

Narzędzie wiedzy czy wymóg formalny? Dysfunkcje systemu ewaluacji polityk publicznych w Polsce

Polski Przegląd Ewaluacyjny
Nr 1(4)/2025
©Autorzy 2025

ISSN 2956-5332
pte.org.pl/o-czasopismie

Paulina Kozłowska

Szkoła Główna Handlowa, ORCID: 0000-0002-6368-8186

Beata Paczek

Politechnika Warszawska, ORCID: 0009-0006-8454-1198

Abstrakt

Artykuł podejmuje krytyczną analizę praktyki ewaluacji polityk publicznych w Polsce, ze szczególnym uwzględnieniem badań realizowanych w kontekście polityki spójności Unii Europejskiej. Na podstawie analizy Bazy Badań Ewaluacyjnych, 62 raportów ewaluacyjnych, bazy rekomendacji oraz wieloletnich doświadczeń autorek zidentyfikowano powtarzalne mechanizmy systemowo osłabiające poznawczą i decyzyjną funkcję ewaluacji. Należą do nich m.in. nadprodukcja badań i inflacja metodyczna, pozorność poznawcza, fragmentaryczność analiz, autocenzura w formułowaniu rekomendacji, brak współwłasności wniosków oraz niedopasowanie czasowe badań do procesów decyzyjnych. Autorki argumentują, że problemy te nie mają charakteru incydentalnego, lecz tworzą stabilny wzorzec instytucjonalny, w którym ewaluacja pełni głównie funkcję formalno-legitymizującą dotychczasowe działania. W artykule poddano także krytycznej refleksji tezę o przełomowym znaczeniu sztucznej inteligencji, wskazując, że bez zmiany logiki projektowania badań technologie te mogą wzmacniać istniejące dysfunkcje.

Słowa kluczowe

ewaluacja, polityka spójności, pozorność poznawcza, rekomendacje, instytucjonalizacja

Knowledge Tool or Formal Requirement? Dysfunctions of the public policy evaluation system in Poland

Abstract

This article offers a critical analysis of public policy evaluation practice in Poland, with particular attention to studies conducted within the framework of the European Union's cohesion policy. Drawing on an analysis of the Evaluation Research Database, 62 evaