Science and Public Policy” 2015, t. 42, nr 3, s. 369–376; J. Górniak, *Efekt przyczynowy w ewaluacji*, [w:] *Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych*, red. A. Haber i M. Szałaj, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2009, s. 189–200.

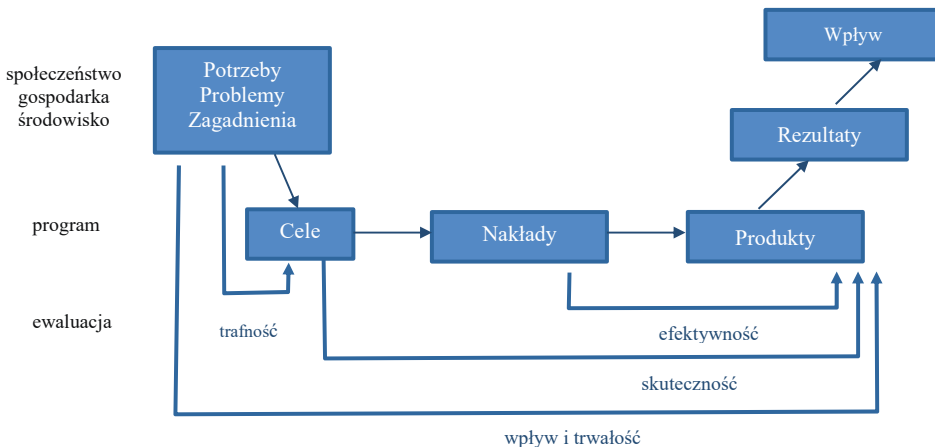
wykorzystują teorie zmiany, aby wskazać różne możliwe drogi, prowadzące do uzyskania efektów<sup>47</sup>.

Przedmiot ewaluacji najczęściej bywa określany jako ewaluand (z łac. *evaluand* – to, co ma być ewaluowane), natomiast w przypadku, gdy obiektem ewaluacji są osoby, często używa się określenia ewaluowana/y (*valuee*)<sup>48</sup>. Przedmioty ewaluacji obejmują bardzo szerokie spektrum:

- programy,
- projekty,
- polityki,
- wnioski grantowe,
- produkty,
- wyposażenie i sprzęt,
- usługi,
- pojęcia, teorie i dane oraz inne rodzaje informacji,
- jednostki,
- organizacje itd.

Rysunek 1.2

Główne kryteria ewaluacji z odniesieniem do faz programu



Źródło: opracowanie własne na podstawie: European Commission, *Evaluation design and management*, op.cit.

<sup>47</sup> OECD, *Evaluation and impact assessment of STI policies*, [w:] *OECD Science, Technology and Innovation Outlook 2016*, OECD, OECD Publishing 2016, [http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016/evaluation-and-impact-assessment-of-sti-policies\\_sti\\_in\\_outlook-2016-13-en](http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016/evaluation-and-impact-assessment-of-sti-policies_sti_in_outlook-2016-13-en).

<sup>48</sup> M. Scriven, *Evaluation thesaurus*, op. cit., s. 48.

Głównym przedmiotem ewaluacji są programy, choć ewaluacja stosuje się do szerokiego wachlarza dyscyplin, aktywności i przedsięwzięć: ewaluacja edukacyjna, usługi i produkty konsumenckie, zasoby ludzkie, planowanie urbanistyczne, polityka naukowa i innowacyjna, pomoc międzynarodowa czy stan środowiska.

Współczesny stan ewaluacji jako dyscypliny naukowej Scriven określa mianem badań transdyscyplinarnych<sup>49</sup>, podobnie jak to jest w przypadku statystyki, teorii pomiaru czy logiki. Jej zadaniem jest tworzenie i rozwijanie narzędzi dla dyscyplin, których przedmiotem badania jest rzeczywistość. Dla badań transdyscyplinarnych charakterystyczne jest po pierwsze to, że mają one zastosowanie nie tylko w dyscyplinach akademickich, ale również w innych dziedzinach życia. Po drugie, w ich ramach wyodrębniają się dwa zasadnicze elementy: „rdzeń akademicki”, odpowiadający „nauce czystej”, który obejmuje zagadnienia najogólniejsze, aspekty teoretyczne, metodologię, aparaturę pojęciową, oraz część aplikacyjną, która często przybiera specyficzną postać w zależności od dziedziny zastosowań<sup>50</sup>.

W literaturze przedmiotu funkcjonuje wiele typologii ewaluacji, wyodrębnionych z uwagi na różne założenia teoretyczne i metodologiczne oraz odmienne kryteria<sup>51</sup>. Olejniczak i in. wyodrębniają następujące wymiary kategoryzowania ewaluacji: ze względu na przedmiot ewaluacji, jej ogólny cel, moment podjęcia badania oraz jego administracyjną organizację, a także zastosowane podejście badawcze.

Ze względu na przedmiot ewaluacji można wyróżnić ewaluację projektu jako badanie pojedynczej interwencji – często nowego, innowacyjnego projektu, które może być elementem większej całości (programu) i zostać wykorzystane jako studium przypadku lub stanowić badanie pilotażowe w celu podjęcia decyzji o kontynuowaniu interwencji w większej skali. Z kolei, ewaluacja programu obejmuje pojedynczy program sektora publicznego lub cały zestaw publicznych interwencji. Ewaluacja polityki, nazywana analizą polityk (*policy analysis*), składa się z szeregu badań i analiz, które zazwyczaj mają wieloletni charakter. Ewaluacja tematyczna/horyzontalna dotyczy zarządzania lub aspektu programu w ujęciu przekrojowym kilkunastu programów. Metaewaluacja (*metaevaluation*) to badanie, którego przedmiotem są inne ewaluacje. Na ogół ma charakter porównawczy i podsumowuje zbiór ewaluacji (dotyczących podobnych programów) lub ocenia poprawność metodologiczną.

Ze względu na ogólny cel ewaluacji Scriven wyróżnił dwa podstawowe rodzaje<sup>52</sup>: ewaluacja formatywna (*formative*) oraz sumatywna. Celem tej pierwszej jest ocena projektu programu, jego planu oraz procesu wdrażania interwencji,

<sup>49</sup> M. Scriven, *Evaluation as a discipline*, op. cit.

<sup>50</sup> C.L. Coryn i J.A. Hattie, *The transdisciplinary model of evaluation*, „Journal of Multi-Disciplinary Evaluation” 2007, t. 3, nr 4, s. 107–114.

<sup>51</sup> K. Olejniczak, M. Kozak i B. Lendzion, red., *Teoria i praktyka ewaluacji interwencji publicznych*, op. cit., s. 28n; E.O. Okoń-Horodyńska, *Smart specialisation as a lever towards innovation: the example of Malopolska region in Poland*, „International Journal of Transitions and Innovation Systems” 2014, t. 3, nr 3, s. 249–269.

<sup>52</sup> M. Scriven, *The Methodology of Evaluation*, op. cit.

dokonana przed jej rozpoczęciem lub przeprowadzana w czasie jej trwania. Pozwala ona osobom zarządzającym programem lepiej ocenić adekwatność celów względem potrzeb społecznych oraz zwiększyć skuteczność i efektywność wdrażania programu. Na ogół ma charakter wewnętrzny. Ewaluacja sumaryczna (*summative*) ocenia rezultaty programu po jego zakończeniu i może być wykorzystana do decyzji dotyczących dalszych kolei programu (modyfikacja, zatrzymanie, itp.). Ma ona charakter opisowy i wyjaśniający, a jej adresatem są interesariusze danego programu.

Ze względu na moment podjęcia badania zasadniczo wyróżnia się: ewaluacje wykonywane przed rozpoczęciem badanej interwencji (*ex ante*), w jej trakcie (*mid-term*) oraz badania po zakończeniu programu będącego przedmiotem analizy i oceny (*ex post*). Celem pierwszej z nich jest optymalizacja alokacji zasobów przez zweryfikowanie struktury, logiki i uzasadnienia programu. Ewaluacja średniookresowa pozwala na ocenę procesu zarządzania, założeń strategii i postępów w osiąganiu rezultatów. Ewaluacja końcowa jest wykonywana po zakończeniu interwencji w celu oceny rezultatów i długookresowych wpływów programu oraz trwałości zmian i wskazania efektów ubocznych (pozytywnych i negatywnych). Niekiedy stosuje się także ewaluację ciągłą (*on going*), która jest wykonywana w dowolnym momencie implementacji programu i ma stanowić wsparcie dla administrowania programem.

Ze względu na administracyjną stronę badania wyróżnia się ewaluację wewnętrzną (prowadzona przez osoby z danej instytucji odpowiedzialnej za wdrażanie analizowanej interwencji, często odrębny dział) oraz zewnętrzną (zlecona poza instytucję prowadzącą dany program, wykonywana przez zespół niezależnych ekspertów).

## 1.2. Społeczne role ewaluacji

Za zasadniczy cel ewaluacji uznaje się podniesienie jakości życia społecznego przez wspomaganie instytucji demokratycznych w doborze, nadzorowaniu, wartościowaniu i doskonaleniu programów i polityk publicznych w celu podniesienia ich jakości i skuteczności oraz uczenia się na ich podstawie przy tworzeniu kolejnych programów<sup>53</sup>. D. Stufflebeam i in. uznają ewaluację wręcz za „najbardziej fundamentalną dyscyplinę dla społeczeństwa”, „istotną charakterystykę ludzkiej kondycji” oraz za jeden z najbardziej wysublimowanych procesów poznawczych, do jakich ludzie są zdolni<sup>54</sup>. Służy ona społeczeństwu, potwierdzając jakość, wartość, postęp, akredytację oraz rozliczalność, dając tym samym podstawę do decyzji o rozwijaniu bądź zakończeniu danego programu.

<sup>53</sup> M.M. Mark, G.T. Henry i G. Julnes, *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs*, Jossey-Bass, San Francisco 2000, s. 3.

<sup>54</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 3.

Zasadnicze trzy powody ewaluacji, jakie można wskazać w odniesieniu do polityk publicznych, to:

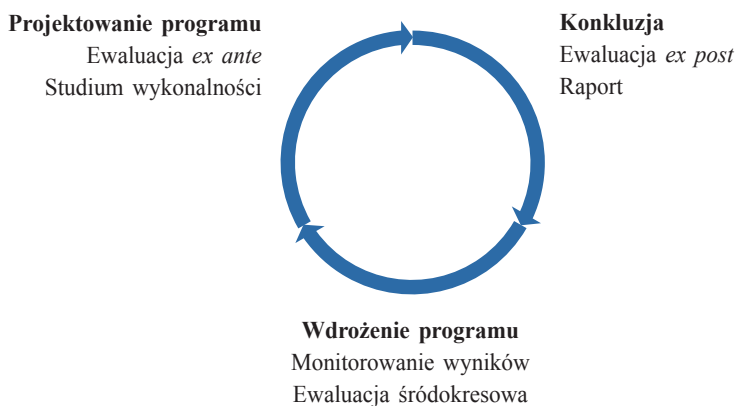
- ich legitymizacja i weryfikacja;
- poprawa jakości interwencji;
- rozliczalność<sup>55</sup>.

Celem publicznych interwencji jest zaspokojenie potrzeb lub rozwiązanie problemów społeczno-ekonomicznych. Ewaluacja może służyć wyartykułowaniu tych potrzeb lub weryfikacji ich rzeczywistego istnienia. Dla realizowanych interwencji dostarcza informacji zwrotnych, co może wskazać sposoby ulepszenia wdrożeń interwencji (pojedyncza pętla uczenia się) lub sposób strategicznej reorientacji interwencji (podwójna pętla uczenia się).

Ewaluacja określa także dla decydentów politycznych oraz obywateli osiągnięte rezultaty w zestawieniu z alokowanymi środkami, co pozwala ustalić rozliczalność poszczególnych programów, projektów czy interwencji publicznych.

**Rysunek 1.3**

**Ewaluacja w cyklu życia programu**



Źródło: opracowanie na podstawie: European Commission, *Evaluation design and management*, op. cit.

Ewaluacja posiada liczne cele<sup>56</sup>:

- ocena jakości i wartości polityki lub programu publicznego;
- usprawnienie projektowania programu i jego realizacji;
- uczenie się i poprawianie funkcjonowania programu;
- generowanie wiedzy, która będzie lepszą podstawą projektowania polityki;
- wsparcie dla funkcji kontrolnej sprawowanej przez demokratyczne instytucje;
- rozliczenie z wykorzystania zasobów i osiągnięcia oczekiwanych wyników;

<sup>55</sup> European Commission, *Evaluation design and management*, op. cit.

<sup>56</sup> R. Szarfenberg, *Zagadnienia terminologiczne*, op. cit., s. 29–30.



- informowanie opinii publicznej o działaniach władz publicznych poprzez upowszechnianie wyników ewaluacji;
  - pomaganie agencjom publicznym w zwiększaniu ich zdolności planistycznych, realizacyjnych i analitycznych.
- Syntetyczne zestawienie normatywnych i pozytywnych funkcji ewaluacji<sup>57</sup>:
- formatywna – usprawniająca realizowany proces, służąca poprawianiu, doskonaleniu;
  - sumatywna – podsumowująca osiągnięte rezultaty, związana z selekcją, certyfikacją, rozliczaniem;
  - psychologiczna lub społeczno-polityczna – związana ze wzbudzeniem motywacji do działania i świadomości;
  - administracyjna – związana ze sprawowaniem władzy i zarządzaniem politykami publicznymi;
  - wspomagająca proces podejmowania decyzji związanych z zarządzaniem programami publicznymi;
  - budowanie partnerstwa i współwłasności;
  - funkcja poznawcza;
  - rozliczanie i wzmocnienie odpowiedzialności;
  - poprawa planowania przez zagwarantowanie racjonalnego uzasadnienia dla danej interwencji;
  - poprawa wdrażania i kontrola jakości m.in. istniejącej struktury, procedur i procesu zarządzania interwencją publiczną;
  - wzmocnienie odpowiedzialności przez zwiększenie skuteczności i wydajności działań publicznych;
  - wspieranie procesów uczenia się m.in. przez budowanie zdolności ewaluacyjnych, weryfikowanie teorii leżących u podstaw danych, polityk czy programów;
  - wzmacnianie partnerstwa i poczucia współwłasności przez realizowanie zasady uprawnienia (*empowerment*), aktywizowanie i wzmacnianie powiązań, relacji, współpracy, interakcji między uczestnikami określonego przedsięwzięcia.
- Obok nich wskazuje się także dysfunkcje czy nieuprawnione funkcje ewaluacji<sup>58</sup>:
- odwlekanie ważnych decyzji;
  - uzasadnianie decyzji, które zostały już podjęte;
  - unikanie lub rozpraszanie odpowiedzialności za podjęte decyzje;
  - poprawę wizerunku osoby lub instytucji zamawiającej badania (ewaluowanie tylko wybranych aspektów programu, żeby ukazać go w lepszym świetle, unikanie ewaluacji pewnych aspektów programów, aby ukryć jego porażkę).

---

<sup>57</sup> Ibidem, s. 30–31.

<sup>58</sup> Ibidem, s. 31.

Zasadniczą rolą ewaluacji w odniesieniu do polityki naukowej i innowacyjnej jest wsparcie rozliczalności wydatków ze środków publicznych oraz intensyfikacja procesu uczenia się instytucji publicznych<sup>59</sup>. Istotnym, choć może mniej zauważalnym rezultatem jest także wpływ na sposób realizacji samego procesu poprzez efekty *ex ante* na sposób działania menadżerów i decydentów, którzy mają świadomość tego, że realizowany przez nich program będzie poddany ewaluacji. Może ona także wpływać na sam produkt przez stymulowanie interakcji pomiędzy poszczególnymi agentami w fazie planowania programu, jego wdrażania lub rekomendacji po zakończeniu. Ewaluacja może również oddziaływać strategicznie w perspektywie długookresowej, wzmacniając wiarygodność i legitymizując określone interwencje w nauce, technologii i innowacjach, co może się okazać kluczowe np. podczas negocjacji międzyresortowych. Nadal jednak jedną ze słabości cyklu polityk publicznych jest ograniczone wykorzystanie wyników ewaluacji<sup>60</sup>.

OECD w raporcie *Governance and implementation of innovation policies* z 2015 roku<sup>61</sup> podkreśla kluczową rolę ewaluacji w projektowaniu i zarządzaniu polityką innowacyjną, co umożliwi decydentom politycznym i analitykom:

- ocenę wkładu, jaki innowacje mają w osiągnięciu celów społecznych i ekonomicznych;
- określenie determinant i przeszkód innowacyjności, co ma zasadnicze znaczenie dla projektowania skutecznych polityk innowacyjnych;
- określenie wpływu całego portfolio projektów i programów na osiągnięcie efektów (np. usuwanie porażek rynkowych, ograniczających dostępność finansów, umiejętności czy technologii);
- określenie efektywności różnych podejść w polityce, co wpływa na racjonalizację alokacji środków;
- ciągle uczenie się i poprawianie projektowania programów i ich administracji;
- inicjowanie debaty publicznej – z udziałem różnorodnych grup interesariuszy – nad adekwatnością i realizacją polityk, programów oraz instytucji, których podstawą są wyniki ewaluacji;
- zwiększenie rozliczalności polityk publicznych.

Istotnym elementem ewaluacji jest „kontrafaktyczna” ocena, na ile pożądane zmiany zaszłyby bez wdrożenia określonej interwencji. Często popełnia się w tej ocenie dwa rodzaje błędów. Pierwszy określany jest mianem „błędem narzucania projektu” (*project fallacy*), polega na przypisywaniu ewaluowanej interwencji wyłącznie lub w całości efektów, które w rzeczywistości mają charakter

---

<sup>59</sup> E. Okoń-Horodyńska, *Narodowy system innowacji w Polsce*, Wydawnictwo Uczelniane Akademii Ekonomicznej, Katowice 1998; OECD, *Evaluation and impact assessment of STI policies*, op. cit.

<sup>60</sup> OECD, *Trends in evaluation and assessment*, [w:] *Synergies for Better Learning*, OECD, OECD Publishing 2013, s. 35–56, [http://www.oecd-ilibrary.org/education/synergies-for-better-learning/trends-in-evaluation-and-assessment\\_9789264190658-5-en](http://www.oecd-ilibrary.org/education/synergies-for-better-learning/trends-in-evaluation-and-assessment_9789264190658-5-en).

<sup>61</sup> OECD, *Governance and implementation of innovation policies*, [w:] *The Innovation Imperative*, OECD Publishing 2015, s. 237–267.

kumulatywny i są rezultatem interakcji wielu czynników. Drugi błąd polega na niedoszacowaniu efektów interwencji z uwagi na wąski zakres prowadzonej ewaluacji lub zbyt krótki jej przedział czasowy.

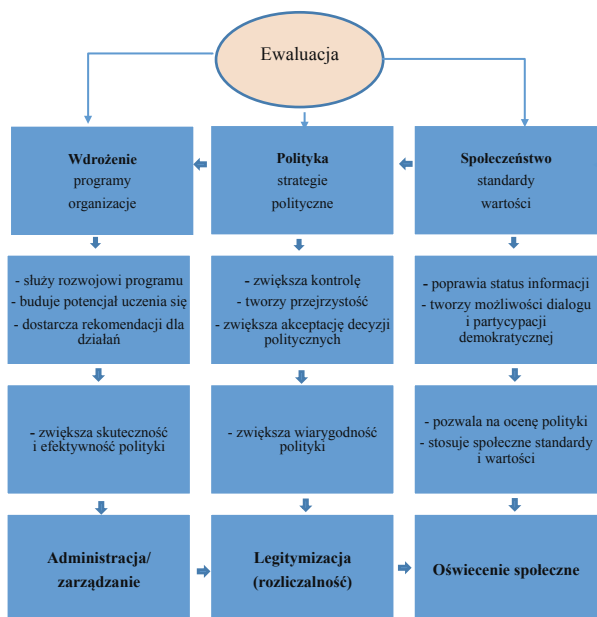
Dwa istotne trendy w rozwijaniu precyzyjnych technik ewaluacji, co w znacznym stopniu ułatwia dynamicznie rozwijające się wspomaganie komputerowe<sup>62</sup>, podnoszą skuteczność wpływu jej wniosków na polityki publiczne:

- identyfikacja zależności kontrfaktycznych (określenie, co by się wydarzyło przy braku danej interwencji);
- wnioskowania przyczynowe (zamiast zwykłych korelacji)<sup>63</sup>.

W podejściu funkcjonalnym ewaluacja odgrywa potrójną rolę społeczną: jako instrument zarządzania od planowania po wdrożenie programu, jako instrument osiągania celów społecznych za pomocą polityk publicznych oraz jako instrument refleksji i krytyki współczesnego społeczeństwa<sup>64</sup>. Rysunek 1.4 przedstawia te trzy funkcje społeczne ewaluacji<sup>65</sup>.

Rysunek 1.4

### Funkcje społeczne ewaluacji



Źródło: opracowanie na podstawie: R. Stockmann, *The Future of Evaluation in Modern Societies*, [w:] *The Future of Evaluation*, red. R. Stockmann i W. Meyer, Palgrave Macmillan UK, London 2016, s. 40.

<sup>62</sup> A. Surdej, red., *Ewaluacja w polityce publicznej*, Difin, Warszawa 2013, s. 11.

<sup>63</sup> J. Górniak, *Efekt przyczynowy w ewaluacji*, op. cit.

<sup>64</sup> R. Stockmann, *The Future of Evaluation in Modern Societies*, [w:] *The Future of Evaluation*, red. R. Stockmann i W. Meyer, Palgrave Macmillan UK, London 2016, s. 39–41.

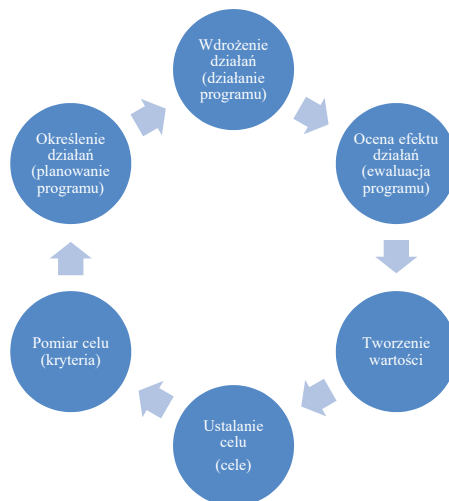
<sup>65</sup> M. Parlett i D. Hamilton, *Evaluation as Illumination: A New Approach to the Study of Innovative Programmes*, op. cit.

### 1.3. Aspekty aksjologiczne ewaluacji

Dominacja stanowiska akognitywizmu<sup>66</sup> w odniesieniu do sądów wartościujących w filozofii nauki w pierwszej połowie XX wieku zasadniczo ograniczała możliwość rozwoju ewaluacji jako odrębnej dyscypliny badawczej<sup>67</sup>. Wprost pojawia się ona w pracach głównego nurtu dopiero wtedy, gdy akognitywizm poddano zasadniczej krytyce, podważając założenie, że wartościowanie pociąga subiektywizm czy arbitralność<sup>68</sup>. E. Suchman<sup>69</sup>, który był zwolennikiem pozytywistycznego podejścia do ewaluacji, za warunek konieczny badań ewaluacyjnych uznał szacowanie wartości. Wartość definiował jako „dowolny aspekt sytuacji, zdarzenia lub przedmiotu, który jest przedmiotem preferencji jako dobry, zły, pożądany, niepożądany itp.”<sup>70</sup>. Cały proces ewaluacji, rozumiany tu jeszcze w sensie Tylera, rozpoczyna się i powraca do kształtowania wartości, jak pokazuje rysunek 1.5.

Rysunek 1.5

Proces ewaluacji Suchmana



Źródło: opracowanie własne na podstawie E.A. Suchman, *Evaluative research: principles and practice in public service and social action programs*, Russell Sage Foundation, New York, NY 1967, s. 34.

<sup>66</sup> P. Kawalec, *Akognitywizm*, [w:] *Powszechna encyklopedia filozofii*, red. A. Maryniarczyk, Polskie Towarzystwo Tomasza z Akwinu, Lublin 2000, t. I, s. 122–128.

<sup>67</sup> M. Scriven, *Evaluation as a discipline*, op.cit., s. 149. Podobną rolę Korporowicz przypisuje konstruzywizmowi, który negując dziedzictwo logicznego empiryzmu, radykalnie podważał założenie o istnieniu obiektywnej rzeczywistości, stojące u podstaw większości dyscyplin naukowych, a przynajmniej przyrodniczych.

<sup>68</sup> M. Scriven, red., *Hard-Won Lessons in Program Evaluation*, „New Directions for Program Evaluation” 1993, t. 58, s. 1–107; M. Scriven, *The nature of evaluation*, „New Directions for Program Evaluation” 1993, t. 58, s. 8 i nast.

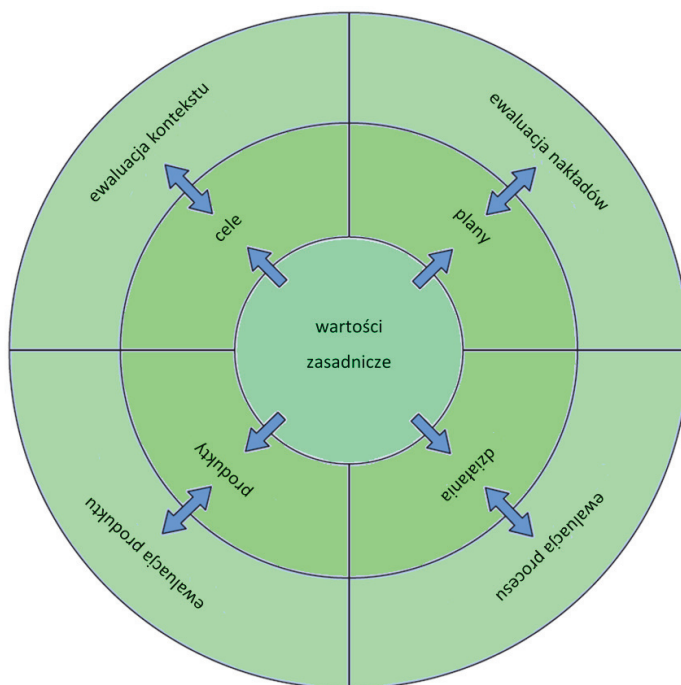
<sup>69</sup> E.A. Suchman, *Evaluative research: principles and practice in public service and social action programs*, Russell Sage Foundation, New York, NY 1967.

<sup>70</sup> Ibidem, s. 67.

Sama etymologia terminu „ewaluacja” wskazuje, że w istotny sposób angażuje ona wartościowanie<sup>71</sup> i nie ma ewaluacji, które byłyby wolne od wartościowania. W zależności od rodzaju ewaluandum wartości te mogą obejmować: efektywność, skuteczność, użyteczność, kosztowność, bezpieczeństwo, legalność itp. Ponadto, sam proces ewaluacji także jest ugruntowany w określonym zbiorze wartości, jakim są np. standardy ewaluacji czy zasady formułowane w przewodnikach dla ewaluatorów przez krajowe lub międzynarodowe stowarzyszenia<sup>72</sup>. Kluczową rolę wartości w ewaluacji przejrzysto ilustruje model CIPP Stufflebeama przedstawiony na rysunku 1.6.

Ryunek 1.6

### Rola wartości w modelu CIPP



Źródło: opracowanie na podstawie: D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 317.

Najmniejsze koło przedstawia wartości zasadnicze, które powinny zostać zdefiniowane i wykorzystane od początku danej ewaluacji. Koło środkowe przedstawia cztery podstawowe fazy ewaluacji, a zewnętrzne – odpowiadające im odmiany ewaluacji. Między odpowiednimi elementami obu ostatnich kół zacho-

<sup>71</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit.

<sup>72</sup> J.N. Hall, J. Ahn i J.C. Greene, *Values Engagement in Evaluation: Ideas, Illustrations, and Implications*, „American Journal of Evaluation” 2012, t. 33, nr 2, s. 195–207.

dzą interakcje zaznaczone dwukierunkowymi strzałkami: wyznaczanie celów wymaga zbadania kontekstu, co z kolei wpływa na uprawomocnienie lub poprawę celów, poprawa planowania kieruje uwagę w stronę ewaluacji nakładów, a to jest odniesieniem dla wzmocnienia planów. Podjęcie działań prowadzi ku ewaluacji procesu, która daje istotne informacje zwrotne dla poprawy wdrażania programu. Rezultaty lub ich brak, a także skutki uboczne zostają wskazane w ewaluacji produktu, która również pomaga w określeniu potrzeb i możliwości osiągnięcia lepszych rezultatów w przyszłości. Jak podkreślają Stufflebeam i in., te wzajemne interakcje umożliwiają osadzenie ewaluacji w zasadniczych wartościach, które są tu rozumiane jako ideały społeczne, grupowe lub jednostkowe. Ewaluacja wymaga, aby ewaluator i zlecający ewaluację określili wartości znajdujące się u podstaw ewaluacji (np. pomoc każdemu dziecku w rozwinięciu swojego potencjału edukacyjnego, zaspokojenie potrzeb dzieci niepełnosprawnych, rozwój potencjału gospodarczego kraju oraz wzrost konkurencyjności itd.). Te wartości z kolei są podstawą wyprowadzenia lub uprawomocnienia poszczególnych kryteriów ewaluacyjnych, a następnie odpowiedniego doboru i konstrukcji instrumentów oraz procedur ewaluacyjnych. Ponadto, określenie wartości stanowi punkt odniesienia przy wykrywaniu nieoczekiwanych mocnych lub słabych stron programu<sup>73</sup>.

#### 1.4. Główne etapy w rozwoju ewaluacji jako dyscypliny badawczej

Początki dużego zainteresowania ewaluacją upatruje się w rozpoczętej przed wybuchem I wojny światowej ewaluacji programów społecznych, zwłaszcza w obszarze edukacji i zdrowia publicznego, w których zaczęto wykorzystywać metody nauk społecznych<sup>74</sup>. Prace te zintensyfikowano w czasie II wojny światowej, gdy ewaluacja wykorzystywana była na potrzeby armii Stanów Zjednoczonych np. do monitorowania morale żołnierzy czy ewaluacji technik propagandowych.

Kolejną ważną fazą w rozwoju ewaluacji były lata 60. i 70., gdy szeroko zakrojonym programom społecznym, jak War on Poverty oraz the Great Society, towarzyszyło wprowadzenie metod badań eksperymentalnych i *quasi*-ekspery-

---

<sup>73</sup> Osobnym zagadnieniem jest etyka badań ewaluacyjnych, por. M. Morris, *Evaluation ethics for best practice: cases and commentaries*, Guilford Press, New York 2008; L. Goodyear, *Resources to Support Ethical Practice in Evaluation: An Interview With the Director of the National Center for Research and Professional Ethics*, „American Journal of Evaluation” 2012, t. 33, nr 3, s. 459–461; Salem Press, *Research & evaluation methods*, 2014, s. 90 i nast.; C. Barnett i L. Camfield, *Ethics in evaluation*, „Journal of Development Effectiveness” 2016, t. 8, nr 4, s. 528–534.

<sup>74</sup> P.H. Rossi, H.E. Freeman i M.W. Lipsey, *Evaluation: a systematic approach*, Sage Publications, Thousand Oaks, Calif 1999, 6th ed.; M.M. Mark, G.T. Henry i G. Julnes, *Evaluation*, op. cit.; C. Fox, R. Grimm i R. Caldeira, *An introduction to evaluation*, SAGE Publications, London 2016.

mentalnych<sup>75</sup>. Związany był z tym rozwój innych technik i planów badawczych w ewaluacji, które wykorzystywały zaawansowane techniki statystyczne. Stopniowo w latach 70. wyłoniła się wspólnota badawcza osób specjalizujących się w zagadnieniach ewaluacji, a w 1976 roku powstało pierwsze czasopismo poświęcone tej problematyce „Evaluation Review”. Kolejnym etapem był przełom tysiąclecia, gdy znacząca liczba organizacji rządowych i pozarządowych zamawiała ewaluacje, czemu towarzyszyło zwiększenie liczby ewaluatorów i powstawanie nowych towarzystw naukowych. W tej fazie, z uwagi na duże zróżnicowanie zadań, jakie stawiano ewaluatorom, znacząco rozwinęły się teorie ewaluacji oraz jej metody i narzędzia<sup>76</sup>. Punkt ciężkości przesunął się z perspektywy ewaluatorów na perspektywę konsumenta, co spowodowało, że ewaluacja przekroczyła granice dyscypliny czysto akademickiej: „Ewaluacja stała się działalnością polityczną i menadżerską, która istotnie wpływa na złożoną mozaikę, z której wyłaniają się decyzje polityczne oraz środki/źródła dla uruchamiania, poszerzania, zmieniania lub podtrzymywania programów polepszających warunki ludzkiego bytowania”<sup>77</sup>. Ewaluacja stała się więc integralnym składnikiem polityki społecznej oraz administracji publicznej<sup>78</sup>.

W wyodrębnianiu się ewaluacji jako odrębnej dyscypliny badawczej E. Vedung wskazuje cztery „fale”, które są powiązane z szerszym kontekstem zmian w postrzeganiu polityk publicznych i które – mimo następowania po sobie – pozostawiają pewien „osad”, a na nim rozgrywają się późniejsze fale<sup>79</sup>. Początek tych fal widzi on w połowie lat 60., gdy – głównie pod wpływem The RAND Corporation – dokonano zasadnicze przesunięcie z analizy *ex ante* możliwych konsekwencji interwencji publicznych w kierunku analizy *ex post* przyjętych lub nawet zakończonych interwencji. Pierwszą z nich, która dominowała w latach 1965–1975, określa on mianem radykalnego racjonalizmu, wynikającego z próby unaukowania polityk publicznych w duchu instrumentalnego racjonalizmu M. Webera. Głównym założeniem tego pozytywistycznego podejścia było jak najpełniejsze wykorzystanie wszystkich dostępnych środków naukowej analizy polityk publicznych w celu odpowiedzi na możliwie wyczerpujący zestaw pytań, co miało przesunąć na drugi plan lub całkowicie usunąć krótkoterminowe planowanie, kształtowane pod wpływem bieżących interesów politycznych i przypadkowo dobranej wiedzy. W tym ujęciu starannie zaplanowane interwencje,

<sup>75</sup> S.I. Donaldson i M.W. Lipsey, *Roles for Theory in Contemporary Evaluation Practice: Developing Practical Knowledge*, [w:] *The SAGE Handbook of Evaluation*, red. I. Shaw, J. Greene i M. Mark, SAGE Publications, London 2006, s. 57–75.

<sup>76</sup> P.H. Rossi, H.E. Freeman i M.W. Lipsey, *Evaluation*, op. cit.

<sup>77</sup> Ibidem, cyt. za: C. Fox, R. Grimm i R. Caldeira, *An introduction to evaluation*, op. cit., s. 16–17.

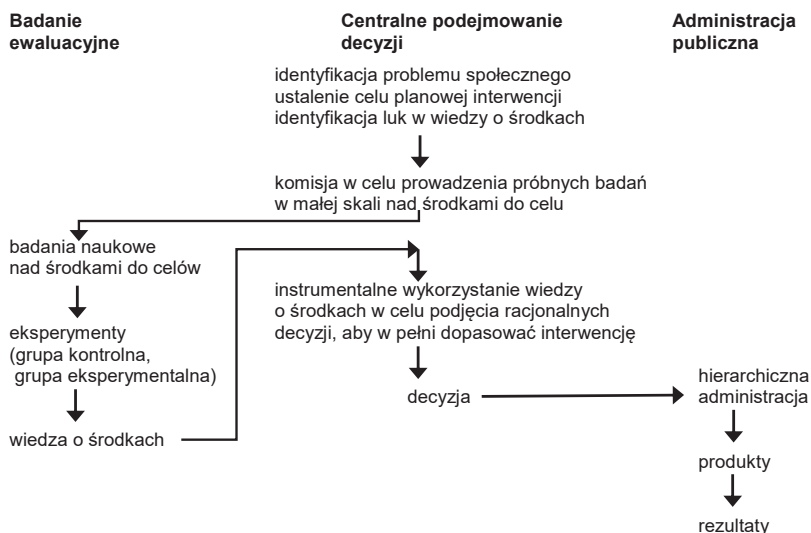
<sup>78</sup> G. Praweńska-Skrzypek i B. Jałocha, *Projektyzacja sektora publicznego w Polsce – implikacje dla organizacji samorządu terytorialnego*, „Zarządzanie Publiczne” 2014, nr 3 (27), s. 273–284; B. Jałocha et al., *Key Competences of Public Sector Project Managers*, „Procedia – Social and Behavioral Sciences” 2014, t. 119, s. 247–256.

<sup>79</sup> E. Vedung, *Four Waves of Evaluation Diffusion*, „Evaluation” 2010, t. 16, nr 3, s. 263–277.

w tym bardzo złożone i długotrwałe, miały być stosunkowo bezproblemowo adoptowane, wdrażane i efektywne. Ewaluacja efektów interwencji oraz jej związek z procesem podejmowania decyzji miały przebiegać zgodnie z modelem, który Vedung określa mianem inżynierskiego modelu interwencji publicznej (por. rysunek 1.7)<sup>80</sup>.

Rysunek 1.7

### Inżynierski model zainicjowania, przeprowadzenia i wykorzystania ewaluacji



Źródło: opracowanie na podstawie E. Vedung, Four Waves of Evaluation Diffusion, „Evaluation” 2010, t. 16, nr 3, s. 263–277.

Zgodnie z nim po odkryciu problemu analizuje się jego zakres, przyczyny oraz konsekwencje. Następnie staje się on przedmiotem decyzji politycznych, które określają zamierzone cele w odniesieniu do tego problemu. Pierwsza faza polega na tym, że w planie eksperymentalnym naukowcy jako obiektywni i neutralni obserwatorzy podejmują się testów w celu określenia najbardziej efektywnych instrumentów do osiągnięcia zamierzonych celów. Na tej podstawie, w drugiej fazie, decydenci polityczni podejmują decyzje o wdrożeniu programu interwencji na dużą skalę przez menadżerów, którzy w sposób neutralny i zgodnie z zamierzeniami decydentów realizują tę interwencję.

W tym podejściu ewaluację traktuje się jako formę neutralnego i obiektywnego badania naukowego, którego wyniki są traktowane jako prawdziwe, neutralne w odniesieniu do wartości i wykorzystane w podejmowaniu decyzji.

<sup>80</sup> H.A. Simon, *Administrative behavior: a study of decision-making processes in administrative organizations*, Free Press, New York 1997, 4th ed.



Vedung podkreśla instrumentalne wykorzystanie ewaluacji w tym przypadku, gdyż nie obejmuje ona formułowania problemów czy określania celów.

Falę drugą, która zaczęła kształtować się w połowie lat 70., Vedung określa jako dialogiczne podejście do ewaluacji<sup>81</sup>. Charakteryzuje się ono przede wszystkim wprowadzeniem pojęcia interesariusza, co miało poszerzyć grono osób zaangażowanych w ewaluację poza wąskim gronem decydentów politycznych, naukowców oraz kadry menadżerskiej i administracyjnej. Ewaluacja miała identyfikować grupy interesariuszy, uwzględniać ich perspektywę oraz przebiegać w dialogu i deliberacji, prowadząc do konsensusu, demokratycznej legitymizacji i politycznej akceptowalności typowych dla racjonalności komunikacyjnej<sup>82</sup>. Interakcja zachodząca między poszczególnymi grupami interesariuszy wiąże się z intensywnym procesem uczenia się zarówno w określaniu własnej perspektywy, jak i perspektyw innych grup. Z uwagi na elementy konstruktywistyczne (nie istnieje prawda obiektywna, gdyż jest wiele rzeczywistości, które są konstrukcjami umysłu, wyniki badania są emergentnym efektem interakcji obserwatora i obserwowanego, metodologię eksperymentalną zastępuje się zaś procesem hermeneutyczno-dialogowym) niekiedy traktuje się to podejście jako przeciwstawne podejściu naukowemu fali pierwszej<sup>83</sup>.

Koniec lat 70. przyniósł kolejną, neoliberalną falę w ewaluacji, która towarzyszyła nowemu zarządzaniu publicznemu<sup>84</sup>. Z drugą falą dzieliła ona brak zaufania do centralnego planowania, typowego dla fali pierwszej. Charakterystycznym jej elementem była pro-rynkowa orientacja na klienta, a towarzyszyła jej decentralizacja, deregulacja i prywatyzacja. W ewaluacji nastąpiło odejście od dotychczasowego skupienia na procesie w stronę wyników, co miało odzwierciedlać biznesowe praktyki w zarządzaniu i na trwałe wprowadziło ewaluację zorientowaną na rozliczalność (*accountability*). Nowe zarządzanie publiczne charakteryzowało się zaufaniem do przywództwa jako podstawy organizacji zorientowanych na wyniki (kontrakty czasowe, premiowanie wydajności, profesjonalizacja), wykorzystaniem pośrednich form kontroli (prywatyzacja i outsourcing usług publicznych, zwiększenie konkurencji, wynagrodzenie premie, systemy kontroli jakości, delegowanie kontroli i odpowiedzialności, zarządzanie przez cele, benchmarking itp.) oraz zorientowaniem na klienta i obywatela (wybór dostawcy, satysfakcja klienta, reprezentacja w procesie podejmowania decyzji itp.).

---

<sup>81</sup> Innym, często stosowanym określeniem było miano „ewaluacji demokratycznej”, a także „fourth generation evaluation”, które wprowadzili E.G. Guba i Y.S. Lincoln, *Fourth generation evaluation*, Sage Publications, Newbury Park, Calif 1989.

<sup>82</sup> Por. M.A. Hajer i H. Wagenaar, red., *Deliberative policy analysis: understanding governance in the network society*, Cambridge University Press, Cambridge, UK, New York, USA 2003.

<sup>83</sup> E.G. Guba i Y.S. Lincoln, *Fourth generation evaluation*, op. cit.; P. Dahler-Larsen, *From Programme Theory to Constructivism*, op. cit.

<sup>84</sup> C. Pollitt i G. Bouckaert, *Public management reform a comparative analysis*, Oxford University Press, Oxford; New York 2000; C. Pollitt, *The essential public manager*, Open University Press, Maidenhead 2003; C. Pollitt, S. van Thiel i V. Homburg, *New public management in Europe: adaptation and alternatives*, Palgrave Macmillan, Basingstoke [England], New York 2007.

Najbardziej znanym i charakterystycznym elementem nowego zarządzania publicznego jest zarządzanie oparte na wynikach. Podstawą procesu jest w tym przypadku informacja zwrotna o osiągnięciu założonych celów, co wiązało się z naciskiem na stałe prowadzenie ewaluacji i monitoringu oraz ewaluację *ex post*. Z uwagi na outsourcing usług publicznych podmioty publiczne przejęły rolę nadzorowania zleceniobiorców, gdzie ewaluacja ma na uwadze przede wszystkim rozliczalność dostawcy usług.

Kolejna fala pojawiała się w połowie lat 90. pod wpływem przemian, jakie zaszły w medycynie, i nosi miano „oparcia na dowodach” (*evidence-based*)<sup>85</sup>. Towarzyszyła ona pojawianiu się międzynarodowych form współpracy (np. The Campbell Collaboration, The Cochrane Collaboration), które dokonywały systematycznego przeglądu dowodów na temat skuteczności wcześniejszych programów interwencji publicznej w zakresie np. pomocy społecznej, zdrowia publicznego czy edukacji. Waga różnych rodzajów dowodów uwarunkowana jest tym, na ile wskazują one zależności przyczynowe, które mogą być podstawą udanych interwencji<sup>86</sup>.

Tabela 1.4

#### Hierarchia dowodów naukowych w metaanalizie

Eksperymenty losowe (podwójnie ślepe)
Badania <i>quasi</i> -eksperymentalne (z wykorzystaniem dopasowania)
Badanie porównawcze przed-i-po
Przekrojowe badania z próbą losową
Ewaluacja procesowa, badanie formatywne i badanie w działaniu
Jakościowe studia przypadków i badania etnograficzne
Opisowe przewodniki i przykłady dobrych praktyk
Opinia profesjonalistów i ekspertów
Opinia użytkownika

Źródło: opracowanie na podstawie: E. Vedung, *Four Waves of Evaluation Diffusion*, „Evaluation” 2010, t. 16, nr 3, s. 263–277.

Ta hierarchia prowadzi do odrzucenia ideału nowego zarządzania publicznego, gdyż opinie profesjonalne i eksperckie są niemal na samym dole, podczas gdy najwyżej cenione są – jak w fali pozytywistycznej – badania eksperymentalne. W tej fali ewaluacja ujmowana jest na podobnej zasadzie co dyscypliny społeczne, w przeciwieństwie do perspektywy politycznej, administracyjnej czy zorientowanej na klienta. Niekiedy twierdzi się, że jest to powrót tradycji pozytywistycznej i racjonalności instrumentalnej, gdzie główna rola ewaluacji polega na osiągnięciu i u powszechnieniu wiedzy o środkach prowadzących do celu.

<sup>85</sup> R. Pawson, *Evidence-based policy: a realist perspective*, SAGE, London 2006; P. Davies, *The State of Evidence-Based Policy Evaluation and its Role in Policy Formation*, „National Institute Economic Review” 2012, t. 219, nr 1, s. R41–R52.

<sup>86</sup> R. Pawson, *Evidence-based policy*, op. cit., s. 49.

Inną specyficzną w tym ujęciu cechą ewaluacji prowadzonych przez organizacje międzynarodowe jest podejście metaprzedmiotowe (zwl. metaanaliza, analiza syntetyzująca czy systematyczne przeglądy). Traktuje się je jako bardziej zobiektywizowane i niezależne niż pojedyncze ewaluacje, dzięki czemu wskazują one na prawdopodobne zależności przyczynowe, a jednocześnie mogą ujawniać luki w dotychczasowej wiedzy i wskazywać nowe obszary badań ewaluacyjnych. Badania ewaluacyjne przebiegają w tej fali odmiennie niż w fali naukowej, gdyż praktykę badań ewaluacyjnych, wykonywanych na zlecenie danego podmiotu, łączy się z badaniami podstawowymi prowadzonymi niezależnie w instytucjach badawczych.

Zaletą ujęcia falowego głównych etapów rozwoju ewaluacji jako dyscypliny jest możliwość odniesienia do szerszych zmian w sposobach postrzegania polityk publicznych, a także regionalnego zróżnicowania efektów sedimentacji. Rysunek 1.8 obrazuje je graficznie za pomocą Ngrams Google<sup>87</sup>.

## 2. Teorie, podejścia i modele ewaluacji

### 2.1. Rola teorii w ewaluacji

Systematyczna refleksja nad procesem poznawczym w czasie badań jest cechą charakterystyczną poznania naukowego w ogóle, a nauk społecznych w szczególności<sup>88</sup>. F. Leeuw i S. Donaldson podkreślają, że w studiach nad ewaluacją panuje pewna wieloznaczność terminu „teoria”. Po pierwsze, jest on używany w sensie ogólnych założeń programu, dotyczących behawioralnych, społecznych czy ekonomicznych mechanizmów, mających wpływ na jego realizację, oraz tego, jak one z kolei wpływają na interwencje podejmowane w ramach tego programu. Często w tym sensie jest on zamiennie używany z takimi terminami, jak „logika programu” czy „model logiczny”<sup>89</sup>. Po drugie, termin „teoria” rozumiany jest jako strategia i taktyki, które ewaluatorzy wykorzystują jako podstawę swojej praktyki, aby zagwarantować adekwatność i efektywność działań, a nie kierować się tylko metodą prób i błędów<sup>90</sup>. Tak rozumiana teoria jest wyznacznikiem planu ewaluacji, doboru personelu, metod i technik czy sposobu przeprowadzenia ewaluacji. Najbardziej znani ewaluatorzy promują swoje teorie jako zalecaną praktykę ewaluacyjną. W obu wymienionych znaczeniach termin

<sup>87</sup> OECD, „Evaluation and impact assessment of STI policies”, op. cit., s. 3.

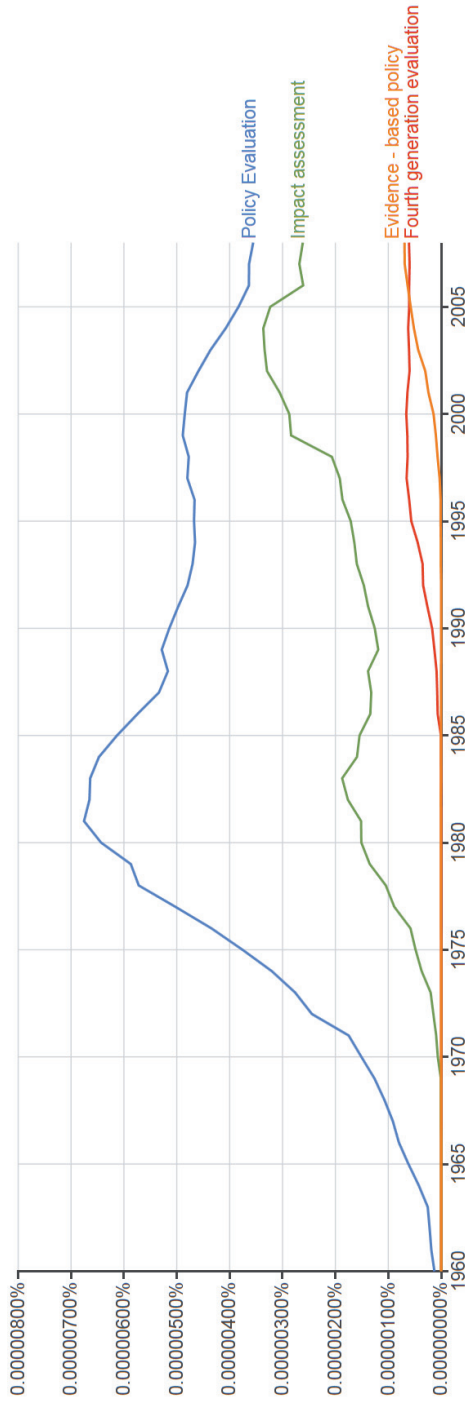
<sup>88</sup> N. Döring i J. Bortz, red., *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, Springer, Berlin Heidelberg 2016, 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage, Springer-Lehrbuch, s. 32.

<sup>89</sup> M. Hansen, M.C. Alkin i T.L. Wallace, *Depicting the logic of three evaluation theories*, „Evaluation and Program Planning” 2013, t. 38, s. 34–43.

<sup>90</sup> W.R. Shadish, T.D. Cook i L.C. Leviton, red., *Foundations of program evaluation: theories of practice*, Sage Publications, Newbury Park 2000, Reprinted, s. 34.

## Rysunek 1.8

## Efekt sedymantacji czterech fal w rozwoju ewaluacji jako dyscypliny badawczej



Źródło: opracowanie własne.

„teoria” jest odnoszony do intencji, oczekiwań i innych procesów poznawczych samych ewaluatorów i znacznie odbiega od rozumienia tego terminu w metodologii nauk społecznych<sup>91</sup>.

W sensie właściwym teorie ewaluacji wyjaśniają konsekwencje polityk, programów i działań ewaluacyjnych w sposób spójny z innymi teoriami z nauk społecznych<sup>92</sup>. Mogą one dotyczyć samego procesu tworzenia polityk publicznych, ich skuteczności i efektywności, dynamiki czy implementacji w organizacjach, jak również wyjaśnienia zachowania się ewaluandów w ich społecznym i instytucjonalnym kontekście podczas wdrażania interwencji publicznej. Jak podkreślają Leeuw i Donaldson, wykorzystanie tego rodzaju teorii, które są dobrze potwierdzone empirycznie i odporne, jest podstawowym wymogiem badań ewaluacyjnych, gdyż dostarczają istotnych przesłanek na temat mechanizmów i kontekstów, istotnych dla polityk i programów, które są przedmiotem badań.

### 2.1.1. Ogólne własności teorii

Odróżnia się ogólne i szczegółowe cechy teorii ewaluacji<sup>93</sup>. Cechy ogólne charakteryzują szeroką klasę ewaluacji niezależnie od kontekstu ich zastosowania, zagadnienia, czasu i miejsca. Ogólne teorie określają etapy procesu ewaluacji oraz zasady oceny i uzasadniania. Szczegółowe cechy natomiast charakteryzują ewaluacje pod podobnymi względami, jednak z uwzględnieniem ewaluandów, zagadnień tematycznych, miejsca i czasu zastosowania. Stufflebeam teorię ewaluacji definiuje jako „spójny zbiór zasad pojęciowych, hipotetycznych, pragmatycznych i etycznych tworzących ogólny schemat dla badań i praktyki ewaluacyjnej”<sup>94</sup>. Zgodnie z tą definicją teoria ewaluacyjna powinna posiadać następujące cechy: całościowa spójność, aparat pojęciowy, sprawdzone hipotezy dotyczące wpływu procedur ewaluacyjnych na osiągnięcie zamierzonych produktów, zasady pragmatyczne (oparte na dotychczasowej praktyce i usystematyzowane w przewodnikach), wymogi etyczne oraz ogólny schemat kierowania praktyką ewaluacyjną i prowadzenia badań nad ewaluacją. Elementami teorii są najczęściej: założenia, kontekst, działania, konsekwencje oraz czynniki zewnętrzne<sup>95</sup>. J. Greene pod-

---

<sup>91</sup> F.L. Leeuw i S.I. Donaldson, *Theory in evaluation: Reducing confusion and encouraging debate*, „Evaluation” 2015, t. 21, nr 4, s. 470.

<sup>92</sup> H.T. Chen i N.C. Turner, *Formal Theory Versus Stakeholder Theory: New Insights From a Tobacco-Focused Prevention Program Evaluation*, „American Journal of Evaluation” 2012, t. 33, nr 3, s. 395–413.

<sup>93</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 45.

<sup>94</sup> Ibidem, s. 50.

<sup>95</sup> M. Mark, *Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation*, [w:] *Fundamental issues in evaluation*, red. N.L. Smith i P.R. Brandon, Guilford Press, New York 2008, s. 111–134; R.L. Miller, *Developing Standards for Empirical Examinations of Evaluation Theory*, op. cit.; J.C. Greene, *Logic and evaluation theory*, „Evaluation and Program Planning”, 2013, t. 38, s. 71–73.

kreśla, że istotnym aspektem ewaluacji jest praktyka społeczna, wobec czego zawsze wykraczać będzie ona poza schematyczność modeli logicznych.

Powyższej definicji nie spełnia większość modeli czy podejść określaných jako „teorie” w szeroko cytowanej przeglądowej pracy pod redakcją M. Alkina<sup>96</sup>, dlatego zdecydowanie częściej używa się określenia „podejście” niż „teoria”<sup>97</sup>. Ponadto, większe znaczenie dla dyscypliny, jaką jest ewaluacja, mają nie ogólne teorie, lecz teorie średniego poziomu lub lokalne, spełniające kryterium aplikacyjności, o którym mowa w kolejnym podpunkcie<sup>98</sup>.

Do elementów aparatu pojęciowego wypracowanych przez teoretyków ewaluacji Stufflebeam i in. zaliczają m.in.: ewaluację formatywną i sumatywną, konstruktywistyczną i responsywną, model CIPP, ewaluację zorientowaną na wykorzystanie wyników, ewaluację partycypacyjną, a także zestaw standardów – obejmujących ewaluację oraz metaewaluację – opracowanych w kolejnych edycjach The Program Evaluation Standards amerykańskiego stowarzyszenia Joint Committee on Standards for Educational Evaluation.

Współcześnie, wobec braku satysfakcjonujących rezultatów ewaluacji, także w politykach publicznych, oraz braku efektów uczenia się na ich podstawie, w teorii ewaluacji coraz większą wagę przywiązuje się do interakcji różnych agentów w poszczególnych obszarach polityki oraz do wpływu czynników instytucjonalnych<sup>99</sup>. Jedną z istotnych propozycji ujęć teoretycznych jest wersja nowego instytucjonalizmu, zorientowana na aktorów, która wskazuje na oddziaływanie ewaluacji na wielu poziomach podejmowania decyzji w procesie tworzenia polityk publicznych<sup>100</sup>.

---

<sup>96</sup> M.C. Alkin, red., *Evaluation roots: tracing theorists' views and influences*, Sage Publications, Thousand Oaks, CA 2004.

<sup>97</sup> J.C. Greene, *Logic and evaluation theory*, op. cit., s. 73.

<sup>98</sup> Ibidem; B. Campbell i A. McGrath, *Where the rubber hits the road: The development of usable middle-range evaluation theory*, [w:] *Social psychology and evaluation*, red. M.M. Mark, S.I. Donaldson i B. Campbell, Guilford Press, New York 2011, s. 346–372.

<sup>99</sup> M.A. Hajer i H. Wagenaar, *Deliberative policy analysis*, op. cit.; E.R.H. Fuchs, *Rethinking the role of the state in technology development: DARPA and the case for embedded network governance*, „Research Policy” 2010, t. 39, nr 9, s. 1133–1147; R. Agranoff, M. McGuire i C. Silvia, *Governance, networks and intergovernmental systems*, [w:] *Routledge handbook of public policy*, red. E. Araral, Routledge, London 2013, s. 361–73; H.F. Hansen, *Systemic evaluation governance. New logics in the Development of organisational fields*, „Scandinavian Journal of Public Administration” 2013, t. 16, nr 3, s. 47–64; S. Borrás i J. Edler, *The Governance of Socio-Technical Systems*, Edward Elgar Publishing 2014, <http://www.elgaronline.com/view/9781784710187.xml>; G. Capano, M. Howlett i M. Ramesh, *Varieties of Governance: Dynamics, Strategies, Capacities*, 2015; OECD, *Governance and implementation of innovation policies*, op. cit.; P. Kawalec, *Problemy koncepcji polityki innowacyjnej jako „przedsiębiorczego państwa”*, „Zagadnienia Naukoznawstwa” 2017, t. 53, nr 2(212), s. 185–206.

<sup>100</sup> J. Streicher, *Evaluations, actors and institutions: the case of research, technology and innovation policy in Austria*, Vienna University of Economics and Business, Vienna 2017, <http://epub.wu.ac.at/5445/>.

### 2.1.2. Tworzenie teorii a kryteria oceny i standardy

Tworzenie teorii ma zasadnicze znaczenie dla postępów w praktyce ewaluacyjnej. Zazwyczaj opiera się ono na studium praktyki ewaluacyjnej, uwzględnieniu kontekstów, tworzeniu nowej aparatury pojęciowej, jej operacjonalizacji i zastosowaniach, a także późniejszym jej modyfikowaniu. Mimo braku reguł tworzenia teorii pomocne są kryteria, które służą do oceny proponowanych teorii.

Oprócz trudności, jakie w tworzeniu teorii napotykają wszystkie nauki społeczne, teoria ewaluacji dodatkowo boryka się z koniecznością uwzględniania różnorodnych kontekstów (społecznych, politycznych, geograficznych, czasowych) programów, które są przedmiotem ewaluacji<sup>101</sup>. Duża zmienność kontekstów, ich niepowtarzalność i nieprzewidywalność utrudnia tworzenie wskazówek działań praktycznych wyprowadzonych z teorii, a także dokonywanie przewidywań. Z tego względu uznaje się, że proces tworzenia teorii ewaluacyjnych musi mieć ciągły charakter i muszą być one regularnie oceniane i aktualizowane<sup>102</sup>, gdyż „kontekst jest aktywnym czynnikiem, kształtującym charakter oraz znaczenia określonych działań ewaluacyjnych i ich konsekwencji (a także założeń)”<sup>103</sup>.

Funkcje, jakie powinna pełnić teoria ewaluacji, wymagają spełnienia przez te ostatnie określonych kryteriów. Niektóre z tych funkcji wprost wskazuje definicja teorii ewaluacji Scrivena: „ogólne ujęcie pewnej dziedziny zjawisk, które prowadzi do sformułowania co najmniej wyjaśnień, a czasem również przewidywań i uogólnień; często, ale niekoniecznie, angażuje ona byty teoretyczne, które nie są bezpośrednio obserwowalne”<sup>104</sup>. Jako kryteria ewaluacji teorii ewaluacyjnych – „najważniejszej ze wszystkich działań naukowych”, wymienia on zatem: moc eksplanacyjną, moc i dokładność predykcyjną, ekonomiczność założeń, owocność zastosowań, posiadanie potwierdzenia w różnych dyscyplinach (np. przez analogię) oraz prostotę twierdzeń.

W. Shadish i in. spośród funkcji teorii ewaluacji eksponują przede wszystkim aplikacyjność: „fundamentalnym celem ewaluacyjnej teorii programu jest określenie wykonalnych praktyk, które ewaluatorzy mogą wykorzystać w celu skonstruowania wiedzy o wartości programów społecznych, która może zostać wykorzystana do zmniejszenia odnośnych problemów społecznych”<sup>105</sup>. Kryteria oceny teorii ewaluacyjnej obejmują więc następujące obszary: programowanie społeczne (natura programów społecznych i ich rola w rozwiązywaniu problemów społecznych), konstruowanie wiedzy (dotyczącej przedmiotu ewaluacji, prawomocne metody generowania tej wiedzy, założenia dotyczące istotnych rodzajów

<sup>101</sup> A.T. Vo i C.A. Christie, *Advancing Research on Evaluation Through the Study of Context: Advancing Research on Evaluation Through the Study of Context*, „New Directions for Evaluation” 2015, t. 2015, nr 148, s. 43–55.

<sup>102</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 58.

<sup>103</sup> J.C. Greene, *Logic and evaluation theory*, op. cit., s. 72.

<sup>104</sup> M. Scriven, *Evaluation thesaurus*, op. cit., s. 360.

<sup>105</sup> W.R. Shadish, T.D. Cook i L.C. Leviton, red., *Foundations of program evaluation*, op. cit., s. 36.

wiedzy), wartościowanie (rola wartości i procesu wartościowania w ewaluacji oraz sposoby konstruowania sądów wartościujących), wykorzystanie wiedzy (cele i sposoby jej wykorzystania, metody zwiększenia efektywności jej wykorzystania), podmioty oraz praktyka społeczna i ewaluacyjna. Dokonany na podstawie tych kryteriów selektywny przegląd dotychczasowych (obejmujący okres do 1991 r.) propozycji teoretycznych wyłonił trzy główne fazy w rozwoju teorii ewaluacyjnych, podkreślając, że brak wśród nich takich, które zadowalająco spełniałyby wszystkie podane kryteria. Niemniej, najbliższej ich spełnienia są teorie Cronbacha i Rossiego.

Zestaw kryteriów dotyczący aspektu związku teorii ewaluacji z praktyką opracowała R. Miller<sup>106</sup>. Zaliczyła do nich: specyficzność operacyjną (wyraźnie teoretycznie ugruntowane jasne wskazówki dla praktyków), zakres zastosowania (wskazanie, w jakich praktycznych okolicznościach oraz dla jakiego rodzaju pytań ewaluacyjnych teoria ma zastosowanie), praktyczną wykonalność (stopień trudności w zastosowaniu idei teoretycznych), dający się wyodrębnić wpływ (wskazuje zamierzone i pożądane wpływy oraz występowanie niezamierzonych oddziaływań) oraz powtarzalność (ten sam wpływ może być powtórzony w innym czasie, w innych zastosowaniach oraz przy innych ewaluatorach)<sup>107</sup>.

Za główne kryteria oceny teorii ewaluacji Stufflebeam i in. uznają wynikające z samej definicji teorii<sup>108</sup>: użyteczność w efektywnym generowaniu weryfikowalnych przewidywań lub propozycji działań ewaluacyjnych oraz ich konsekwencji, formułowanie rzetelnych, prawomocnych i wykonalnych wskazań dla etycznego prowadzenia efektywnych ewaluacji programów. Bardziej szczegółowe kryteria obejmują: przejrzystość i kompleksowość założeń, oszczędność, odporność, generalizowalność oraz moc heurystyczną.

### 2.1.3. Pluralizm teorii

Dzieje ewaluacji, jak wskazano w punktach 1.1 oraz 1.4, charakteryzuje duże zróżnicowanie, zarówno co do przedmiotu, sposobu podejścia, jak i charakteru ogólnych, teoretycznych założeń. Stufflebeam i in.<sup>109</sup> jako uzasadnienie pluralizmu wśród teorii ewaluacji wskazują istnienie licznych podejść teoretycznych: pozytywistyczne, egzystencjalistyczne, konstruktywistyczne, obiektywistyczne oraz postmodernistyczne. Związane są one przede wszystkim z różnym sposobem konceptualizacji samej ewaluacji, ale rzutują na rozstrzygnięcia metodologiczne,

<sup>106</sup> R.L. Miller, *Developing Standards for Empirical Examinations of Evaluation Theory*, op. cit.

<sup>107</sup> Przykładem zastosowania tych kryteriów do oceny jednej z teorii ewaluacji jest: M.C. Alkin, *When is a theory a theory? A case example*, „Evaluation and Program Planning” 2017, t. 63, s. 141–142.

<sup>108</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 52; F. Lawrenz i D. Huffman, *Methodological pluralism: The gold standard of STEM evaluation*, „New Directions for Evaluation” 2006, nr 109, s. 19–34.

<sup>109</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 58–59.



gdyż odrzucenie pluralizmu wiąże się z narzucaniem jednego rodzaju metody, np. badań eksperymentalnych<sup>110</sup>.

W przypadku teorii, które reprezentują zbliżone orientacje, F. Leeuw i S. Donaldson<sup>111</sup> proponują integrowanie elementów różnych teorii w celu stworzenia teorii wyższego rzędu, np. przez połączenie takich elementów tych teorii, które generują potwierdzalne w doświadczeniu predykcje. G. Westhorp<sup>112</sup> natomiast, biorąc pod uwagę przypadki zróżnicowanego oddziaływania programu, który okazuje się skuteczny w jednych grupach, wykazując długookresowy wpływ, a nieskuteczny w innych, wykazując tylko wpływ krótkookresowy, proponuje warstwowe podejście do łączenia teorii<sup>113</sup>. Podobnie jak w poprzednim przypadku podstawowym wymogiem jest pokrewieństwo orientacji oraz tematyki różnych podejść teoretycznych. Teoria „wyższej” warstwy dla teorii niższej warstwy – na zasadzie przyczynowania odgórnego – podaje uwarunkowania, które jednocześnie charakteryzują odrębny mechanizm przyczynowy. Zaletą podejścia warstwowego jest przekroczenie ograniczeń teoretycznych jednego tylko podejścia i uwzględnienie bardziej adekwatnego zbioru zmiennych w badaniach.

## 2.2. Klasyfikacje teorii ewaluacji

Mimo stosunkowo krótkiej historii ewaluacji jako odrębnej dyscypliny badawczej posiada ona bogatą paletę podejść teoretycznych i praktycznych. Z uwagi na to zróżnicowanie często omawia się różne systemy klasyfikacyjne teorii ewaluacji, aby pełniej uchwycić tę wielość i odmienność<sup>114</sup>. W kolejnych podpunktach zaprezentowane zostały cztery różne sposoby klasyfikacji teorii, które opierają się też na różnych kryteriach. Podejście Guby i Lincolna, określone jako „generacyjne”, skupia się na liniowej i czasowej sekwencji kolejnych generacji teorii ewaluacji. Geneza teorii jest tylko fragmentarycznie wyznacznikiem „drzewa klasyfikacyjnego” Alkina, w którym poszczególne konary i gałęzie reprezentują przede wszystkim pokrewieństwo założeń teoretycznych. Klasyfikacja systematyczna Fitzpatricka, Sandersa i Worthena zasadniczo nie uwzględnia kryterium chronologicznego, lecz skupia się na celu ewaluacji oraz sposobach wykorzystania jej wyników. Ostatnia z omówionych klasyfikacji Stockmana i Meyera jako główne kryterium uporządkowania podejść teoretycznych wybiera funkcje społeczne ewaluacji.

<sup>110</sup> F. Lawrenz i D. Huffman, *Methodological pluralism*, op. cit., s. 20.

<sup>111</sup> F.L. Leeuw i S.I. Donaldson, *Theory in evaluation*, op. cit.

<sup>112</sup> G. Westhorp, *Using complexity-consistent theory for evaluating complex systems*, „Evaluation” 2012, t. 18, nr 4, s. 405–420.

<sup>113</sup> Podobne podejście zaproponował E. Lieberman: E.S. Lieberman, *Nested Analysis as a Mixed-Method Strategy for Comparative Research*, „American Political Science Review” 2005, t. 99, nr 03, s. 435–452.

<sup>114</sup> R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, Palgrave Macmillan, Basingstoke 2013, rozdział 3.

### 2.2.1. Generacyjna (Guba, Lincoln)

Wprowadzona w klasycznej monografii Guby i Lincoln<sup>115</sup> klasyfikacja stanowiła podsumowanie serii wcześniejszych prób uporządkowania teorii ewaluacji<sup>116</sup>, co było utrudnione z uwagi na duże rozproszenie tej tematyki wśród różnych dyscyplin (badania nad edukacją, psychologia, socjologia, ekonomia itp.) w początkowym okresie rozwoju ewaluacji. Posługuje się ona kategorią „generacji”, czyli podejść do ewaluacji kolejnych pokoleń, na oznaczenie etapów radykalnych zmian, jakie się dokonały, określając przy tym własne podejście jako generację czwartą. Nie ma ona także ambicji wyczerpania wszystkich możliwych podejść, lecz raczej zarysowania punktów odniesienia dla alternatywnych teorii ewaluacji. Mimo to ma pewne zalety. Jedną z nich jest prostota, także odzwierciedlona w nazewnictwie kolejnych generacji, hasłowo wyodrębnionych wokół:

- pomiaru,
- opisu,
- oceny oraz
- negocjacji.

Za wyróżnik pierwszej generacji uznano zbieranie danych ilościowych oraz ilościowe próby pomiaru za pomocą testów psychometrycznych (W. Wundt, F. Galton). Do tej generacji przypisane zostały nazwiska również innych autorów (np. Rice, Binet, Otis, Hildreth), którzy wprost nie identyfikowali swoich badań jako ewaluacji. Główne podstawy teoretyczne tej generacji tworzyła scjentyficznie zorientowana szkoła J.F. Herbarta, zwłaszcza prace J.M. Rice’a, który prowadził badania nad amerykańskim systemem edukacji pod koniec XIX w., posługując się m.in. porównawczymi testami wyników kształcenia dla wielu szkół amerykańskich, które można uznać za pierwowzór badań OECD Program for International Student Assessment (PISA)<sup>117</sup>.

Do drugiej generacji zaliczony został przede wszystkim Tyler jako zwolennik opisywania wzorców mocnych i słabych stron w odniesieniu do ustalonych celów, co ma stanowić główne zadanie ewaluacji<sup>118</sup>. W ten sposób zmieniona została rola pomiaru – już nie głównego zadania ewaluacji, lecz jej narzędzia, jednego spośród wielu innych<sup>119</sup>.

<sup>115</sup> E.G. Guba i Y.S. Lincoln, *Fourth generation evaluation*, op. cit.

<sup>116</sup> Por. m.in. B. Worthen i J. Sanders, *Educational Evaluation. Theory and Practice*, „Evaluation Review” 1973, t. 13, nr 1, s. 18–31; W.J. Popham, *Educational evaluation*, Prentice-Hall, Englewood Cliffs, New York 1975; E.R. House, *Assumptions Underlying Evaluation Models*, „Educational Researcher” 1978, t. 7, nr 3, s. 4–12.

<sup>117</sup> R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 111; najnowsze wydanie (31.08.2017): OECD, *PISA 2015 Assessment and Analytical Framework*, OECD Publishing 2017, PISA, [http://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework\\_9789264281820-en](http://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework_9789264281820-en).

<sup>118</sup> E.G. Guba i Y.S. Lincoln, *Fourth generation evaluation*, op. cit., s. 28.

<sup>119</sup> R.W. Tyler, *General Statement on Evaluation*, „The Journal of Educational Research” 1942, t. 35, nr 7, s. 492–501.

Wyznacznikiem trzeciej generacji jest wyeksponowanie roli oceny w ewaluacji. Wśród przedstawicieli tej generacji zostali uwzględnieni: M. Scriven, R. Stake czy D. Stufflebeam. W znanym artykule *The countenance of educational evaluation* z 1967 roku Stake wprowadził – obok macierzy opisu, charakteryzującej program i uzyskane w jego realizacji wyniki – nowy element w teorii ewaluacji, jakim była macierz oceny (*judgement matrix*), w której dokonuje się porównania zamierzonej jakości i zalet programu z faktycznie uzyskanymi wynikami oraz standardami.

Czwarta generacja wyrasta z idei Stake'a ewaluacji responsywnej, która w istotny sposób uwzględnia perspektywy grup interesariuszy oraz bierze pod uwagę zróżnicowanie wartości. Proponowane przez siebie podejście Guba i Lincoln określa się jako „naturalistyczne” z uwagi na analogię do badań etnograficznych, gdzie ewaluator bada przebieg programu bez ingerowania w jego przebieg, ograniczania go czy manipulowania. Ewaluator uczy się od ewaluowanych ich perspektywy, kategorii, jakimi opisują świat, oraz ich sposobu wyjaśniania. Bardzo ważne jest zidentyfikowanie istotnych grup interesariuszy, z którymi opracowuje się wstępny zestaw pytań w ramach tzw. kół hermeneutycznych. Wyróżnikiem tego podejścia jest prezentowanie i społeczne konstruowanie wyników ewaluacji podczas procesu ich negocjowania z interesariuszami, co jest jednocześnie pierwszym krokiem w kolejnym cyklu ewaluacji. Dwoma podstawowymi założeniami, tworzącymi „pień” drzewa, są rozliczalność i kontrola (*accountability and control*) oraz szeroko rozumiana metodologia badań społecznych (*social inquiry*).

### 2.2.2. Drzewo klasyfikacyjne (Alkin, Christie)

Uporządkowanie nurtów w teorii ewaluacji w postaci drzewa oparte jest na założeniu o istnieniu wspólnych elementów, które następnie uległy dywersyfikacji. Na wspólnym pniu osadzone są trzy główne konary, których podstawą zróżnicowania są następujące kategorie: użycie, metody oraz wartościowanie. Za inicjatora podejścia skoncentrowanego na użyciu wyników ewaluacji Alkin i Christie uznali Stufflebeama, który ewaluację traktował zasadniczo jako źródło wiedzy i informacji dla osób podejmujących decyzje (model CIPP). Patton wykorzystanie wyników odnosi natomiast przede wszystkim do interesariuszy, a rolę ewaluatora postrzega w kategoriach uczenia się oraz wspierania rozwoju samego programu (ewaluacja rozwojowa, *developmental evaluation*).

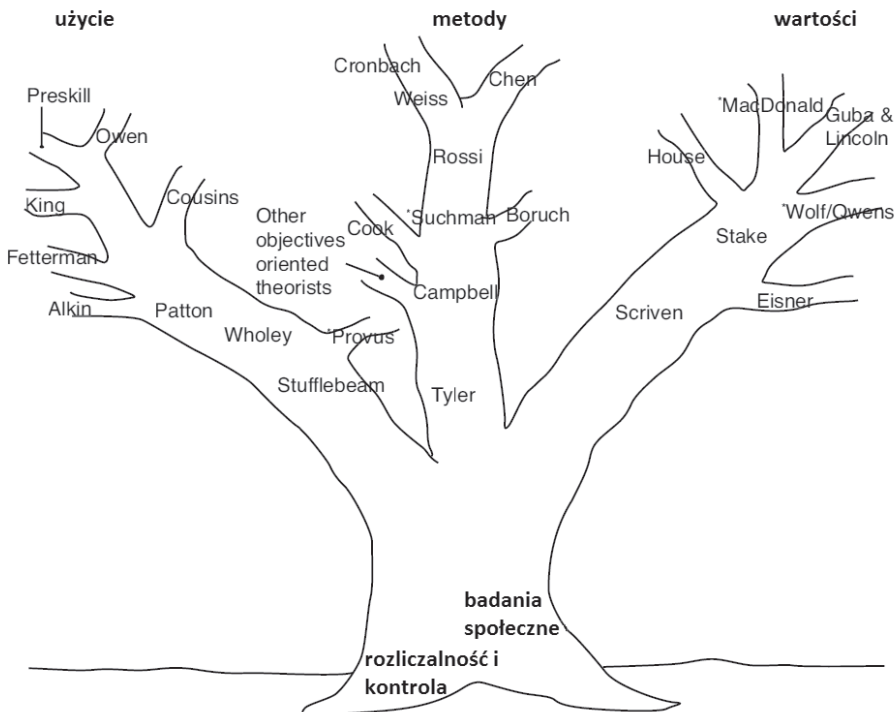
Bezpośrednią kontynuacją empirycznych badań społecznych jest konar „metod”, którego rozwój zapoczątkowały prace Tylera. Jego główny wkład upatruje się tu w teoretycznym ufundowaniu ewaluacji zorientowanej na cele, którą w swoich pracach rozwijali później m.in. Bloom, Hammond, Metfessel i Popham. Wpływ ten jednak nie obejmował autorów, którzy również umieszczeni zostali na tym samym konarze: Campbell (koncepcja *quasi*-eksperymentalnego planu

badan), Suchman (ewaluacja jako dyscyplina naukowa) czy Rossi (ewaluacja kierunkowana teorią programu)<sup>120</sup>.

Trzeci główny konar czerpie swoje podstawy z teorii wartościowania Scrivena<sup>121</sup>. Ewaluacja w tym ujęciu polega przede wszystkim na wskazaniu pozytywnych i negatywnych wartości związanych z programem i jego realizacją nie tylko na potrzeby decyzji podejmowanych w odniesieniu do danego programu, lecz również na potrzeby dyskursu publicznego. Wśród kontynuatorów tego podejścia wskazani są m.in. Stake oraz Guba i Lincoln.

**Rysunek 1.9**

**Drzewo klasyfikacyjne Alkina i Christie**



Źródło: opracowanie na podstawie: A.C. Alkin i C.A. Christie, *Evaluation theory tree*, [w:] *Evaluation roots: tracing theorists' views and influences*, red. M.C. Alkin, Sage Publications, Thousand Oaks, CA 2004, s. 13.

<sup>120</sup> Szerzej argumentację przeciw bezpośredniemu wiązaniu Tylera z Campbellem, Suchmanem czy Rossim Stockman i Meyer prezentują w: R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 123n.

<sup>121</sup> W.R. Shadish, T.D. Cook i L.C. Leviton, red., *Foundations of program evaluation*, op. cit., s. 91.

Drzewo Alkina i Christie należy traktować jako propozycję uporządkowania typologicznego, które nie wyklucza obecności danego nazwiska na kilku konarach. Dotyczy to zwłaszcza klasyków teorii ewaluacji, których wkład nie ogranicza się tylko do jednego z podanych kryteriów. Jak podkreślają Stockman i Meyer<sup>122</sup>, to uporządkowanie trudno uznać za adekwatne szczególnie w odniesieniu do rozwoju teorii i praktyki ewaluacji na gruncie europejskim, gdzie mamy do czynienia ze znacznie większym zróżnicowaniem kontekstów narodowych, sektorowych czy dziedzinowych niż w kontekście dziejów ewaluacji w USA. Ponadto, metafora drzewa nie oddaje we właściwy sposób oddziaływań i wzajemnych inspiracji, do jakich dochodziło między autorami, którzy zostali ulokowani na odrębnych konarach (np. Scriven, Patton i Stufflebeam).

### 2.2.3. Systematyzacyjna (*Fitzpatrick, Sanders, Worthen*)

Uporządkowanie systematyzacyjne jest motywowane praktycznym podejściem do ewaluacji, mającym charakter eklektyczny i uzasadnianym pragmatycznie. Fitzpatrick, Sanders i Worthen<sup>123</sup> wyodrębniają cztery zasadnicze kategorie podejść do ewaluacji:

- zorientowane na charakterystykę programu;
- zorientowane na podejmowanie decyzji dotyczących programu;
- zorientowane na kompleksową ocenę jakości programu, usługi lub produktu;
- zorientowane na uczestnictwo interesariuszy.

Do pierwszej kategorii zalicza się podejścia, w których ewaluacja jest wykorzystywana przede wszystkim do określenia, czy zamierzone cele programu zostały osiągnięte. Podejście zainicjowane przez Tylera Sanders poszerzył o badanie spójności i struktury logicznej. Druga kategoria obejmuje przede wszystkim podejścia ukierunkowane na perspektywę menadżerską kierowania i administrowania projektem w sposób, który umożliwi osiągnięcie jego zaplanowanych celów. Głównym przedstawicielem tego podejścia jest Stufflebeam (model CIPP), lecz obejmuje ono także odmienne podejścia, jak model Alkina czy Pattona. Podkreśla się tu formatywną rolę ewaluacji, gdyż pozwala ona na bardziej elastyczne kierowanie programem jeszcze w trakcie jego realizacji. Głównym zagrożeniem tego podejścia jest ignorowanie perspektywy interesariuszy oraz zbytne zaangażowanie ewaluatora w realizację programu.

Kompleksowa ocena jakości programu na ogół dotyczy ewaluacji realizowanych w perspektywie konsumenta (np. *Consumer Reports* czy *Stiftung Warentest*), najczęściej na zlecenie organizacji konsumenckich lub rządu. Ocena ma głównie charakter porównawczy dostępnych alternatyw i często bazuje na liście niezbędnych kryteriów, którą w badaniach programów edukacyjnych zainicjował

<sup>122</sup> R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 134.

<sup>123</sup> J.L. Fitzpatrick, J.R. Sanders i B.R. Worthen, *Program evaluation: alternative approaches and practical guidelines*, Pearson Education, Upper Saddle River, New York 2011, 4th ed.

Scriven. Tego rodzaju podejście jest na ogół mocno ograniczone przez ewaluanda, a także dość sztywno określony zestaw kryteriów oraz wykorzystywanie głównie ewaluacji sumatywnej. Szczególną grupę w tej kategorii tworzą ewaluacje oparte na wiedzy eksperckiej, a ich najbardziej reprezentatywnym przykładem są akredytacje i certyfikacje.

Ewaluacje, które podkreślają zaangażowanie interesariuszy, często motywowane są przez dążenie do zmiany społecznej i wyrównania szans dla nieuprzywilejowanych czy wykluczonych grup interesariuszy. Odrzuca się tu<sup>124</sup> kategorie charakterystyczne dla naukowego podejścia do ewaluacji, upatrując zasadniczy cel ewaluacji w uprawnianiu interesariuszy (*empowerment*), dzięki czemu sami będą mogli doprowadzić do zmiany swojej sytuacji społecznej<sup>125</sup>.

Pewnym ograniczeniem klasyfikacji Fitzpatricka i in. jest brak uwzględnienia perspektywy dynamiki teorii ewaluacji, a także założenie o wzajemnym wykluczaniu się poszczególnych podejść, w związku z czym problematyczne byłoby tu umieszczenie jednego nazwiska w więcej niż jednej kategorii. Ponadto, dominującą perspektywą dla ujęcia systematyzacyjnego są badania ewaluacyjne w pedagogice i psychologii.

#### 2.2.4. Funkcjonalna (Stockman, Meyer)

Podstawowym założeniem funkcjonalnej typologii Stockmana i Meyera jest to, że ewaluacje odzwierciedlają ogólne trendy zmian społecznych, dla których współcześnie charakterystyczne są: racjonalność, sprawiedliwość, przejrzystość i uczestnictwo<sup>126</sup>. To one właśnie są odniesieniem dla legitymizacji polityk publicznych w dyskursie demokratycznym. Stockman i Meyer wyróżniają w tym kontekście cztery podstawowe rodzaje funkcji ewaluacji, traktując je jak wymiary wspólnej przestrzeni, w której funkcjonuje ewaluacja:

- dostarczenie wiedzy (wiedza);
- wspomaganie kontroli (kontrola);
- zainicjowanie rozwoju i procesów uczenia się (rozwój);
- legitymizacja wdrażanych programów (uprawomocnienie).

W przypadku wiedzy chodzi nie tylko o dowody naukowe (dotyczące programu, potrzeb interesariuszy, zależności przyczynowej jako podstawy interwencji itd.), lecz także o inne rodzaje wiedzy, które są istotne dla podejmowanych decyzji, dotyczących danej interwencji publicznej. Taka wiedza może wynikać

---

<sup>124</sup> D.M. Fetterman, *Foundations of empowerment evaluation*, SAGE Publications, Thousand Oaks, CA 2001.

<sup>125</sup> Mark i in. nie zgadzają się z podejściem, że tylko duże zaangażowanie interesariuszy może prowadzić do zmiany społecznej (*social betterment*), gdyż za konieczne uznają również adresowanie ewaluacji do decydentów: M.M. Mark, G.T. Henry i G. Julnes, *Evaluation*, op. cit.

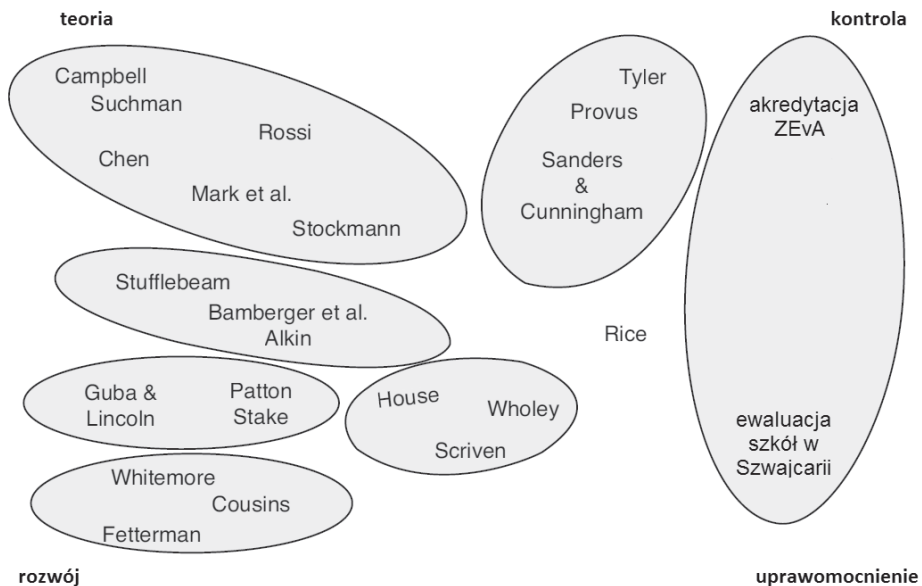
<sup>126</sup> R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 73–76.

z analizy wewnętrznych struktur i procesów, jakie zachodzą w systemie politycznym i administracyjnym. Funkcja kontroli dotyczy osiągania efektów (efektywność, skuteczność, trwałość itd.) i ma odniesienie do różnych wymiarów działania (prawny, ekonomiczny, społeczny, administracyjny itd.). Dwie pierwsze funkcje dostarczają wiedzy i informacji, które z kolei umożliwiają rozwój programu i dialog między różnymi grupami interesariuszy (sponsorzy programu, instytucje wdrażające, grupy docelowe itd.). Dla rozwoju programu istotna jest transparentność w komunikacji wyników, gdyż warunkuje ona procesy uczenia się, które umożliwiają rozwijanie programu. Legitymizacja programu jest szczególnie istotna w przypadku istnienia alternatywnych programów i jest uwarunkowana ewaluacją sumatywną, która przekazuje informacje o osiągniętych rezultatach, wpływie programu, jego efektywności, trwałości itd.

W ramach określonej funkcjonalnie czterowymiarowej przestrzeni ewaluacji Stockman i Meyer przeprowadzają charakterystykę różnych zagadnień: przedstawicieli różnych podejść do ewaluacji (rysunek 1.10), polityki edukacyjnej oraz procesów i procedur w ewaluacji.

**Rysunek 1.10**

**Systematyzacja funkcjonalna podejść do ewaluacji**



Źródło: R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 159.

## 2.3. Główne typy podejść do ewaluacji

### 2.3.1. Zorientowane na poprawę i rozliczalność

Wspólną cechą tej grupy podejść do ewaluacji jest ambicja pełnej oceny wartości programu, uwzględnienia wszystkich istotnych pytań oraz kryteriów (jakość, zalety, wpływ, uczciwość, znaczenie, wykonalność, efektywność, bezpieczeństwo, równość itd.). Przy ocenie jakości programu zwykle uwzględniają one potrzeby interesariuszy, a często również wszystkie – nie tylko wyznaczone przez cele programu – rezultaty i wpływy realizacji programu. Na ogół dominuje nastawienie obiektywizujące do ewaluacji, założenie istnienia niezależnej rzeczywistości jest podstawą poszukiwania jednoznacznych odpowiedzi na pytania ewaluacyjne. Podejścia do ewaluacji zorientowane na poprawę i rozliczalność wykorzystują też bogate spektrum metod ilościowych i jakościowych w celu dokonania krzyżowej weryfikacji ustaleń oraz podkreślają dużą rolę precyzji i kompleksowości ewaluacji. Wymagają one przy tym wprowadzenia jednoznacznych definicji jakości i zalet programu, precyzyjnego doboru i analizy danych dotyczących wszystkich istotnych zmiennych.

Do zasadniczych celów tego podejścia należy: 1) doprowadzenie do poprawy jakości życia społecznego lub zwiększenie rozliczalności przez wpływanie na proces decyzyjny dotyczący danego programu; 2) ułatwienie konsumentom dokonania właściwego wyboru danego programu, produktu lub usługi; 3) pomoc instytucjom akredytującym w przeprowadzeniu certyfikacji instytucji, programów lub pracowników. Jego ograniczeniem może być niewykonalność w pełnym zamierzonym zakresie, a także nadmiar ustaleń w stosunku do podmiotu, który zleca tego rodzaju ewaluację.

Tego rodzaju podejście reprezentują:

1. Badania zorientowane na proces decyzji oraz rozliczalność.
2. Badania zorientowane na konsumenta.
3. Akredytacja i certyfikacja.

Badania zorientowane na proces decyzji i rozliczalność łączą retrospektywną ocenę wartości programu z możliwością jego poprawy, posługując się zarówno ewaluacją formatywną, jak i sumatywną. Ich reprezentatywnym modelem jest UTOS Cronbacha<sup>127</sup> oraz CIPP Stufflebeama<sup>128</sup>, w którym istotną rolę w całym procesie ewaluacji odgrywa perspektywa klienta administrującego procesem, będącym przedmiotem ewaluacji. Model CIPP obejmuje: kontekst programu (zwłaszcza potrzeby motywujące zmiany lub ulepszenia), nakłady (strategie, plan

---

<sup>127</sup> L.J. Cronbach i K. Shapiro, *Designing evaluations of educational and social programs*, Jossey-Bass, San Francisco 1982, A Joint publication in the Jossey-Bass series in social and behavioral science & in higher education.

<sup>128</sup> D.L. Stufflebeam, *The CIPP Model for Program Evaluation*, [w:] *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, red. G.F. Madaus, M.S. Scriven i D.L. Stufflebeam, Springer Netherlands, Dordrecht 1983, s. 117–141.



operacyjny, zasoby, uzgodnienia dotyczące realizacji interwencji), proces (wdrożenie interwencji oraz jego koszty) i produkty (pozytywne i negatywne rezultaty programu). Głównym zadaniem ewaluacji jest tu określenie, czy zidentyfikowane potrzeby odbiorców programu są zaspokajane w sposób efektywny (także kosztowo), etyczny oraz co najmniej tak dobry, jak alternatywne programy. Z uwagi na konieczność utrzymywania stałego kontaktu z interesariuszami to podejście jest bliskie ewaluacji uczestniczącej. Może być to jednak powodem utrudnień w ewaluacji lub obciążenia uzyskanych wyników, zwłaszcza gdy w grę wchodzi kontekst polityczny bądź gdy ewaluatorzy zaczynają zbytnio identyfikować się z programem, tracąc obiektywną perspektywę.

W badaniach zorientowanych na konsumenta, które zainicjowały prace Scrivena, ewaluator przyjmuje rolę „oświeconego klienta”. Głównym celem ewaluacji jest pomoc w tworzeniu produktów lub usług, które będą miały wysoką jakość i użyteczność dla ich odbiorców, konsumentów (np. uczniowie, ich rodzice, nauczyciele, podatnicy itp.). Ewaluacja pomaga w tym przypadku konsumentom w określeniu i ocenie jakości, zalet oraz znaczenia konkurencyjnych programów, usług i produktów. Zakłada się przy tym istnienie obiektywnej rzeczywistości, która ma być podstawą – często trudnej – oceny i wskazania rozwiązań najlepszych wśród dostępnych alternatyw. Podstawowym wymogiem jest poddawanie ewaluacji metaewaluacji<sup>129</sup>. Metodologia tego podejścia zakłada wykorzystanie „logiki ewaluacji” Scrivena<sup>130</sup>, która obejmuje cztery zasadnicze kroki: ustalenie kryteriów jakości, skonstruowanie standardów, pomiar ewaluandum i porównanie ze standardami oraz syntezę wyników w postaci oceny jakości, zalet lub znaczenia ewaluandum. Taka ewaluacja wymaga doświadczonego badacza, który ma dostęp do niezbędnych źródeł informacji. Jej ograniczeniem jest słabe zwykle angażowanie interesariuszy oraz niewielkie możliwości poprawy danego programu, a przeprowadzona zbyt wcześnie może zahamować kreatywność twórców programu.

Akredytacja i certyfikacja obejmują przede wszystkim instytucje i jednostki świadczące usługi w celu określenia, czy spełniają one wymogi danej profesji oraz czy realizują swoją funkcję społeczną<sup>131</sup>. Są ukierunkowane przez ustalone wcześniej kryteria akredytacji lub certyfikacji. Zwykle poprzedza je raport sporządzony przez sam podmiot, który następnie jest weryfikowany przez ewaluatora. Instytucja akredytująca lub certyfikująca zamyka cały proces, podejmując decyzję o udzieleniu akredytacji lub certyfikatu na określony okres bądź też decyzję warunkową lub negatywną. Tę odmianę ewaluacji zapoczątkowała amerykańska the College Entrance Examination Board, a później the Cooperative Study of Secondary School Standards, w pierwszych dekadach XX wieku. Słabą stroną

<sup>129</sup> M. Scriven, *Meta-Evaluation Revisited*, „Journal of MultiDisciplinary Evaluation” 2009, t. 6, nr 11, s. iii–viii.

<sup>130</sup> M. Scriven i C.L.S. Coryn, *The logic of research evaluation*, „New Directions for Evaluation” 2008, nr 118, s. 89–105.

<sup>131</sup> M. Hughes i S. Kushner, *Accreditation*, [w:] *Encyclopedia of Evaluation*, red. S. Mathison, Sage Publications, Thousand Oaks, CA 2005, s. 4–7.

tego podejścia jest podatność na zniekształcanie wyników podczas etapu wizytacji, gdy instytucja prezentuje wyłącznie udane elementy swojej działalności lub stara się wpłynąć na ewaluatorów w celu uzyskania pozytywnego raportu.

### 2.3.2. Zorientowane na rozwój społeczny

Podejścia zorientowane na rozwój społeczny motywowane są przede wszystkim zapewnieniem wszystkim grupom społecznym równego dostępu do programów społecznych i edukacyjnych. Poprzez ewaluację programów dążą do wsparcia grup wykluczonych. Charakterystyczne dla nich jest podejście konstruktywistyczne oraz wykorzystanie metod jakościowych, a w przypadku podejść transformacyjnych<sup>132</sup> – korzystanie z metod mieszanych. Ogólne założenia obejmują filozofię postmodernistyczną, pluralizm kulturowy, relatywizm moralny i istnienie wielości rzeczywistości. Kładą też nacisk na zaangażowanie interesariuszy w proces tworzenia i interpretacji wyników. Do tego rodzaju podejść zalicza się:

1. Ewaluacja responsywna lub zorientowana na interesariuszy.
2. Ewaluacja konstruktywistyczna.
3. Deliberatywna ewaluacja demokratyczna.

Ograniczeniem tych podejść jest możliwość narażenia wyników ewaluacji na ich nadużywanie przez silnie zaangażowanych interesariuszy, a także – przy zbytym zaangażowaniu samego ewaluatora – na konflikt interesów i utratę obiektywnej perspektywy, koniecznej do trafnej oceny jakości programu. Inicjatorzy tego rodzaju podejść to m.in.: R. Stake<sup>133</sup>, E. Guba, Y. Lincoln<sup>134</sup>, E. House<sup>135</sup>, D. Mertens<sup>136</sup>.

Najbardziej charakterystycznym podejściem dla tej grupy jest ewaluacja responsywna, która ma charakter pluralistyczny, interaktywny, holistyczny oraz jest zorientowana na subiektywną interpretację, dla której nie poszukuje się rozstrzygającej perspektywy, stąd jest ona także określana jako relatywistyczna. Głównym celem ewaluacji jest ustalenie całościowego obrazu („oblicza” – *countenance*) programu przez wskazanie różnych i rozbieżnych perspektyw na wartości, które stoją u podstaw programu i do których realizacji on prowadzi, włącznie z jego efektami ubocznymi

<sup>132</sup> D.M. Mertens, *Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods*, Sage, Los Angeles 2010, 3. wyd.; D.M. Mertens i M. Tarsilla, *Mixed Methods Evaluation*, [w:] *The Oxford handbook of multimethod and mixed methods research inquiry*, red. S.N. Hesse-Biber i B. Johnson, Oxford University Press, Oxford; New York 2015, s. 426–446; D.M. Mertens, *The Role of Evaluation in Social Change: Perspectives on the Future of Evaluation from the USA*, [w:] *The Future of Evaluation*, red. R. Stockmann i W. Meyer, Palgrave Macmillan UK, London 2016, s. 214–227.

<sup>133</sup> R.E. Stake, *Program evaluation, particularly responsive evaluation*, [w:] *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, red. G.F. Madaus, M.S. Scriven i D.L. Stufflebeam, Springer Netherlands, Dordrecht 1983, s. 287–310.

<sup>134</sup> E.G. Guba i Y.S. Lincoln, *Fourth generation evaluation*, op. cit.

<sup>135</sup> E.R. House, *Deliberative democratic evaluation*, [w:] *Encyclopedia of Evaluation*, red. S. Mathison, Sage Publications, Thousand Oaks, CA 2005, s. 104–108.

<sup>136</sup> D.M. Mertens, *Research and evaluation in education and psychology*, op. cit.

(zamierzonymi i niezamierzonymi). To podejście jest szczególnie uwrażliwione na zapewnienie równego dostępu do programu dla grup wykluczonych lub nieuprzywilejowanych. Silnie akcentuje ono także współpracę z różnymi grupami interesariuszy, dla których ewaluator staje się istotnym doradcą na drodze rozumienia i ulepszania programu, mającego zaspokajać ich potrzeby. Wykorzystuje się na ogół metody słabo sformalizowane, oparte na ciągłej komunikacji między ewaluatorem a interesariuszami, studia przypadków, obserwację, narrację, a nawet socjodramę. Podejście to zainicjowały prace Stake'a<sup>137</sup>, a kontynuował je Guba w swoim podejściu konstruktywistycznym (naturalistycznym). Odrzucił on metody badań zbliżone do nauk przyrodniczych (zwł. randomizowany eksperyment), charakteryzując wraz z Lincolnem swoje podejście jako „czwartą generację” (por. podpunkt 1.4).

Deliberatywna ewaluacja postuluje stosowanie zasad demokratycznych przy ustalaniu wyników, traktując ewaluację jako wpływową instytucję społeczną, która istotnie oddziałuje na demokratyzację przez generowanie rzetelnych i wiarygodnych ustaleń<sup>138</sup>. Interesariusze aktywnie uczestniczą w procesie przygotowywania ostatecznych wyników ewaluacji, co ma zagwarantować, że żadna istotna perspektywa nie została w nich pominięta. Sami twórcy tego podejścia traktują je jednak jako pewien normatywny ideał, trudny do zrealizowania w praktyce<sup>139</sup>.

### 2.3.3. Podejścia eklektyczne

Niektórzy ewaluatorzy nie podejmują wyraźnego zobowiązania do przyjęcia założeń jednego z podejść i pragmatycznie wykorzystują różne podejścia. W odróżnieniu od pseudoewaluacji starają się jednak spełnić kryteria ewaluacji. Do tych podejść zalicza się m.in. podejście form ewaluacyjnych Owena<sup>140</sup>, klastrową ewaluację stosowaną przez W.K. Kellogg Foundation<sup>141</sup>, ewaluację zorientowaną na wykorzystanie wyników Pattona<sup>142</sup> oraz różne odmiany ewaluacji uczestniczącej<sup>143</sup>.

---

<sup>137</sup> R.E. Stake, *The countenance of educational evaluation*, op. cit.; R.E. Stake, *Program evaluation, particularly responsive evaluation*, op. cit.; K.M. Sturges i C. Howley, *Responsive Meta-Evaluation: A Participatory Approach to Enhancing Evaluation Quality*, „American Journal of Evaluation” 2017, t. 38, nr 1, s. 126–137.

<sup>138</sup> M.A. Hajer i H. Wagenaar, *Deliberative policy analysis*, op. cit.; W.M. Trochim et al., *The Evaluation of Large Research Initiatives: A Participatory Integrative Mixed-Methods Approach*, „American Journal of Evaluation” 2008, t. 29, nr 1, s. 8–28.

<sup>139</sup> E.R. House, *Assumptions Underlying Evaluation Models*, op. cit.; E.R. House, *Deliberative democratic evaluation*, op. cit.

<sup>140</sup> J.M. Owen, *Program evaluation: forms and approaches*, Guilford Press, New York 2007, 3rd ed.

<sup>141</sup> R. Millett, *W.K. Kellogg Foundation cluster evaluation model of evolving practices*, W.K. Kellogg Foundation, Battle Creek, MI 1995.

<sup>142</sup> M.Q. Patton, K. McKegg i N. Wehipeihana, red., *Developmental evaluation exemplars: principles in practice*, The Guilford Press, New York 2016.

<sup>143</sup> J.B. Cousins i L.M. Earl, *The Case for Participatory Evaluation*, „Educational Evaluation and Policy Analysis” 1992, t. 14, nr 4, s. 397–418.

### 2.3.4. Pseudoewaluacje i quasi-ewaluacje

Pseudoewaluacje to działania, w których dokonuje się manipulacji na wynikach, jak ich selektywne ujawnianie, nadmierne uogólnienie lub fałszowanie czy zniekształcanie samego procesu, np. podczas kształcenia ewaluatorów bądź wdrażania procedur ewaluacyjnych w organizacji. Motywem takich działań może być chęć zysku albo cele polityczne<sup>144</sup>. Do pseudoewaluacji zalicza się: ewaluację jako *public relations* (celem nie jest poszukiwanie prawdy o programie, lecz upowszechnienie jego pozytywnego obrazu), politycznie kontrolowane badania (ich wyniki udostępniane są tylko wybranym grupom odbiorców), „uległą” (*pandering*) ewaluację („ustala” wyniki zgodne z oczekiwaniami zamawiającego w zamian za korzyści), ewaluację jako pretekst (znajdowanie uzasadnienia dla z góry narzuconych wyników), wzmacnianie pozycji pod pozorem ewaluacji (grupa wykluczona, mając zbyt dalekosiężny wpływ na ewaluację, ogłasza swoją ocenę jako raport niezależnego ewaluatora) oraz selektywną informację zwrotną dla konsumentów (na ogół selektywne zestawienie ocen produktu lub usługi przez klientów, które jest publikowane przez podmiot działający komercyjnie). Często zdarza się, że dane podejście wykorzystuje kilka różnych form pseudoewaluacji<sup>145</sup>.

*Quasi-ewaluacja* to zdaniem Stufflebeam i in. takie podejście, które zasadniczo spełnia wymogi rzetelnej ewaluacji, jednak jest zbyt zawężone pod względem zakresu, jakiego dotyczą pytania, lub zastosowanego doboru metod. Może być to skutkiem posługiwania się definicją celów programu, które zawężają zbiór istotnych zmiennych, lub skupienia się wyłącznie na pozytywnych (odpowiednio: negatywnych) aspektach udanego (nieudanego) programu. Jednym z argumentów przemawiających za zawężeniem podejścia do ewaluacji jest to, że daje ono możliwość uzyskania precyzyjnej odpowiedzi w stosunkowo krótkim czasie, co jest bardziej efektywne niż ewaluacja długookresowa, całościowa, mierząca kompleksowo wpływ programu<sup>146</sup>. Największym wyzwaniem dla *quasi-ewaluacji* jest ich adekwatne zaprezentowanie, aby uniknąć nieporozumień i mylenia ich z pełnozakresowymi kompleksowymi ewaluacjami. Często także *quasi-ewaluacje* są zbyt zawężone, w związku z czym nie spełniają standardów poprawnej ewaluacji. Przykłady *quasi-ewaluacji* obejmują: zainicjowane przez Tylera badania oparte na celach programu (określające, na ile zostały one faktycznie osiągnięte), studia dobrych praktyk (określające udane elementy programu

---

<sup>144</sup> Stockam i Meyer wprowadzają tu odrębną funkcję „taktyczną” instrumentalnego wykorzystania ewaluacji, por. R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 76.

<sup>145</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 117–132.

<sup>146</sup> L.J. Cronbach i R.E. Snow, *Aptitudes and instructional methods: a handbook for research on interactions*, Irvington Publishers: distributed by Halsted Press, New York 1977.

w celu ich kontynuacji i rozwoju)<sup>147</sup>, ewaluację rezultatów jako wartości dodanej (dostarczającej informacji dla projektowania programu, ale też istotnej dla jego rozliczalności)<sup>148</sup>, badania eksperymentalne i *quasi*-eksperymentalne (ukierunkowane przede wszystkim na ustalenie zależności przyczynowych)<sup>149</sup>, analizę kosztów (określającą, na ile inwestycja w program przyniosła korzyść w postaci osiągnięcia założonych celów programu oraz szerszych korzyści społecznych)<sup>150</sup>, ewaluację opartą na teorii (ukierunkowaną na ocenę zasadności programu na podstawie przesłanek teoretycznych oraz zrozumienia jego sukcesu/porażki)<sup>151</sup> oraz metaanalizę (w celu zebrania, podsumowania, analizy i wyciągnięcia wniosków o efektach interwencji publicznej na podstawie dostępnych wiarygodnych badań danej interwencji)<sup>152</sup>.

Badania eksperymentalne odgrywają szczególną rolę w ustalaniu zależności przyczynowych w badaniach<sup>153</sup>. Mimo szerokiego zakresu stosowania badań eksperymentalnych w ewaluacji<sup>154</sup> (usługi, zatrudnienie, opieka zdrowotna, programy kulturowe, edukacja itd.) istnieje szereg wątpliwości dotyczących ich stosowania<sup>155</sup>. Po pierwsze, grupa kontrolna pozbawiona jest dostępu do korzyści, jakie daje interwencja. Po drugie, bardzo trudno jest zachować odrębność obu grup, np. w przypadku programów edukacyjnych realizowanych w różnych szkołach. Po trzecie, informacje z prowadzonej ewaluacji nie są wykorzystywane do poprawy programu w trakcie jego trwania. Po czwarte, istnieje wiele sytuacji, w których nie można zastosować tego rodzaju planu badań.

---

<sup>147</sup> R.O. Brinkerhoff, *The success case method: find out quickly what's working and what's not*, Berrett-Koehler, San Francisco, CA 2003.

<sup>148</sup> W.L. Sanders i S.P. Horn, *Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research*, „Journal of Personnel Evaluation in Education” 1998, t. 12, nr 3, s. 247–256.

<sup>149</sup> W.R. Shadish, T.D. Cook i D.T. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin, Boston 2001.

<sup>150</sup> H.M. Levin i P.J. McEwan, *Cost-effectiveness analysis: methods and applications*, Sage Publications, Thousand Oaks, CA 2001, 2. wyd.

<sup>151</sup> C.L.S. Coryn et al., *A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009*, „American Journal of Evaluation” 2011, t. 32, nr 2, s. 199–226.

<sup>152</sup> M. Borenstein, red., *Introduction to meta-analysis*, John Wiley & Sons, Chichester, U.K 2009.

<sup>153</sup> P. Kawalec, *Przyczyna i wyjaśnianie: studium z filozofii i metodologii nauk*, Wydawnictwo KUL, Lublin 2006.

<sup>154</sup> R.F. Boruch, *Randomized field trials in education*, [w:] *International handbook of educational evaluation*, red. T. Kellaghan i D.L. Stufflebeam, Kluwer Academic Publishers, Norwell, MA 2003, s. 107–124.

<sup>155</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 252–254.

### 3. Metody i plan badań ewaluacyjnych

#### 3.1. Metody badań ewaluacyjnych

Ewaluacja korzysta przede wszystkim z metod badań, jakie stosują nauki społeczne<sup>156</sup>. Posługuje się podobnymi metodami zbierania, generowania, analizy i interpretacji oraz reprezentowania danych dotyczących jakości programu, praktyki lub polityki publicznej. Metody badań ewaluacyjnych przechodziły podobne fale jak dyskusje w naukach społecznych. Początkowo dominowały metody ilościowe z naciskiem na ustalanie zależności przyczynowych w planie badań eksperymentalnym lub *quasi*-eksperymentalnym oraz na standaryzację pomiaru zjawisk społecznych. W latach 1970–1980 toczyła się ożywiona dyskusja dotycząca wyboru między metodami ilościowymi a jakościowymi na kanwie zwrotu interpretatywnego w naukach społecznych. Wskazywano na kontekstowy, konstrukcjonistyczny oraz wartościujący charakter poznania rzeczywistości społecznej. Lata 90. przyniosły akceptację prawomocności różnych tradycji metodologicznych w badaniach ewaluacyjnych, a także próby tworzenia mieszanych planów badań, które integrowały różne rodzaje metod<sup>157</sup>. Za specyfiką metodologiczną ewaluacji uznaje się przede wszystkim jej ocenny charakter, związany z wartościującymi kategoriami, jak „efektywny” czy „udany”, kontekst polityczny, wynikający z konieczności współpracy z różnorodnymi prawomocnymi interesariuszami, oraz relacyjność, zakorzenioną w stosunkach, jakie zachodzą między grupami interesariuszy a ewaluatorem w konkretnych kontekstach<sup>158</sup>.

W badaniach ewaluacyjnych łączy się na ogół wiele metod i technik badawczych, które są dobierane i adaptowane do wybranych badań ewaluacyjnych z uwagi na określone kryteria: rodzaj interwencji publicznej, cel badania ewaluacyjnego, zakres badania ewaluacyjnego, rodzaj badania ewaluacyjnego odnoszącego się do cyklu polityki publicznej oraz rodzaj badania ewaluacyjnego odnoszącego się do cyklu procesu ewaluacji, który obejmuje projektowanie, pozyskiwanie danych, ich analizę i formułowanie wniosków<sup>159</sup>.

---

<sup>156</sup> I.F. Shaw, J. Greene i M. Mark, *Handbook of evaluation*, op. cit., s. 20; Salem Press, *Research & evaluation methods*, op. cit.

<sup>157</sup> J.C. Greene, V.J. Caracelli i W.F. Graham, *Toward a Conceptual Framework for Mixed-Method Evaluation Designs*, „Educational Evaluation and Policy Analysis” 1989, t. 11, nr 3, s. 255–274; H. Chen, *Applying mixed methods under the framework of theory-driven evaluations*, „New Directions for Evaluation” 1997, nr 74, s. 61–72; L.J.C. Riggan, *Advances in mixed-method evaluation: A synthesis and comment*, „New Directions for Evaluation” 1997, nr 74, s. 87–94.

<sup>158</sup> C.A. Christie i D.N. Fleischer, *Insight Into Evaluation Practice: A Content Analysis of Designs and Methods Used in Evaluation Studies Published in North American Evaluation-Focused Journals*, „American Journal of Evaluation” 2010, t. 31, nr 3, s. 326–346.

<sup>159</sup> B. Turowski i M. Zawicki, *Funkcje, etapy, metody i narzędzia ewaluacji*, [w:] *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, red. S. Mazur, UE Katowice, MSAP, Kraków 2007, s. 46.

Na etapie projektowania ewaluacji definiowany jest zakres badań, określone są pytania badawcze oraz koncepcja planu badań. Odpowiednio do założeń określonej interwencji publicznej dobiera się wskaźniki i określa grupy beneficjentów, które będą przedmiotem badania<sup>160</sup>. Opracowuje się też narzędzia ewaluacyjne (kwestionariusze, wzorce do analizy danych statystycznych). Do metod i technik wykorzystywanych na tym etapie należą: analiza SWOT, ewaluacja formatywna, konsultacje z interesariuszami, mapa oddziaływania, metody uczestniczące, modele logiczne oraz studium wykonalności.

Sposoby pozyskiwania danych z dostępnych źródeł są uwarunkowane przyjętą koncepcją ewaluacji. Do stosowanych tu metod i technik należą: analiza kosztów i korzyści, autoewaluacja, badania beneficjentów, badania społeczne, benchmarking, konsultacje z interesariuszami, schematy eksperymentalne i *quasi*-eksperymentalne, studia przypadku, techniki obserwacyjne, wykorzystanie danych administracyjnych lub wtórnych, wywiady indywidualne, zogniskowany wywiad grupowy.

Na etapie analizy dokonuje się interpretacji zebranych danych. Są one zestawiane i porównywane (np. tabele, wykresy, rysunki). Następnie w trakcie właściwej analizy weryfikuje się postawione hipotezy, określa relacje zmiennych, zwłaszcza przyczynowo-skutkowe, lub szacowanie efektów interwencji publicznej (zamierzonych i niezamierzonych). Wykorzystuje się metody i techniki analizy danych, jak: analiza efektywności kosztowej, analiza nakładów i produktów, analiza regresji, autoewaluacja przez uczestników programu, metoda delficka, modele ekonometryczne, modele logiczne, ocena strategicznego wpływu na środowisko, panel ekspertów, schematy eksperymentalne i *quasi*-eksperymentalne, studia przypadku.

Formułowanie wniosków jest kluczowym etapem procesu ewaluacji, uwarunkowanego poziomem złożoności ewaluandum, precyzją pytań ewaluacyjnych oraz zakresem zgromadzonych i przeanalizowanych danych. Wnioski obejmują formułowanie ocen, które jako czynność wartościująca wymaga odniesienia wyników analiz do określonych wartości i standardów. Przy dokonywaniu wartościowania także korzysta się z metod i technik, które je wspierają, jak: analiza efektywności kosztowej, analiza kosztów i korzyści, porównawcza analiza wielokryterialna, benchmarking, ocena wpływu na gospodarkę, ocena wpływu na środowisko.

W badaniach ewaluacyjnych często wykorzystuje się połączenia metod i technik w formie tzw. triangulacji, która obejmuje trzy poziomy danych:

1. Gromadzenie dostępnych danych z różnych źródeł (np. wyniki badań ankietowych, analiza dokumentacji lub studiów przypadku).
2. Zbieranie danych od różnych grup respondentów (np. zarządzający programem, beneficjenci) w celu uchwycenia różnych perspektyw interesariuszy programu.

---

<sup>160</sup> Por. zwł. R. Todeschini i A. Baccini, *Handbook of bibliometric indicators: quantitative tools for studying and evaluating research*, Wiley-VCH Verlag, Weinheim 2016.

3. Przeprowadzanie ewaluacji (gromadzenie danych, ich analiza) przez zespół badawczy jako forma wymiany i konfrontacji poglądów, wzajemnej kontroli oraz uzgadniania treści formułowanych wniosków<sup>161</sup>.

### 3.2. Plany badań ewaluacyjnych

Plan badań ewaluacyjnych ma charakter zbliżony do planu badań dowolnego projektu badawczego z zakresu nauk społecznych. Określa on ewaluandum, cele oraz zadania, wskazuje sposób przeprowadzenia ewaluacji, określa grupy osób, które wezmą udział w badaniu, oraz odbiorców wyniku badania. Specyfiką badań ewaluacyjnych jest zorientowanie na aplikację oraz to, że są one ograniczone jako badania zlecone na zamówienie, co zmniejsza autonomię badaczy<sup>162</sup>.

Ogólne zasady tworzenia planu badań ewaluacyjnych sformułował Suchman<sup>163</sup>:

- dobry plan badań powinien być dopasowany do celu ewaluacji (niezależnie od tego, czy ma ona charakter naukowy);
- nie istnieje jedyny poprawny plan, gdyż hipotezy można badać różnymi metodami w ramach różnych planów;
- każdy plan badań jest kompromisem, wynikającym z wielu praktycznych uwarunkowań;
- plan nie jest algorytmem, lecz raczej serią wskazówek służących utrzymaniu głównego kierunku działania ewaluacyjnego.

Do podstawowych planów badań ewaluacyjnych zalicza się:

- plan eksperymentalny i *quasi*-eksperymentalny;
- plan jakościowy;
- plan mieszany.

Współcześnie wprowadza się jednak bardziej kompleksowe całości, jakimi są zbiory planów określane jako „podejścia do planu badań”<sup>164</sup>. Ich zestawienie prezentuje tabela 1.5.

---

<sup>161</sup> B. Turowski i M. Zawicki, *Funkcje, etapy, metody i narzędzia ewaluacji*, op. cit., s. 58.

<sup>162</sup> R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 214.

<sup>163</sup> E.A. Suchman, *Evaluative research*, op. cit.

<sup>164</sup> E. Stern et al., *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*, op. cit.



Tabela 1.5

## Podejścia do planu badań ewaluacyjnych

Podejścia do planu badań	Szczegółowe odmiany	Podstawa wnioskowań przyczynowych
Eksperymentalne	Eksperymenty randomizowane, <i>quasi</i> -eksperymenty, eksperymenty naturalne.	Wnioskowania kontrfaktyczne, współwystępowanie przyczyny i skutków.
Statystyczne	Modelowanie statystyczne, badania długookresowe, ekonometria.	Korelacja między przyczyną a skutkiem lub między zmiennymi, wpływ (zwykle dających się wyizolować) wielu zmiennych na pojedynczy skutek. Kontrola konfunderów.
Oparte na teorii	Plany procesu przyczynowego: teoria zmiany, śledzenie procesu, analiza udziału, ścieżki wpływu.  Plany mechanizmu przyczynowego: ewaluacja realistyczna, analiza dyskryminacyjna.	Identyfikacja/konfirmacja procesów lub łańcuchów przyczynowych.  Badanie występowania wspierających czynników i mechanizmów w kontekście.
Oparte na przypadkach	Interpretatywne: naturalistyczne, teoria ugruntowana, etnografia.  Ustrukturyzowane: konfiguracje, jakościowa analiza porównawcza, pogłębione studium przypadku, symulacje i analiza sieciowa.	Porównania między oraz w obrębie przypadków kombinacji czynników przyczynowych.  Analityczna generalizacja oparta na teorii.
Uczestniczące	Plany normatywne: ewaluacja partycypacyjna lub demokratyczna, ewaluacja uprawniająca.  Plany sprawcze: uczenie się przez działanie, dialog polityczny, kolaboratywne badanie w działaniu.	Uprawomocnienie przez uczestników, ich działania i doświadczane skutki są spowodowane przez program.  Przyjęcie, dostosowanie i zobowiązanie do danego celu.
Studia syntezujące	Metaanaliza, synteza narracyjna, synteza oparta na badaniu realistycznym.	Akumulacja i agregacja w obrębie kilku perspektyw (statystycznej, opartej na teorii, etnograficznej itp.).

Źródło: E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies i B. Befani, *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*, „DFID: Department for International Development” 2012, s. 24.

### 3.2.1. Plan eksperymentalny i quasi-eksperymentalny

Główną zaletą stosowania planów badań wykorzystujących metody ilościowe jest uzyskanie wyników obiektywnych, opisanych na skalach matematycznych, interpretowalnych w sposób ścisły, jak również porównywalnych z innymi badaniami ilościowymi<sup>165</sup>. Wynika to głównie ze standaryzacji narzędzi badawczych oraz z wykorzystania statystycznych metod analizy danych. Metody ilościowe w badaniach ewaluacyjnych wykorzystuje się w celu określenia zależności między zmiennymi oraz częstotliwości analizowanego zjawiska. Podstawowym planem ilościowym jest plan eksperymentalny, który służy do badania zależności przyczynowo-skutkowej, gdy kontrolowane są zakłócające wpływy czynników ubocznych<sup>166</sup>. Najczęściej stosowany jest eksperyment klasyczny, charakteryzowany trzema parami składowych: zmiennymi niezależnymi i zależnymi, pretestem i posttestem oraz grupą eksperymentalną i kontrolną<sup>167</sup>, do których przydziela się jednostki na drodze randomizacji. Celem eksperymentu jest pomiar wpływu zmiennej niezależnej na zmienną zależną, tj. porównanie sytuacji, kiedy oddziaływało bodźcem (działania podejmowane w ramach interwencji publicznej), z sytuacją, kiedy tego bodźca nie było. Pretestem w eksperymencie klasycznym jest pomiar zmiennej zależnej. Z kolei posttest to ponowny pomiar zmiennej zależnej, ale w następstwie oddziaływania bodźca. Eksperyment klasyczny polega także na tym, że obserwacji poddawana jest – poza grupą eksperymentalną – także grupa kontrolna, na którą nie oddziałuje się bodźcem. Zaletą eksperymentu klasycznego jest możliwość dokonania oceny, czy określona interwencja powoduje rzeczywiste zmiany społeczne w porównaniu do sytuacji, kiedy interwencje tego rodzaju nie są podejmowane. Należy przy tym zaznaczyć, że tak rozumiany schemat eksperymentalny ma zastosowanie wobec relatywnie prostych, niezbyt skomplikowanych działań (interwencji). Ewaluatorzy na ogół nie uczestniczą w inicjowaniu i wdrażaniu interwencji, a dostęp do jej efektów mają tylko w określonym przedziale czasu, co ogranicza czynnik kontroli nad przebiegiem całego procesu.

Z uwagi na liczne ograniczenia w praktycznej realizacji planu eksperymentalnego ewaluacja przebiega według zbliżonego planu, który jednak nie spełnia wszystkich wymogów. Plan *quasi*-eksperymentalny zasadniczo nie zachowuje randomizacji, gdyż przypisanie do grup kontrolnej i eksperymentalnej dokonuje się

---

<sup>165</sup> Ł. Widła, *Zastosowanie metod ilościowych w ewaluacji*, [w:] *Teoria i praktyka ewaluacji interwencji publicznych*, red. K. Olejniczak, M. Kozak, i B. Lendzion, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008, s. 150–178; K. Zvoch, *Modern Quantitative Methods for Evaluation Science: Recommendations for Essential Methodological Texts*, „American Journal of Evaluation” 2014, t. 35, nr 3, s. 430–440.

<sup>166</sup> R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 216.

<sup>167</sup> B. Turowski i M. Zawicki, *Funkcje, etapy, metody i narzędzia ewaluacji*, op. cit., s. 54.

na podstawie decyzji badacza. Pojawia się jednak w ten sposób możliwość, że obserwowany efekt spowodowany został przez alternatywną przyczynę. Stąd konieczność uwzględniania teorii błędu, w której szacuje się prawdopodobieństwo popełnienia błędu z uwagi na alternatywne do zakładanej hipotezy wyjaśniające efekt.

### 3.2.2. Plan jakościowy

Użyteczność planu wykorzystującego wyłącznie metody jakościowe była dość długo kwestionowana w ewaluacji<sup>168</sup>. Poza wpływem pozytywizmu na nauki społeczne ważną rolę w przypadku ewaluacji odgrywał fakt, iż jedne z pierwszych badań w tym zakresie były przeprowadzone przez E. Thorndike'a za pomocą ilościowych metod eksperymentalnych. Koncentrowano się na ewaluacjach sumatywnych z planem eksperymentalnym lub *quasi*-eksperymentalnym, dokonując oceny skuteczności programu przez ilościowe wskaźniki zebrane w grupach, które brały udział w programie (grupa eksperymentalna), i takich, które były poza programem (grupa kontrolna). Pomijano natomiast kwestię jakości całego procesu, tzn. problem znaczenia, jaki dany proces miał dla uczestników. Zwolennicy metod jakościowych podkreślają, że dla wyjaśnienia przyczyn zachodzenia pewnych zjawisk nie wystarczają dane ilościowe i trzeba odwołać się do danych, które można zebrać tylko metodami jakościowymi<sup>169</sup>. Ponadto, kontekst, w ramach którego zachodzi dany proces, może być bardzo ważny dla wyjaśnienia tego, co zaszło. W krytyce ilościowych metod nauk społecznych zwracano uwagę, że eksperyment losowy w badaniach społecznych jest tworem sztucznym i oddalonym od rzeczywistych sytuacji społecznych<sup>170</sup>. Proces badań w planie jakościowym jest rozumiany jako pętla uczenia się i ma na celu zrozumienie rzeczywistych procesów społecznych w ich naturalnym kontekście. Stąd też metody badawcze powinny minimalizować interwencję w te naturalne sytuacje i procesy. Często metody jakościowe wykorzystuje się jako uzupełnienie dla danych ilościowych, co pozwala zbudować narrację o programie, która da szerszy horyzont do interpretowania danych ilościowych. Jeśli ewaluator nie ma pełnej wiedzy na temat wydarzeń istotnych z punktu widzenia ogólnej oceny programu – wstępne badania jakościowe umożliwią mu zdobycie ogólnej orientacji w przebiegu programu, wskażą istotne problemy, pozwolą określić, które są mniej ważne. Tego typu wstępne badanie jest użyteczne także wówczas, gdy ewaluator porusza się

---

<sup>168</sup> S. Mandes, *Metody jakościowe w ewaluacji*, [w:] *Teoria i praktyka ewaluacji interwencji publicznych*, red. K. Olejniczak, M. Kozak i B. Lendzion, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008, s. 129–149.

<sup>169</sup> L. Goodyear, red., *Qualitative inquiry in evaluation: from theory to practice*, Jossey-Bass, San Francisco, CA 2014, First edition, Research methods for the social sciences 29.

<sup>170</sup> R. Stockmann i W. Meyer, *Functions, methods and concepts in evaluation research*, op. cit., s. 223.

po słabo rozpoznanym terenie, np. gdy program realizowany jest po raz pierwszy lub po raz pierwszy dzieje się to w danym środowisku społecznym. Przeprowadzenie badania jakościowego przed realizacją ankiety wskazane jest również w sytuacji, gdy problem jest słabo poznany lub ewaluatorzy mają niewielkie doświadczenie w jego zakresie. Szybkie badanie jakościowe, zrealizowane przy wykorzystaniu zogniskowanych wywiadów pogłębionych w grupie potencjalnych respondentów, pozwoli uniknąć wielu błędów i rozczarowań. Może być bowiem tak, iż duża liczba braków odpowiedzi na dane pytanie nie wynika z faktu, że respondenci nie mieli zdania na temat, którego dotyczyło, mogli po prostu go nie zrozumieć, uznać za niejasne lub, co gorsza, obraźliwe. Weryfikacja zrozumiałości pytań ankiety jest szczególnie wskazana w przypadku badań na dużych próbach (czyli badań kosztownych) oraz realizowanych w grupach społecznych o niższych kompetencjach kulturowych.

Badania jakościowe są użyteczne także po zakończeniu badań ilościowych. Interpretując dane uzyskane z ankiety, niejednokrotnie napotkać można zaskakujące wyniki. Ewaluator może oczywiście opierać się na domysłach i intuicji, wyjaśniając przyczyny, dla których respondenci odpowiadali tak, a nie inaczej, jednak tego typu argumenty będą narażone na krytykę ze strony odbiorców badań. Jeśli np. w jednej miejscowości dany trener jest oceniany dużo słabiej niż w pozostałych, to można przypuszczać, że doszło tam do jakichś zdarzeń istotnych dla uzyskanej oceny. Tylko badanie jakościowe – wywiad z trenerem oraz uczestnikami szkolenia – może dać odpowiedź na pytanie, co się stało i dlaczego uzyskano takie, a nie inne wyniki.

### 3.2.3. *Plan mieszany*

Plan badań, w którym w jednym procesie badawczym wykorzystuje się różne rodzaje metod w celu uzyskania odpowiedzi na pytanie badawcze, jest jedną z odmian stosowania niejednorodnych metod mnogich (*mutli-method*). Specyfiką metod mieszanych jest integrowanie wyników różnych metod w celu uzyskania jednej, kompleksowej odpowiedzi na pytanie badawcze<sup>171</sup>. Głównym argumentem za stosowaniem takiego planu badań jest wzajemne dopełnianie się wyników badań, np. odpornej zależności przyczynowej w dużej próbie badanej ilościowo oraz zrozumienie mechanizmu tej zależności w małej próbie badanej jakościowo<sup>172</sup>. W praktyce badań ewaluacyjnych jest on wybierany z uwagi na

---

<sup>171</sup> P. Kawalec, *Metody mieszane w kontekście procesu badawczego w naukoznawstwie*, „Zagadnienia Naukoznawstwa” 2014, t. 50, nr 1 (199), s. 3–22. Artykuł zawiera zarys najważniejszych etapów w dziejach metod mieszanych oraz przegląd ich podstawowych odmian.

<sup>172</sup> E.S. Lieberman, *Nested Analysis as a Mixed-Method Strategy for Comparative Research*, op. cit.

pragmatyczne rozstrzygnięcie dylematów teoretycznych<sup>173</sup> oraz konfliktu perspektyw paradygmatycznych w naukach społecznych<sup>174</sup>.

Różnego rodzaju metody mogą być stosowane sekwencyjnie (wyniki zastosowania jednego rodzaju metody są punktem wyjścia w zastosowaniu kolejnego rodzaju metody) lub współbieżnie (równolegle prowadzone badania różnymi metodami z interakcją uzyskiwanych za ich pomocą wyników). Często jest jednak sytuacja, gdy mimo zaangażowania różnego rodzaju metod tylko jedno z nich, np. ilościowe<sup>175</sup>, dominują w określaniu ostatecznego wyniku badania<sup>176</sup>. Stanowi to o elastyczności tego rodzaju badania i daje możliwości dopasowania do danej sytuacji i kontekstu badania ewaluacyjnego<sup>177</sup>. Zastosowanie mieszanego planu badań okazuje się szczególnie pomocne w przypadku kompleksowych badań ewaluacyjnych, które obejmują złożony i długookresowy wpływ programu<sup>178</sup>.

### 3.2.4. Studia przypadków

Studium przypadku stanowi szerszą analizę określonego zjawiska, rozpatrywanego z różnych perspektyw i rozpoznawanego dzięki wykorzystaniu różnych metod gromadzenia danych – tak ilościowych, jak i jakościowych. W badaniach ewaluacyjnych przypadkiem bywa najczęściej kraj, projekt bądź program. Na studium przypadku składa się pojedynczy przypadek lub ich zbiór. Studia przypadku stanowią bogate źródło informacji o ewaluowanych przedsięwzięciach.

---

<sup>173</sup> V.J. Caracelli i J.C. Greene, *Crafting mixed-method evaluation designs*, „New Directions for Evaluation” 1997, t. 1997, nr 74, s. 21; J.C. Greene, V.J. Caracelli i W.F. Graham, *Toward a Conceptual Framework for Mixed-Method Evaluation Designs*, op. cit.; R.B. Johnson i A.J. Onwuegbuzie, *Mixed Methods Research: A Research Paradigm Whose Time Has Come*, „Educational Researcher” 2004, t. 33, nr 7, s. 14–26; D.M. Mertens, *Research and evaluation in education and psychology*, op. cit.

<sup>174</sup> P. Kawalec, *Metody mieszane w kontekście procesu badawczego w naukoznawstwie*, op. cit., s. 9–15; D.M. Mertens i M. Tarsilla, *Mixed Methods Evaluation*, op. cit.

<sup>175</sup> M.M. Mark, *Mixed and Multimethods in Predominantly Quantitative Studies, Especially Experiments and Quasi-Experiments*, [w:] *The Oxford handbook of multimethod and mixed methods research inquiry*, red. S.N. Hesse-Biber i B. Johnson, Oxford University Press, Oxford; New York 2015, s. 21–41; S.N. Hesse-Biber, D. Rodriguez i N.A. Frost, *A Qualitatively Driven Approach to Multimethod and Mixed Methods Research*, [w:] *The Oxford handbook of multimethod and mixed methods research inquiry*, red. S.N. Hesse-Biber i B. Johnson, Oxford University Press, Oxford, New York 2015, s. 3–20.

<sup>176</sup> Pełniejsza typologia planów mieszanych, zob. P. Kawalec, *Metody mieszane w kontekście procesu badawczego w naukoznawstwie*, op. cit., s. 15.

<sup>177</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 559.

<sup>178</sup> W.L. Sanders i S.P. Horn, *Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research*, op. cit.; E. Norgbey, *Debate on the Appropriate Methods for Conducting Impact Evaluation of Programs within the Development Context*, „Journal of MultiDisciplinary Evaluation” 2016, t. 12, nr 27, s. 58–66.

Pozwalają zilustrować wielowątkowe i złożone zależności oraz procesy towarzyszące ewaluowanym przedsięwzięciom. Dlatego też są one uznawane za właściwą metodę do przedstawiania kompleksowych analiz zjawisk wielowymiarowych. Studia przypadków mają szczególne zastosowanie w ewaluacjach tematycznych lub przy tworzeniu zestawień i ewaluacji zbiorów projektów.

Studium przypadku jest poprzedzone określeniem kryteriów selekcji, zakresu danych poddanych analizie, rodzaju wniosków, jakie mają być wyprowadzone w wyniku opracowanego opisu studium przypadku. Studium przypadku przedstawiane w narracyjnej formie znacząco wzbogaca raporty ewaluacyjne, nadając im bardziej czytelny i przystępny charakter<sup>179</sup>.

### 3.3. Główne fazy w badaniu ewaluacyjnym

Proces ewaluacyjny szeroko rozumiany<sup>180</sup> obejmuje jako systematyczne badanie wartości programów i działań społecznych, lecz również implikuje system oddziaływań społecznych pomiędzy interesariuszami programu na poszczególnych etapach. Zalicza się do nich (por. tabela 1.6):

- wstępna diagnozy oczekiwań i potrzeb wszystkich interesariuszy;
- planowanie celów oraz form ewaluacji;
- projektowanie obszarów, pytań kluczowych, kryteriów, metod, próby badawczej, sposobu doboru i przygotowania ewaluatorów, a także kwestii związanych z zarządzaniem procesem ewaluacji;
- zbieranie danych;
- ich analiza;
- raportowanie i popularyzacja wyników ewaluacji.

Jak podkreśla Korporowicz, szczególnie ważne rezultaty kompleksowego procesu ewaluacyjnego ujawniają się dopiero wtedy, gdy przedmiotem ewaluacji jest nie tylko rezultat programu, ale całościowo pojęty okres jego realizacji.

Każda faza procesu ewaluacji ma specyficzną dynamikę społeczną, swoich aktorów oraz realizuje się w konkretnym kontekście, a także generuje właściwe sobie problemy etyczne.

Celem fazy planowania ewaluacji jest decyzja o podjęciu badania ewaluacyjnego, a także określenie jego ogólnych założeń. Decyzja taka na ogół wynika z wymogów prawnych lub z potrzeb administratora danego programu. Jeśli ewaluacja jest zlecona zewnętrznemu zespołowi ekspertów, to konieczne jest przygotowanie specyfikacji ewaluacji (np. w formie SIWZ)<sup>181</sup>.

<sup>179</sup> B. Turowski i M. Zawicki, *Funkcje, etapy, metody i narzędzia ewaluacji*, op. cit.

<sup>180</sup> L. Korporowicz, *Badania ewaluacyjne*, op. cit., s. 71–73.

<sup>181</sup> K. Olejniczak, M. Kozak i B. Lenzion, *Teoria i praktyka ewaluacji interwencji publicznych*, op. cit.

Tabela 1.6

## Etapy procesu ewaluacji

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

Źródło: opracowanie na podstawie K. Olejniczak, M. Kozak i B. Lenzio, *Teoria i praktyka ewaluacji interwencji publicznych*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008, s. 102.

Celami pierwszej fazy badania ewaluacyjnego – strukturalizacji – są: sprecyzowanie pytań badawczych, określenie kryteriów i wskaźników, wybór narzędzi zbierania danych itd. Właściwe zrozumienie programu – jego ograniczeń, specyfiki i kontekstu – warunkuje dalszą analizę programu. Celem fazy obserwacji jest zebranie danych (pierwotne, wtórne) niezbędnych do analizy, np. przez wywiady, badania ankietowe, badania społeczne, wizyty w terenie, studia przypadków.

Faza analizy i interpretacji zebranych danych prowadzi do oszacowania efektów programu, zweryfikowania postawionych hipotez oraz ostatecznej oceny programu i jego efektów, głównie na podstawie metod statystycznych oraz porównawczych. Faza oceny jest natomiast kluczowym elementem ewaluacji,

dokonywanym w odniesieniu do uprzednio sformułowanych pytań ewaluacyjnych i wybranych kryteriów. Stwierdza się w niej faktyczną jakość danej interwencji. Jednocześnie formułowane są wstępne wnioski i rekomendacje z analizy mające służyć poprawie programu. W przypadku oceny i formułowania rekomendacji częste jest włączanie w nią interesariuszy programu.

Ważnym elementem pracy nad raportem końcowym jest prezentacja i konsultacje na temat wstępnej wersji raportu końcowego, dotyczące kwestii merytorycznych (zawartość rekomendacji) i technicznych (forma prezentacji – struktura raportu). Taki dialog prowadzony między wykonawcą a zlecającym (przy zachowaniu bezstronności i niezależności wykonawcy – badacza) znacząco przyczynia się do późniejszej skuteczności raportu – wdrażania rekomendacji, zainteresowania wynikami badań i szerszego odzewu. Momentem kończącym formalnie badanie ewaluacyjne jest przekazanie zlecającemu przez wykonawcę ostatecznej wersji raportu.

## 4. Instytucjonalny wymiar ewaluacji

### 4.1. Instytucjonalizacja ewaluacji a proces uczenia się organizacji

Z uwagi na zmiany społeczne, jakie dokonują się współcześnie, systematyczna ewaluacja stała się nieodzownym elementem funkcjonowania zarówno organizacji, jak i polityk publicznych<sup>182</sup>. Instytucjonalizacja ewaluacji oznacza zdefiniowanie systemu ewaluacji, jego trwałe wdrożenie na wszystkich szczeblach i regularne wykorzystywanie w organizacji<sup>183</sup>. Pozwala to zrationalizować i zoptymalizować procesy wartościowania, podejmowania decyzji i rozwoju w organizacji<sup>184</sup>.

W odniesieniu do polityk publicznych przejawem tworzenia systemu ewaluacji jest powoływanie odrębnych, dedykowanych instytucji, których zadaniem jest systematyczna ewaluacja w celu zbierania i syntezy dowodów na potrzeby cyklu polityk publicznych<sup>185</sup>. Przykładami są the What Works Centres (UK),

---

<sup>182</sup> S.S. Nagel, red., *Handbook of public policy evaluation*, Sage Publications, Thousand Oaks, Calif 2002; S. Højlund, *Evaluation use in the organizational context – changing focus to improve theory*, „Evaluation” 2014, t. 20, nr 1, s. 26–43.

<sup>183</sup> J. Kozłowski, *Ewaluacja instytucji naukowych w Polsce w świetle porównań międzynarodowych i konsultacji*, Departament Strategii MNiSW, Warszawa 2010; S. Højlund, *Evaluation Use in Evaluation Systems: The Case of the European Commission*, Department of Business and Politics. DBP, Frederiksberg 2015.

<sup>184</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 673–674; H. Preskill i S. Boyle, *A Multidisciplinary Model of Evaluation Capacity Building*, „American Journal of Evaluation” 2008, t. 29, nr 4, s. 443–459.

<sup>185</sup> K. Warwick i A. Nolan, *Evaluation of Industrial Policy*, OECD Science, Technology and Industry Policy Papers, 2015, 74, [http://www.oecd-ilibrary.org/science-and-technology/evaluation-of-industrial-policy\\_5jz181jh0j5k-en](http://www.oecd-ilibrary.org/science-and-technology/evaluation-of-industrial-policy_5jz181jh0j5k-en).

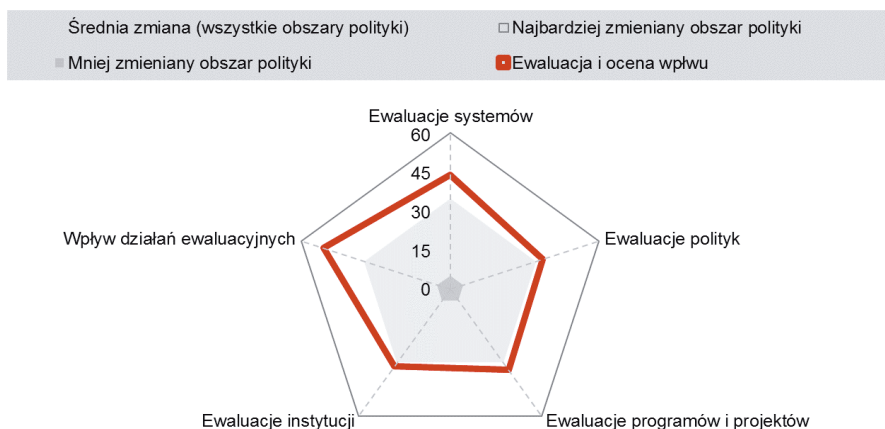


the Washington State Institute for Public Policy in the United States (USA) czy Centrum Ewaluacji i Analiz Polityk Publicznych UJ.

Raport OECD<sup>186</sup> zwraca uwagę na duże zróżnicowanie instytucjonalizacji ewaluacji polityk publicznych. Jako przykłady krajów, które są szczególnie aktywne w dziedzinie ewaluacji polityki naukowej i innowacyjnej, wymienia się USA – także Japonię – gdzie powstały programy w zakresie „Nauki o nauce i polityce innowacyjnej” (SciSIP – skrót od *Science of science and innovation policy*). Analogiczny program w zakresie SciSIP realizowany jest w Norwegii (FORINNPOL). Z uwagi jednak na kompleksowość interwencji i towarzyszącej jej ewaluacji często tego rodzaju inicjatywy mają charakter jednorazowy.

**Rysunek 1.11**

**Ewaluacja i ocena wpływu w innych obszarach zmian polityk STI, 2014–2016.  
Procent polityk, które zostały wprowadzone, zmienione lub odwołane  
w tym okresie**



Źródło: opracowanie na podstawie: OECD, *Evaluation and impact assessment of STI policies*, [w:] *OECD Science, Technology and Innovation Outlook 2016*, OECD, OECD Publishing 2016, [http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016/evaluation-and-impact-assessment-of-sti-policies\\_sti\\_in\\_outlook-2016-13-en](http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016/evaluation-and-impact-assessment-of-sti-policies_sti_in_outlook-2016-13-en), s. 6.

Uwaga: EC/OECD STI Policy survey 2016 ma na celu dwuroczny przegląd głównych zmian w portfolio narodowych polityk oraz zarządzania w STI. Przegląd ten bazuje na opracowaniu pod egidą the OECD Committee for Scientific and Technological Policy (CSTP) mapowania *policy mix* dla innowacji i obejmuje szeroki zakres obszarów polityki. W 2016 roku objęte nim zostały 54 gospodarki, łącznie z krajami OECD, główne gospodarki wschodzące (np. Argentyna, Brazylia, Chiny, Kolumbia, Kostaryka, Egipt, Indie, Indonezja, Malezja, Peru, Rosja, Afryka Południowa i Tajlandia), kraje członkowskie EU spoza OECD

<sup>186</sup> OECD, *Evaluation and impact assessment of STI policies*, op. cit.

oraz Komisja Europejska. Łącznie kraje ujęte w ankiecie STIP objęły ok. 98% globalnego B+R. Odpowiedzi udzielają delegaci CSTP oraz delegaci European Research and Innovation Committee (ERAC) dla krajów EU spoza OECD. Jest to eksperymentalny wskaźnik, zainspirowany analizą demografii firm (np. *churning rate*). Obejmuje on główne inicjatywy w politykach, które zostały wdrożone, odwołane lub znacząco zmienione w latach 2014–2016 jako odsetek wszystkich inicjatyw w politykach, które były aktywne na początku tego okresu. Choć proste zliczanie nie wyjaśnia wielkości i wpływu zmian w politykach, wskaźnik *churning rate* w polityce STI odzwierciedla jej cel oraz działania w określonych obszarach polityki i określonych okresach czasu.

## 4.2. Metaewaluacja

Pojęcie metaewaluacji do literatury wprowadził pod koniec lat 60. Scriven w odpowiedzi na problem niekompatybilności różnych ewaluacji zrealizowanych na zamówienie the Urban Institute<sup>187</sup>. Metaewaluacja jest procesem formatywnej lub sumatywnej ewaluacji innych ewaluacji<sup>188</sup>. Zasadniczą rolę metaewaluacji Scriven porównuje do *peer review*, dzięki czemu przed publikacją wyników uzyskuje się opinię eksperta na jej temat. Istotną rolę odgrywają w niej pytania o trafność przyjętych przez ewaluatora założeń, koncentrujące się głównie na tym, co zostało pominięte w ewaluacji. Dla interesariuszy metaewaluacja służy również ocenie rzetelności, wiarygodności, efektywności kosztowej i aplikowalności ewaluacji. Systematyczne metaewaluacje – przeprowadzane zarówno wewnątrznie przez ewaluującego, jak i zewnątrznie (tabela 1.7) – są również kluczowe dla rozwoju ewaluacji jako dyscypliny przez ocenę jakości oraz ujawnianie ograniczeń dotychczasowych ewaluacji.

Podstawowe rodzaje działań, jakie obejmuje metaewaluacja, są następujące: dobór ekspertów do przeprowadzenia metaewaluacji, identyfikacja i zaangażowanie interesariuszy, uzgodniony dobór standardów dla ocenianej ewaluacji lub systemu, określenie pytań, formalne ustalenia lub negocjowanie kontraktu na przeprowadzenie metaewaluacji, zebranie istotnych dostępnych informacji, zbieranie nowych informacji w razie potrzeby, analiza i synteza zebranych informacji, wyprowadzenie wniosków dotyczących ewaluacji na podstawie odnośnych standardów, zasad i kryteriów, komunikacja wyników w postaci raportu lub prezentacji, w zależności od danej sytuacji – wspomaganie zleceniodawcy lub interesariuszy w interpretacji wyników i ich zastosowaniu<sup>189</sup>.

Proces ewaluacyjny może stanowić temat różnorodnej refleksji, być przedmiotem świadomości zawodowej ewaluatora, ale także obiektem analizy z punktu

<sup>187</sup> M. Scriven, *Meta-Evaluation Revisited*, op. cit., s. iii.

<sup>188</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 631.

<sup>189</sup> Ibidem, s. 643.

widzenia konkretnych nauk społecznych<sup>190</sup>. Naukowa refleksja nad ewaluacją oferuje też instrumenty opisu jej rzeczywistego funkcjonowania w:

- realnych strukturach świata;
- organizacjach i systemach politycznych;
- w relacjach między wszystkimi uczestnikami ewaluacji;
- przez dokonywane wybory metodologiczne, tworzony przez nią język, sposoby upowszechniania wyników;
- towarzyszącą jej perswazję i standardy;
- kodeksy etyczne ewaluatora;
- odniesienia do innych rodzajów działalności badawczej lub analitycznej;
- sposoby własnego uprawomocnienia;
- systemy kształcenia, „inicjacji” i weryfikacji zawodowej.

Tabela 1.7

#### Schemat przekrojowy dla różnych odmian metaewaluacji

Meta-ewaluacja	Formacyjna	Sumatywna
Wewnętrzna	Ewaluator systematycznie ocenia i ulepsza przyjętą definicję grupy odbiorców, pytania, plany, budżet, wynegocjowane uzgodnienia, narzędzia, techniki zbierania i analizy danych, raporty itd.	Do końcowego raportu ewaluator dołącza potwierdzenie, w jakim zakresie ewaluacja spełnia profesjonalne standardy (np. dotyczące jej użyteczności, wykonalności, adekwatności, dokładności, rozliczalności itp.).
Zewnętrzna	Niezależny metaewaluator, powołany przez zleceniodawcę, monitoruje przebieg ewaluacji oraz w razie potrzeby dostarcza ewaluatorowi informacji zwrotnej o mocnych i słabych stronach procesu ewaluacji.	Niezależny metaewaluator, powołany przez zleceniodawcę, przygotowuje sumatywny metaewaluacyjny raport o zakończonej ewaluacji, oceniając jej użyteczność, wykonalność, adekwatność, dokładność, rozliczalność itp. Ten raport powinien być udostępniony wszystkim użytkownikom, do których trafia również sama ewaluacja.

Źródło: opracowanie na podstawie D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 635.

Korporowicz zwraca uwagę na analizę ewaluacji w perspektywie wyróżnionych przez Habermasa form wiedzy. Pierwsza z wymienionych przez Habermasa form wiedzy wynika z empiryczno-analitycznych zainteresowań badaczy. Koncentruje się na precyzyjnym określeniu celów konkretnych działań społecznych lub gospodarczych, a następnie na arsenale realizujących je środków. Wiedza ta ma umożli-

<sup>190</sup> L. Korporowicz, *Badania ewaluacyjne*, op. cit., s. 75–76.

liwić najbardziej racjonalny sposób wykorzystania środków bez większej refleksji nad słusnością wyboru samych celów. Jest to zasadniczo wiedza o charakterze technokratycznym, dlatego cenione są umiejętności kontrolowania, przewidywania, ścisłego pomiaru, a wartościami są skuteczność, efektywność, minimalizacja kosztów. W tak pojętym modelu wiedzy bardzo wyraźnie rysują się zadania, jakie spełniać ma funkcjonalna w stosunku do niego ewaluacja. Zasadniczym kryterium ewaluacji staje się stopień realizacji zamierzonych celów określany poprzez techniki badania mierzalnych osiągnięć i określania racjonalności sposobu ich realizacji. Rola ewaluatora polega na wytwarzaniu sprawdzalnej, możliwie zobiektywizowanej wiedzy o poziomie oczekiwanych osiągnięć poprzez dostęp do wyspecjalizowanych danych i operacyjną umiejętność ich wykorzystania. Ewaluator jest w tej sytuacji dostarczycielem ukierunkowanej i przygotowanej według precyzyjnego wzoru informacji. Sam wzbrania się przed jakąkolwiek świadomą ingerencją w ewaluowaną rzeczywistość, zakładając, że jest w stanie wystarczająco się od niej zdystansować. Na wynik jego studiów nie powinna mieć wpływu jego własna pozycja społeczna, jego własne przekonania i miejsce w analizowanej rzeczywistości, która jest zasadniczo „odrębna”, „zewnętrzna”, a on sam istnieje od niej niezależnie.

Drugi typ opisywanej przez Habermasa wiedzy określany jest przez niego jako historyczno-hermeneutyczny i wyrasta z potrzeby zrekonstruowania sposobu interpretacji i sposobów rozumienia rzeczywistości przez zaangażowanych w nią ludzi. Wiedza tego typu ma dać rozeznanie, jakie znaczenia nadają ludzie swych działaniom i wytworom, aby uchwycić ich świat niejako „od środka” w celu zidentyfikowania rzeczywistych funkcji konkretnych zachowań. Wiele z nich, nie różniąc się od siebie w swych przejawach „zewnętrznych”, wyrażać może zupełnie inne intencje i wartości. Pominięcie tego prostego faktu uchronić ma od rozminięcia się z realnością społeczną, kulturową i międzyludzką. Zadania ewaluacji jawią się w tej perspektywie zupełnie inaczej. Ewaluacja odsłonić ma parametry „sytuacyjne”, „znaczeniowe” i motywacyjne, opisując realność doświadczeń ludzkich. Ewaluator dokonać więc musi koniecznej interpretacji danych, w tym także zachowań indywidualnych i zbiorowych ludzi uczestniczących w ewaluowanych przedsięwzięciach, bazując nie tyle na swoim, ile zrekonstruowanym systemie wartości i znaczeń, jakie wyrażają ewaluowani<sup>191</sup>. Ewaluacja pokazuje wówczas nie uśredniony, ale zróżnicowany system interpretacyjny uczestniczących w programie osób i grup. Celem ewaluacji nie jest wtedy ujawnienie jednej, „zobiektywizowanej” prawdy, ale raczej pokazanie społecznej percepcji analizowanej rzeczywistości, która jawi się zawsze poprzez czyjeś „prawdy”. Ewaluator nie jest zewnętrznym sędzią i nie nakłada na rzeczywistość zewnętrznie przyjętych kryteriów. Analizuje także swą własną rolę w tym, co rekonstruuje, starając się mówić językiem ewaluowanych i odsłaniać ich własną perspektywę wartościowania.

<sup>191</sup> G. Praweńska-Skrzypek, red., *Partycypacja obywatelska w życiu społeczności lokalnej. Stan, bariery, rekomendacje*, Międzynarodowe Centrum Rozwoju Demokracji, Kraków 1996.

Trzeci rodzaj wiedzy ma według Habermasa charakter refleksyjno-krytyczny. Jej zadaniem jest odsłonięcie ukrytych, choć stałych zależności przejawiających się w szeroko rozumianych interakcjach społecznych oraz zainicjowanie związanej z tym „odkryciem” autorefleksji. Ewaluator ujawnia w takim przypadku niewyrażane jawnie założenia programowe działań, niedostrzegane stosunki władzy. Ten typ ewaluacji spełnia funkcje emancypujące jej uczestników poprzez inicjowanie autorefleksji, odsłanianie mało dostrzegalnych elementów rzeczywistości społecznej, które nie zawsze sprzyjają rozwojowi, ujawnianiu możliwości, a przyczyniają się do powstawania mechanizmów marginalizacji osób i grup. Z przytoczonych analiz wynika, iż zupełnie dopuszczalnym i wartościowym celem ewaluacji nie musi być „pomiar” wybranego zjawiska, umiejętności czy procesu. Nie musi być nim nawet najczęściej praktykowane określenie skuteczności lub efektywności ewaluowanego programu. Nie zawsze więc ewaluacja projektowana jest, aby wywrzeć wymierny i doraźny wpływ na procesy decyzyjne. Ujawnienie i zrozumienie ewaluowanej rzeczywistości może być jednym z akceptowanych celów działań ewaluacyjnych. Realizacja tego celu może mieć wielki wpływ na świadomość decydentów, ale także na uczestników ewaluowanego programu i pośrednio okazać się czynnikiem istotnie wpływającym na całą serię decyzji, choć w sposób trudny do precyzyjnego uchwycenia. W tym właśnie sensie ewaluator może odgrywać rolę „medium”, ukazując, „oświełając” lub wręcz wyzwalać istotną społecznie wiedzę, wspomagając procesy zmiany społecznej, inicjując innowacje lub jedynie racjonalizując konkretne procedury. Racjonalizacja ta obejmować może różne stopnie poszerzenia typowej funkcji ewaluacji, jaka uformowała się jeszcze w latach 30. i zmierzała początkowo jedynie do określenia relacji możliwie precyzyjnie zoperacjonalizowanych rezultatów konkretnego programu w stosunku do zakładanych w nim celów.

Role metaewaluacji w rozwijaniu ewaluacji jako dyscypliny badawczej można scharakteryzować następująco:

- stworzenie platform współpracy i komunikacji dla ewaluatorów przez dostarczenie dowodów empirycznych w miejsce stylu retorycznego;
- pomoc w dokumentowaniu i zrozumieniu osiągnięć ewaluacji;
- ułatwienie oceny rzeczywistych możliwości ewaluacji;
- inspirowanie udoskonalenia praktyki ewaluacyjnej przez wskazanie okoliczności, w których ewaluacja zawodzi;
- zwiększenie poczucia profesjonalizmu wśród ewaluatorów przez wykazanie, że ewaluacja jest pełnoprawnym przedmiotem systematycznych badań;
- pomoc w przekształceniu ogólnych i abstrakcyjnie sformułowanych standardów i wskazówek w bardziej empirycznie ukierunkowane informacje o korzyściach ze stosowania różnych praktyk ewaluacyjnych<sup>192</sup>.

---

<sup>192</sup> M. Mark, *Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation*, op. cit., s. 115.

Uwzględniając cel badań ewaluacji oraz wykorzystane w niej metody, można wyodrębnić kilka zasadniczych rodzajów tych studiów. Analiza przyczynowa odnosi się do wykorzystania metod estymacji rezultatów programu lub identyfikacji testowania mechanizmów, które generują te rezultaty. Opis skupia się na metodach pomiaru cech klientów, świadczonych usług lub perspektywy klientów na istotne zmienne reprezentujące rezultaty programu, jednak bez wskazywania nowych kategorii czy ustalania zależności przyczynowych. Wartościowanie odnosi się do ewaluacji, które oceniają wartości związane z danym programem m.in. przez ustalanie zmiennych dotyczących procesu i rezultatów najistotniejszych z perspektywy interesariuszy.

Tabela 1.8

## Rodzaje studiów nad ewaluacją

	Opis	Klasyfikacja	Analiza przyczynowa	Wartościowanie
Cel	Pomiar cech klientów, usług, rezultatów	Kategoryzacja (klientów, usług, itp.)	Szacowanie efektów, określanie mechanizmów	Określanie wartości przyjętych przez interesariuszy i opinię publiczną
Typowe metody	Pomiar, obserwacja	Analiza klastrowa, taksonomie	Eksperymenty, <i>quasi</i> -eksperymenty, studia przypadków	Deliberacja grupowa, ankiety
Ilość (% całości)	95,5%	–	3,5%	1%
Koszt (% całości)	20%	–	70%	10%

Źródło: opracowanie na podstawie: M. Mark, *Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation*, [w:] *Fundamental issues in evaluation*, red. N.L. Smith i P.R. Brandon, Guilford Press, New York 2008, s. 122.

Stufflebeam i in. wskazują szereg kompetencji, które powinien posiadać metaewaluator<sup>193</sup>:

- znajomość alternatywnych zbiorów standardów ewaluacji oraz umiejętność doboru, dostosowywania i aplikowania ich w sposób odpowiedni do danego przypadku ewaluacji;
- kompetencje metodologiczne wraz ze znajomością ewaluandum;
- uczciwość, rzetelność i szacunek dla jednostek oraz społeczeństwa;
- umiejętność negocjowania kontraktu obejmującego pełny zakres metaewaluacji;
- umiejętność komunikacji i współpracy z interesariuszami.

<sup>193</sup> D.L. Stufflebeam i C.L.S. Coryn, *Evaluation theory, models, and applications*, op. cit., s. 637 i nast.

Jednym z ważnym przykładów metaewaluacji jest program realizowany pod auspicjami The TAFTIE Task Force on Benchmarking the Impact, Efficiency and Effectiveness of Innovation Agencies, w którym uczestniczy 11 agencji innowacji oraz instytut badawczy Technopolis<sup>194</sup>. Jednym z planowanych efektów jest stworzenie – przez benchmarking metodologii i technik ewaluacyjnych – standardów zawartych w modelu referencyjnym, które następnie zostaną wykorzystane w ocenie działań podejmowanych przez partycypujące agencje innowacji. Model referencyjny ma obejmować: specyfikację modelu logicznego, definicję pytań badawczych, zarządzanie procesem ewaluacji, wybór metodologii, analizę wyników oraz komunikację publiczną (publikacja). Projekt TAFTIE, na podstawie benchmarkingu najlepszych praktyk z kilku krajów, ma określić kryteria jakości dla instrumentów polityki innowacyjnej. Te standardy są z kolei podstawą oceny ewaluacji wykonanych przez poszczególne partycypujące agencje innowacji.

Często popełnianym błędem jest mylenie metaewaluacji z metaanalizą, choć dotyczą one całkowicie odmiennych zagadnień. Metaewaluacja jest oceną jakości i zalet danej ewaluacji, systemu ewaluacyjnego lub narzędzi ewaluacji. Metaanaliza natomiast jest formą ilościowej syntezy wyników badań, dotyczących jednego pytania badawczego, np. efektów danej interwencji, która była badana w kilku eksperymentach.

### 4.3. Praktyka oparta na dowodach naukowych

Zapoczątkowana w medycynie idea systematycznego wykorzystania dowodów naukowych stała się popularna od połowy lat 90. także w politykach, zarządzaniu oraz praktyce, które określa się jako „oparte na dowodach naukowych” (*evidence-based*)<sup>195</sup>. Idea ugruntowania decyzji dotyczących praktyki w nauce okazała się uniwersalna i możliwa do zastosowania zarówno w skali makro (poziom społeczeństw i obszarów życia zbiorowego), mezo (strategia i struktura danej organizacji), jak i mikro (jednostka)<sup>196</sup>.

Jak wskazano w podpunkcie 1.1, za najlepszy rodzaj dowodów empirycznych uznaje się eksperymenty z losowo dobranymi grupami eksperymentalną i kontrolną, przeprowadzone na reprezentatywnych próbach, których wyniki określone zostały w analizie syntetycznej i metaanalizie<sup>197</sup>. Analiza syntetyczna i metaanaliza oparte

<sup>194</sup> K. Warwick i A. Nolan, *Evaluation of Industrial Policy*, op. cit.

<sup>195</sup> R. Szarfenberg, *Dowody naukowe jako podstawa polityki społecznej, zarządzania społecznego i pracy socjalnej*, „Problemy Polityki Społecznej” 2011, nr 15, s. 13–28; M. Mark, *Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation*, op. cit.; P. Davies, *The State of Evidence-Based Policy Evaluation and its Role in Policy Formation*, op. cit.

<sup>196</sup> F. Adam i P. Humphreys, red., *Encyclopedia of decision making and decision support technologies*, Information Science Reference, Hershey, PA 2008.

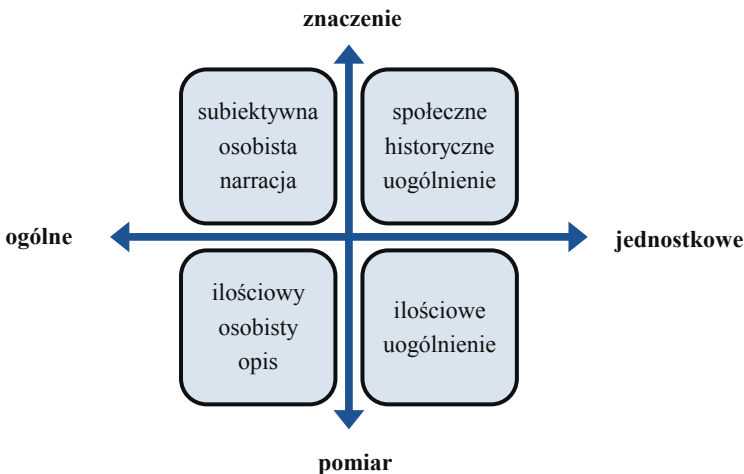
<sup>197</sup> T. Dehue, *Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design*, „The American Journal of

są na założeniu, że wyniki wielu eksperymentów sprawdzających daną interwencję pozwalają ustalić odporną zależność przyczynową, którą można ekstrapolować na większą skalę. W dalszej kolejności wykorzystuje się badania *quasi*-eksperymentalne, przedeksperymentalne (tylko po interwencji, albo przed i po interwencji, bez grup porównawczych), sondaże oraz badania jakościowe<sup>198</sup> bądź stosuje się mieszany plan badań z wykorzystaniem różnych rodzajów dowodów<sup>199</sup>.

Szarfenberg różnicę między rodzajami dowodów empirycznych w naukach społecznych przedstawia na dwuwymiarowym wykresie. Uwzględnia on z jednej strony dylemat podejścia idiograficznego (opis jednostkowy) vs nomotetycznego (prawo naukowe), a z drugiej spór interpretywizmu (znaczenie) vs naturalizm (pomiar). Typologię prezentuje rysunek 1.12.

Rysunek 1.12

### Rodzaje dowodów z uwagi na stopień ogólności i rodzaj metod



Źródło: opracowanie na podstawie: R. Szarfenberg, *Dowody naukowe jako podstawa polityki społecznej, zarządzania społecznego i pracy socjalnej*, „Problemy Polityki Społecznej” 2011, nr 15, s. 13–28.

Polityki oparte na dowodach na ogół realizowane są według następujących etapów:

1. Przekształcenie potrzeb informacyjnych wynikających z istotnych dla interwencji publicznej decyzji w pytania, na które można odpowiedzieć, odwołując się do wyników badań empirycznych.

Psychology” 2001, t. 114, nr 2, s. 283–302; R. Rogers-Dillon, *The welfare experiments: politics and policy evaluation*, Stanford Law and Politics, Stanford, CA 2004.

<sup>198</sup> R. Szarfenberg, *Dowody naukowe jako podstawa polityki społecznej, zarządzania społecznego i pracy socjalnej*, op. cit.

<sup>199</sup> S.N. Hesse-Biber, *Mixed methods research: merging theory with practice*, Guilford Press, New York 2010; S.N. Hesse-Biber, D. Rodriguez i N.A. Frost, *A Qualitatively Driven Approach to Multimethod and Mixed Methods Research*, op. cit.



2. Zebranie najlepszych dostępnych wyników badań naukowych istotnych dla uzyskania odpowiedzi na postawione pytania.
3. Krytyczna ocena zebranych dowodów pod względem ich rzetelności, poziomu stwierdzonego wpływu interwencji oraz użyteczności.
4. Zastosowanie wyników oceny zebranych dowodów w decyzjach dotyczących interwencji z uwzględnieniem kontekstu oraz wartości i preferencji klienta.
5. Ewaluacja skuteczności i efektywności realizacji kroków wcześniejszych etapów w celu ich udoskonalenia w przyszłości.

#### 4.4. Komunikowanie wyników ewaluacji

Rola komunikowania wyników ewaluacji wynika z wartości dodanej, jaką może ona tworzyć w procesie podejmowania decyzji oraz w wymiarze instytucjonalnym<sup>200</sup>. Z uwagi na wagę tego działania istotne jest tworzenie planu komunikacji na możliwie wczesnym etapie projektowania ewaluacji. Powinien on uwzględniać konieczność zrównoważenia dwóch przeciwstawnych kierunków działania: z jednej strony zapewnienie możliwie jak najszerszego odbioru wyników ewaluacji, a z drugiej – zachowanie atmosfery zaufania, która warunkuje interakcje między różnymi grupami interesariuszy ewaluacji. Dlatego istotne jest uwzględnienie zróżnicowania kanałów komunikacyjnych w zależności od grupy odbiorców. Syntetyczne zestawienie prezentuje tabela 1.9.

Tabela 1.9

Zróżnicowanie kanałów komunikacji wyników ewaluacji

	Wywiad, spotkanie	Wstępna wersja raportu	Poufna notatka	Raport publiczny	Synteza	Prezentacja	Broszura artykuł	Konferencja prasowa
Zamawiający			+	+	+	+		
Zarząd projektu, menadżerowie, administracja	+	+		+		+		
Instytucje europejskie, krajowe, regionalne				+	+		+	
Obywatele, dziennikarze					+		+	+

Źródło: opracowanie na podstawie: European Commission, *Evaluation design and management*, op. cit.

<sup>200</sup> European Commission, *Evaluation design and management*, op. cit.

Badanie faktycznie stosowanych metod w komunikacji wskazuje, że zarówno w wewnętrznej, jak i zewnętrznej ewaluacji dominują (w nawiasach podano częstości dla ewaluacji wewnętrznej / zewnętrznej według następującej skali: 0 – nigdy, 1 – rzadko, < 26% przypadków, 2 – okazjonalnie 26–50%, 3 – często, 51–75%, 4 – rutynowo, > 75%):

- raport końcowy (3,5/3,8);
- podsumowanie typu *executive summary* (3,3/3,6);
- ustna prezentacja dla zleceniodawcy (2,9/3,2),
- raport pośredni (2,3/2,7);
- prezentacja konferencyjna (1,7/2,2);
- artykuł w czasopiśmie naukowym (1,1/2,2);
- wewnętrzny newsletter (1,2/1,6);
- notatka prasowa (0,8/1,0);
- zewnętrzny newsletter (0,9/0,8);
- konferencja prasowa (0,5/0,4);
- wystąpienie telewizyjne lub radiowe (0,3/0,5);
- psychodrama (0,1/ < 0,1)<sup>201</sup>.

W ocenie badanych szczególnie efektywne były krótkie formy komunikacji, jak np. streszczenia typu *executive summary*, wykorzystanie elementów graficznych (wykresy, tabele) oraz jasny język, a także ciągła komunikacja od samego początku realizacji ewaluacji z angażowaniem interesariuszy programu. Udana komunikacja wspomaga:

- zrozumienie idei ewaluacji oraz docenienie jej wyników;
- zrozumienie programu ewaluowanego;
- podejmowanie działań w odpowiedzi na wyniki ewaluacji.

Raport końcowy powinien spełniać wymogi naukowe, określać przesłanki ewaluacji, jej cel, główne pytania, plan i metody badań, wskazywać źródła informacji oraz wyjaśniać analizę i interpretację danych. Rekomendacje powinny być precyzyjnie osadzone we wnioskach i treści raportu. W przypadku wniosków o negatywnym wydźwięku i krytycznych rekomendacjach raport często zawiera również plan naprawczy, który wspiera proces uczenia się organizacji będącej zleceniodawcą ewaluacji. Poszczególne typy ewaluacji mają właściwy sobie sposób rozpowszechniania i wykorzystania: ewaluacja wstępna jest wykorzystywana do korekty planu interwencji, średniookresowa i bieżąca – są używane do zmian w zarządzaniu programem, do poprawy programu w trakcie jego trwania i ewentualnie do projektowania nowego, jeśli ewaluacje są publikowane pod koniec jego trwania. Ewaluacje końcowe docierają do szerokiego grona interesariuszy i mogą odgrywać rolę społecznego uprawomocnienia programu lub interwencji.

---

<sup>201</sup> R.T. Torres, H.S. Preskill i M.E. Piontek, *Communicating and Reporting: Practices and Concerns of Internal and External Evaluators*, „American Journal of Evaluation” 1997, t. 18, nr 2, s. 105–125.

Ważnymi elementami przy rozpowszechnianiu wyników ewaluacji są:

- jasna strategia rozpowszechniania wyników ewaluacji – określenie grup docelowych i sposobów komunikowania wniosków (powinno być to zrobione już na etapie projektowania ewaluacji, zaś raport z badania powinien być dostosowany właśnie do tych wymagań);
- szeroki dostęp do wyników badania – publikowanie wyników na stronach internetowych instytucji zarządzającej programem, informacja dla beneficjentów, zorganizowanie dyskusji wokół wyników badania;
- zaprojektowanie systemu informacji dla kontynuacji badania (*follow up* wdrażania rekomendacji i śledzenie ich efektów)<sup>202</sup>.

### Prace cytowane

- Adam F. i Humphreys P., red., *Encyclopedia of decision making and decision support technologies*, Information Science Reference, Hershey, PA 2008.
- Agranoff R., McGuire M. i Silvia C., *Governance, networks and intergovernmental systems*, [w:] *Routledge handbook of public policy*, red. E. Araral, Routledge, London 2013, s. 361–73.
- Alkin M.C., red., *Evaluation roots: tracing theorists' views and influences*, Sage Publications, Thousand Oaks, CA 2004.
- Alkin M.C., *When is a theory a theory? A case example*, „Evaluation and Program Planning” 2017, t. 63, s. 141–142.
- Barnett C. i Camfield L., *Ethics in evaluation*, „Journal of Development Effectiveness” 2016, t. 8, nr 4, s. 528–534.
- Besharov D.J., Baehler K.J. i Klerman J.A., red., *Improving public services: international experiences in using evaluation tools to measure program performance*, Oxford University Press, New York 2017.
- Bjørnholt B. i Larsen F., *The politics of performance measurement: 'Evaluation use as mediator for politics'*, „Evaluation” 2014, t. 20, nr 4, s. 400–411.
- Borenstein M., red., *Introduction to meta-analysis*, John Wiley & Sons, Chichester, U.K. 2009.
- Borrás S. i Edler J., *The Governance of Socio-Technical Systems*, Edward Elgar Publishing 2014, <http://www.elgaronline.com/view/9781784710187.xml>.
- Borrás S., i Højlund S., *Evaluation and policy learning: The learners' perspective: Evaluation and policy learning*, „European Journal of Political Research” 2015, t. 54, nr 1, s. 99–120.
- Boruch R.F., *Randomized field trials in education*, [w:] *International handbook of educational evaluation*, red. T. Kellaghan i D.L. Stufflebeam, Kluwer Academic Publishers, Norwell, MA 2003, s. 107–124.
- Bozeman B. i Link A.N., *Toward an assessment of impacts from US technology and innovation policies*, „Science and Public Policy” 2015, t. 42, nr 3, s. 369–376.
- Brinkerhoff R.O., *The success case method: find out quickly what's working and what's not*, Berrett-Koehler, San Francisco, CA 2003.

<sup>202</sup> K. Olejniczak, *Mechanizmy wykorzystania ewaluacji*, op. cit., s. 101n.

- Campbell B. i McGrath A., *Where the rubber hits the road: The development of usable middle-range evaluation theory*, [w:] *Social psychology and evaluation*, red. M.M. Mark, S.I. Donaldson, i B. Campbell, Guilford Press, New York 2011, s. 346–372.
- Capano G., Howlett M. i Ramesh M., *Varieties of Governance: Dynamics, Strategies, Capacities*, 2015.
- Caracelli V.J. i Greene J.C., *Crafting mixed-method evaluation designs*, „New Directions for Evaluation” 1997, nr 74, s. 19–32.
- Chen H., *Applying mixed methods under the framework of theory-driven evaluations*, „New Directions for Evaluation” 1997, nr 74, s. 61–72.
- Chen H.T. i Turner N.C., *Formal Theory Versus Stakeholder Theory: New Insights From a Tobacco-Focused Prevention Program Evaluation*, „American Journal of Evaluation” 2012, t. 33, nr 3, s. 395–413.
- Christie C.A. i Fleischer D.N., *Insight Into Evaluation Practice: A Content Analysis of Designs and Methods Used in Evaluation Studies Published in North American Evaluation-Focused Journals*, „American Journal of Evaluation” 2010, t. 31, nr 3, s. 326–346.
- Coryn C.L. i Hattie J.A., *The transdisciplinary model of evaluation*, „Journal of MultiDisciplinary Evaluation” 2007, t. 3, nr 4, s. 107–114.
- Coryn C.L.S., Noakes L.A., Westine C.D. i Schröter D.C., *A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009*, „American Journal of Evaluation” 2011, t. 32, nr 2, s. 199–226.
- Coryn C.L.S., Wilson L.N., Westine C.D., Hobson K.A., Ozeki S., Fiekowsky E.L., Greenman G.D. i Schröter D.C., *A Decade of Research on Evaluation: A Systematic Review of Research on Evaluation Published Between 2005 and 2014*, „American Journal of Evaluation” 2017, s. 1–19.
- Cousins J.B. i Earl L.M., *The Case for Participatory Evaluation*, „Educational Evaluation and Policy Analysis” 1992, t. 14, nr 4, s. 397–418.
- Cronbach L.J., *Course Improvement Through Evaluation*, „Teachers College Record” 1963, t. 64, nr 8, s. 672–683.
- Cronbach L.J. i Shapiro K., *Designing evaluations of educational and social programs*, Jossey-Bass, San Francisco 1982, A Joint publication in the Jossey-Bass series in social and behavioral science & in higher education.
- Cronbach L.J. i Snow R.E., *Aptitudes and instructional methods: a handbook for research on interactions*, Irvington Publishers : distributed by Halsted Press, New York 1977.
- Dahler-Larsen P., *From Programme Theory to Constructivism: On Tragic, Magic and Competing Programmes*, „Evaluation” 2001, t. 7, nr 3, s. 331–349.
- Dahler-Larsen P., *The Changing Role of Evaluation in a Changing Society*, [w:] *The Future of Evaluation*, red. R. Stockmann i W. Meyer, Palgrave Macmillan UK, London 2016, s. 24–35.
- Davies P., *The State of Evidence-Based Policy Evaluation and its Role in Policy Formation*, „National Institute Economic Review” 2012, t. 219, nr 1, s. R41–R52.
- Dehue T., *Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design*, „The American Journal of Psychology” 2001, t. 114, nr 2, s. 283–302.

- Derrick G.E. i Samuel G.N., *The Evaluation Scale: Exploring Decisions About Societal Impact in Peer Review Panels*, „Minerva” 2016, t. 54, nr 1, s. 75–97.
- Donaldson S.I. i Lipsey M.W., *Roles for Theory in Contemporary Evaluation Practice: Developing Practical Knowledge*, [w:] *The SAGE Handbook of Evaluation*, red. I. Shaw, J. Greene i M. Mark, SAGE Publications, London 2006, s. 57–75.
- Döring N. i Bortz J., red., *Forschungsmethoden und Evaluation in den Sozial – und Humanwissenschaften*, Springer, Berlin Heidelberg 2016, 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage, Springer-Lehrbuch.
- European Commission, *Evaluation design and management*, Office for Official Publications of the European Communities, Luxembourg 1999.
- European Commission, *Evaluation matters: The Evaluation Policy for European Union Development Co-operation*, Publications Office of the European Union, Luxembourg 2014.
- Fetterman D.M., *Foundations of empowerment evaluation*, SAGE Publications, Thousand Oaks, CA 2001.
- Fitzpatrick J.L., Sanders J.R. i Worthen B.R., *Program evaluation: alternative approaches and practical guidelines*, Pearson Education, Upper Saddle River, N.J 2011, 4th ed.
- Fox C., Grimm R. i Caldeira R., *An introduction to evaluation*, SAGE Publications, London 2016.
- Fuchs E.R.H., *Rethinking the role of the state in technology development: DARPA and the case for embedded network governance*, „Research Policy” 2010, t. 39, nr 9, s. 1133–1147.
- Gates E. i Dyson L., *Implications of the Changing Conversation About Causality for Evaluators*, „American Journal of Evaluation” 2017, t. 38, nr 1, s. 29–46.
- Gates E.F., *Learning from seasoned evaluators: Implications of systems approaches for evaluation practice*, „Evaluation” 2017, t. 23, nr 2, s. 152–171.
- Goodyear L., red., *Qualitative inquiry in evaluation: from theory to practice*, Jossey-Bass, San Francisco, CA 2014, First edition, Research methods for the social sciences 29.
- Goodyear L., *Resources to Support Ethical Practice in Evaluation: An Interview With the Director of the National Center for Research and Professional Ethics*, „American Journal of Evaluation” 2012, t. 33, nr 3, s. 459–461.
- Górnjak J., *Efekt przyczynowy w ewaluacji*, [w:] *Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych*, red. A. Haber i M. Szalaj, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2009, s. 189–200.
- Greene J.C., *Logic and evaluation theory*, „Evaluation and Program Planning” 2013, t. 38, s. 71–73.
- Greene J.C., Caracelli V.J. i Graham W.F., *Toward a Conceptual Framework for Mixed-Method Evaluation Designs*, „Educational Evaluation and Policy Analysis” 1989, t. 11, nr 3, s. 255–274.
- Guba E.G. i Lincoln Y.S., *Fourth generation evaluation*, Sage Publications, Newbury Park, Calif 1989.
- Hajer M.A. i Wagenaar H., red., *Deliberative policy analysis: understanding governance in the network society*, Cambridge University Press, Cambridge, UK; New York, USA 2003.
- Hall J.N., Ahn J. i Greene J.C., *Values Engagement in Evaluation: Ideas, Illustrations, and Implications*, „American Journal of Evaluation” 2012, t. 33, nr 2, s. 195–207.

- Hansen H.F., *Systemic evaluation governance. New logics in the Development of organizational fields*, „Scandinavian Journal of Public Administration” 2013, t. 16, nr 3, s. 47–64.
- Hansen M., Alkin M.C. i Wallace T.L., *Depicting the logic of three evaluation theories*, „Evaluation and Program Planning” 2013, t. 38, s. 34–43.
- Hesse-Biber S.N., *Mixed methods research: merging theory with practice*, Guilford Press, New York 2010.
- Hesse-Biber S.N., Rodriguez D. i Frost N.A., *A Qualitatively Driven Approach to Multimethod and Mixed Methods Research*, [w:] *The Oxford handbook of multimethod and mixed methods research inquiry*, red. S.N. Hesse-Biber i B. Johnson, Oxford University Press, Oxford, New York 2015, s. 3–20.
- Højlund S., *Evaluation Use in Evaluation Systems: The Case of the European Commission*, Department of Business and Politics. DBP, Frederiksberg 2015.
- Højlund S., *Evaluation use in the organizational context – changing focus to improve theory*, „Evaluation” 2014, t. 20, nr 1, s. 26–43.
- Hoskin K., *The Examination, Disciplinary Power and Rational Schooling*, „History of Education” 1979, t. 8, nr 2, s. 21–135.
- House E.R., *Assumptions Underlying Evaluation Models*, „Educational Researcher” 1978, t. 7, nr 3, s. 4–12.
- House E.R., *Deliberative democratic evaluation*, [w:] *Encyclopedia of Evaluation*, red. S. Mathison, Sage Publications, Thousand Oaks, CA 2005, s. 104–108.
- Hughes M. i Kushner S., *Accreditation*, [w:] *Encyclopedia of Evaluation*, red. S. Mathison, Sage Publications, Thousand Oaks, CA 2005, s. 4–7.
- Jałocha B., Krane H.P., Ekambaram A. i Praweńska-Skrzypek G., *Key Competences of Public Sector Project Managers*, „Procedia – Social and Behavioral Sciences” 2014, t. 119, s. 247–256.
- Johnson R.B. i Onwuegbuzie A.J., *Mixed Methods Research: A Research Paradigm Whose Time Has Come*, „Educational Researcher” 2004, t. 33, nr 7, s. 14–26.
- Joint Committee on Standards for Educational Evaluation, *The program evaluation standards: how to assess evaluations of educational programs*, Sage Publications, Thousand Oaks 1994, 2. wyd.
- Kawalec P., *Akognitywizm*, [w:] *Powszechna encyklopedia filozofii*, red. A. Maryniarczyk, Polskie Towarzystwo Tomasza z Akwinu, Lublin 2000, t. I, s. 122–128.
- Kawalec P., *Metody mieszane w kontekście procesu badawczego w naukoznawstwie*, „Zagadnienia Naukoznawstwa” 2014, t. 50, nr 1 (199), s. 3–22.
- Kawalec P., *Problemy koncepcji polityki innowacyjnej jako „przedsiębiorczego państwa”*, „Zagadnienia Naukoznawstwa” 2017, t. 53, nr 2(212), s. 185–206.
- Kawalec P., *Przyczyna i wyjaśnianie: studium z filozofii i metodologii nauk*, Wydawnictwo KUL, Lublin 2006.
- Korporowicz L., *Badania ewaluacyjne*, [w:] *Zarządzanie badaniami naukowymi i pracami rozwojowymi w jednostkach naukowych*, red. P. Kawalec i P. Lipski, Wydawnictwo LBS, Lublin 2008, t. 2, s. 65–76.
- Kotarba B., *Znaczenie kontekstu sytuacyjnego w lokalnej polityce publicznej*, „Wrocławskie Studia Politologiczne” 2016, t. 21, s. 119–130.
- Kozłowski J., *Ewaluacja instytucji naukowych w Polsce w świetle porównań międzynarodowych i konsultacji*, Departament Strategii MNiSW, Warszawa 2010.

- Lawrenz F. i Huffman D., *Methodological pluralism: The gold standard of STEM evaluation*, „New Directions for Evaluation” 2006, t. 2006, nr 109, s. 19–34.
- Leeuw F.L. i Donaldson S.I., *Theory in evaluation: Reducing confusion and encouraging debate*, „Evaluation” 2015, t. 21, nr 4, s. 467–480.
- Levin H.M. i McEwan P.J., *Cost-effectiveness analysis: methods and applications*, Sage Publications, Thousand Oaks, CA 2001, 2. wyd.
- Lieberman E.S., *Nested Analysis as a Mixed-Method Strategy for Comparative Research*, „American Political Science Review” 2005, t. 99, nr 3, s. 435–452.
- Madaus G.F. i Stufflebeam D.L., *Educational Evaluation: Classic Works of Ralph W. Tyler*, Kluwer Academic Publishers, Boston 1989.
- Madaus G.F. i Stufflebeam D.L., *Program evaluation: a historical overview*, [w:] *Evaluation models: viewpoints on educational and human services evaluation*, red. D.L. Stufflebeam, G.F. Madaus i T. Kellaghan, Kluwer Academic Publishers, Boston 2000, 2. wyd., s. 3–18.
- Mandes S., *Metody jakościowe w ewaluacji*, [w:] *Teoria i praktyka ewaluacji interwencji publicznych*, red. K. Olejniczak, M. Kozak i B. Lendzion, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008, s. 129–149.
- Mark M., *Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation*, [w:] *Fundamental issues in evaluation*, red. N.L. Smith i P.R. Brandon, Guilford Press, New York 2008, s. 111–134.
- Mark M.M., *Mixed and Multimethods in Predominantly Quantitative Studies, Especially Experiments and Quasi-Experiments*, [w:] *The Oxford handbook of multimethod and mixed methods research inquiry*, red. S.N. Hesse-Biber i B. Johnson, Oxford University Press, Oxford; New York 2015, s. 21–41.
- Mark M.M., Henry G.T. i Julnes G., *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs*, Jossey-Bass, San Francisco 2000.
- Mertens D.M., *Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods*, Sage, Los Angeles 2010, 3. wyd.
- Mertens D.M., *The Role of Evaluation in Social Change: Perspectives on the Future of Evaluation from the USA*, [w:] *The Future of Evaluation*, red. R. Stockmann i W. Meyer, Palgrave Macmillan UK, London 2016, s. 214–227.
- Mertens D.M. i Tarsilla M., *Mixed Methods Evaluation*, [w:] *The Oxford handbook of multimethod and mixed methods research inquiry*, red. S.N. Hesse-Biber i B. Johnson, Oxford University Press, Oxford, New York 2015, s. 426–446.
- Miller R.L., *Developing Standards for Empirical Examinations of Evaluation Theory*, „American Journal of Evaluation” 2010, t. 31, nr 3, s. 390–399.
- Miller R.L., King J.A., Mark M.M. i Caracelli V., *The Oral History of Evaluation: The Professional Development of Robert Stake*, „American Journal of Evaluation” 2016, t. 37, nr 2, s. 287–294.
- Millett R., *W.K. Kellogg Foundation cluster evaluation model of evolving practices*, W.K. Kellogg Foundation, Battle Creek, MI 1995.
- Morris M., *Evaluation ethics for best practice: cases and commentaries*, Guilford Press, New York 2008.
- Nagel S.S., red., *Handbook of public policy evaluation*, Sage Publications, Thousand Oaks, Calif 2002.

- Norgbey E., *Debate on the Appropriate Methods for Conducting Impact Evaluation of Programs within the Development Context*, „Journal of MultiDisciplinary Evaluation” 2016, t. 12, nr 27, s. 58–66.
- Ochsner M., Hug S.E. i Daniel H.-D., red., *Research Assessment in the Humanities*, Springer International Publishing, Cham 2016.
- OECD, *Evaluation and impact assessment of STI policies*, [w:] *OECD Science, Technology and Innovation Outlook 2016*, OECD, OECD Publishing 2016, [http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016/evaluation-and-impact-assessment-of-sti-policies\\_sti\\_in\\_outlook-2016-13-en](http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016/evaluation-and-impact-assessment-of-sti-policies_sti_in_outlook-2016-13-en).
- OECD, *Glossary of Key Terms in Evaluation and Results-Based Management*, OECD-DAC, Paris 2002.
- OECD, *Governance and implementation of innovation policies*, [w:] *The Innovation Imperative*, OECD Publishing 2015, s. 237–267.
- OECD, *PISA 2015 Assessment and Analytical Framework*, OECD Publishing 2017, PISA, [http://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework\\_9789264281820-en](http://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework_9789264281820-en).
- OECD, *Trends in evaluation and assessment*, [w:] *Synergies for Better Learning*, OECD, OECD Publishing 2013, s. 35–56, [http://www.oecd-ilibrary.org/education/synergies-for-better-learning/trends-in-evaluation-and-assessment\\_9789264190658-5-en](http://www.oecd-ilibrary.org/education/synergies-for-better-learning/trends-in-evaluation-and-assessment_9789264190658-5-en).
- Okoń-Horodyńska E., *Narodowy system innowacji w Polsce*, Wydawnictwo Uczelniane Akademii Ekonomicznej, Katowice 1998.
- Okoń-Horodyńska E.O., *Smart specialisation as a lever towards innovation: the example of Malopolska region in Poland*, „International Journal of Transitions and Innovation Systems” 2014, t. 3, nr 3, s. 249–269.
- Olejniczak K., *Mechanizmy wykorzystania ewaluacji: studium ewaluacji średniookresowych INTERREG III*, Wydawnictwo Naukowe SCHOLAR, Warszawa 2008.
- Olejniczak K., Kozak M. i Lenzion B., red., *Teoria i praktyka ewaluacji interwencji publicznych*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008.
- Owen J.M., *Program evaluation: forms and approaches*, Guilford Press, New York 2007, 3rd ed.
- Parlett M. i Hamilton D., *Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes*, Centre for Research in Educational Sciences, University of Edinburgh, Edinburgh 1972.
- Patton M.Q., McKegg K. i Wehipeihana N., red., *Developmental evaluation exemplars: principles in practice*, The Guilford Press, New York, NY 2016.
- Pawson R., *Evidence-based policy: a realist perspective*, SAGE, London 2006.
- Perez C., Vellinga P., de Buck P., Faulstich M., Giroaurd N., Gregori G., Hudson C., et al., *Changing gear in R&I: Green growth for jobs and prosperity in the EU. Report of the European Commission expert group “R&I policy framework for green growth & jobs”*, European Commission, Strasbourg 2016.
- Pollitt C., *The essential public manager*, Open University Press, Maidenhead 2003.
- Pollitt C., i Bouckaert G., *Public management reform a comparative analysis*, Oxford University Press, Oxford; New York 2000.
- Pollitt C., Thiel S. van i Homburg V., *New public management in Europe: adaptation and alternatives*, Palgrave Macmillan, Basingstoke [England], New York 2007.
- Popham W.J., *Educational evaluation*, Prentice-Hall, Englewood Cliffs, N.J 1975.



- Prawelska-Skrzypek G., red., *Partycypacja obywatelska w życiu społeczności lokalnej. Stan, bariery, rekomendacje*, Międzynarodowe Centrum Rozwoju Demokracji, Kraków 1996.
- Prawelska-Skrzypek G. i Jałocha B., *Projektyzacja sektora publicznego w Polsce–implikacje dla organizacji samorządu terytorialnego*, „Zarządzanie Publiczne” 2014, nr 3 (27), s. 273–284.
- Preskill H. i Boyle S., *A Multidisciplinary Model of Evaluation Capacity Building*, „American Journal of Evaluation” 2008, t. 29, nr 4, s. 443–459.
- Reale E., Avramov D., Canhial K., Donovan C., Flecha R., Holm P., Larkin C. et al., *A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research*, „Research Evaluation” 2017.
- Riggin L.J.C., *Advances in mixed-method evaluation: A synthesis and comment*, „New Directions for Evaluation” 1997, nr 74, s. 87–94.
- Rogers-Dillon R., *The welfare experiments: politics and policy evaluation*, Stanford Law and Politics, Stanford, CA 2004.
- Rossi P.H., Freeman H.E., i Lipsey M.W., *Evaluation: a systematic approach*, Sage Publications, Thousand Oaks, Calif 1999, 6th ed.
- Salem Press, *Research & evaluation methods*, 2014.
- Sanders W.L. i Horn S.P., *Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research*, „Journal of Personnel Evaluation in Education” 1998, t. 12, nr 3, s. 247–256.
- Scriven M., *Evaluation as a discipline*, „Studies in Educational Evaluation” 1994, t. 20, nr 1, s. 147–166.
- Scriven M., *Evaluation thesaurus*, Sage Publications, Newbury Park, CA 1991, 4. wyd.
- Scriven M., red., *Hard-Won Lessons in Program Evaluation*, „New Directions for Program Evaluation” 1993, t. 58, s. 1–107.
- Scriven M., *Meta-Evaluation Revisited*, „Journal of MultiDisciplinary Evaluation” 2009, t. 6, nr 11, s. iii–viii.
- Scriven M., *The Methodology of Evaluation*, [w:] *Perspectives of curriculum evaluation*, red. R.W. Tyler, R.M. Gagne i M. Scriven, Rand-McNally, Chicago 1967, s. 39–83.
- Scriven M., *The nature of evaluation*, „New Directions for Program Evaluation” 1993, t. 58, s. 5–48.
- Scriven M. i Coryn C.L.S., *The logic of research evaluation*, „New Directions for Evaluation” 2008, nr 118, s. 89–105.
- Shadish W.R., Cook T.D. i Campbell D.T., *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin, Boston 2001.
- Shadish W.R., Cook T.D. i Leviton L.C., red., *Foundations of program evaluation: theories of practice*, Sage Publications, Newbury Park 2000, Reprinted.
- Shaw I.F., Greene J. i Mark M., red., *Handbook of evaluation: policies, programs and practices*, Sage Publications, London 2013.
- Simon H.A., *Administrative behavior: a study of decision-making processes in administrative organizations*, Free Press, New York 1997, 4th ed.
- Stake R.E., *Program evaluation, particularly responsive evaluation*, [w:] *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, red. G.F. Madaus, M.S. Scriven, i D.L. Stufflebeam, Springer Netherlands, Dordrecht 1983, s. 287–310.

- Stake R.E., *The countenance of educational evaluation*, Teachers College Record 1967, t. 68, s. 523–540.
- Standardy ewaluacji*, Polskie Towarzystwo Ewaluacyjne, Warszawa 2008.
- Stern E., Stame N., Mayne J., Forss K., Davies R. i Befani B., *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*, „DFID: Department for International Development” 2012.
- Stockmann R. i Meyer W., *Functions, methods and concepts in evaluation research*, Palgrave Macmillan, Basingstoke 2013.
- Stockmann R., *The Future of Evaluation in Modern Societies*, [w:] *The Future of Evaluation*, red. R. Stockmann i W. Meyer, Palgrave Macmillan UK, London 2016, s. 36–50.
- Streicher J., *Evaluations, actors and institutions: the case of research, technology and innovation policy in Austria*, Vienna University of Economics and Business, Vienna 2017, <http://epub.wu.ac.at/5445/>.
- Strona Independent Evaluation Group Banku Światowego, <http://ieg.worldbankgroup.org/what-monitoring-and-evaluation>
- Stufflebeam D.L., *The CIPP Model for Program Evaluation*, [w:] *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, red. G.F. Madaus, M.S. Scriven i D.L. Stufflebeam, Springer Netherlands, Dordrecht 1983, s. 117–141.
- Stufflebeam D.L. i Coryn C.L.S., *Evaluation theory, models, and applications*, Jossey-Bass & Pfeiffer Imprints, Wiley, San Francisco 2014, 2. wyd.
- Sturges K.M. i Howley C., *Responsive Meta-Evaluation: A Participatory Approach to Enhancing Evaluation Quality*, „American Journal of Evaluation” 2017, t. 38, nr 1, s. 126–137.
- Suchman E.A., *Evaluative research: principles and practice in public service and social action programs*, Russell Sage Foundation, New York 1967.
- Surdej A., *Ewaluacja w polityce publicznej*, Difin, Warszawa 2013.
- Szarfenberg R., *Dowody naukowe jako podstawa polityki społecznej, zarządzania społecznego i pracy socjalnej*, „Problemy Polityki Społecznej” 2011, nr 15, s. 13–28.
- Szarfenberg R., *Zagadnienia terminologiczne*, [w:] *Ewaluacja w służbach społecznych*, red. B. Szatur-Jaworska, Mazowieckie Centrum Polityki Społecznej, Warszawa 2010, s. 24–37.
- Szatur-Jaworska B., red., *Ewaluacja w służbach społecznych*, Mazowieckie Centrum Polityki Społecznej, Warszawa 2010.
- Theiss M., *Historyczne źródła i czynniki rozwoju ewaluacji*, [w:] *Ewaluacja w służbach społecznych*, red. B. Szatur-Jaworska, Mazowieckie Centrum Polityki Społecznej, Warszawa 2010, s. 15–22.
- Todeschini R., i Baccini A., *Handbook of bibliometric indicators: quantitative tools for studying and evaluating research*, Wiley-VCH Verlag, Weinheim 2016.
- Torres R.T., Preskill H.S. i Piontek M.E., *Communicating and Reporting: Practices and Concerns of Internal and External Evaluators*, „American Journal of Evaluation” 1997, t. 18, nr 2, s. 105–125.
- Trochim W.M., Marcus S.E., Mâsse L.C., Moser R.P. i Weld P.C., *The Evaluation of Large Research Initiatives: A Participatory Integrative Mixed-Methods Approach*, „American Journal of Evaluation” 2008, t. 29, nr 1, s. 8–28.

- Turowski B. i Zawicki M., *Funkcje, etapy, metody i narzędzia ewaluacji*, [w:] *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, red. S. Mazur, UE Katowice, MSAP, Kraków 2007, s. 30–58.
- Tyler R.W., *General Statement on Evaluation*, „The Journal of Educational Research” 1942, t. 35, nr 7, s. 492–501.
- Vedung E., *Four Waves of Evaluation Diffusion*, „Evaluation” 2010, t. 16, nr 3, s. 263–277.
- Vo A.T., i Christie C.A., *Advancing Research on Evaluation Through the Study of Context: Advancing Research on Evaluation Through the Study of Context*, „New Directions for Evaluation” 2015, t. 2015, nr 148, s. 43–55.
- Warwick K., i Nolan A., *Evaluation of Industrial Policy*, OECD Science, Technology and Industry Policy Papers, 2015, [http://www.oecd-ilibrary.org/science-and-technology/evaluation-of-industrial-policy\\_5jz181jh0j5k-en](http://www.oecd-ilibrary.org/science-and-technology/evaluation-of-industrial-policy_5jz181jh0j5k-en).
- Westthorp G., *Using complexity-consistent theory for evaluating complex systems*, „Evaluation” 2012, t. 18, nr 4, s. 405–420.
- White H., *An introduction to the use of randomised control trials to evaluate development interventions*, „Journal of Development Effectiveness” 2013, t. 5, nr 1, s. 30–49.
- Widła Ł., *Zastosowanie metod ilościowych w ewaluacji*, [w:] *Teoria i praktyka ewaluacji interwencji publicznych*, red. K. Olejniczak, M. Kozak, i B. Lendzion, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008, s. 150–178.
- Worthen B. i Sanders J., *Educational Evaluation. Theory and Practice*, „Evaluation Review” 1973, t. 13, nr 1, s. 18–31.
- Znanięcki F., *Education and social change*, red. E. Hałas, P. Lang, Frankfurt am Main 1998.
- Zvoch K., *Modern Quantitative Methods for Evaluation Science: Recommendations for Essential Methodological Texts*, „American Journal of Evaluation” 2014, t. 35, nr 3, s. 430–440.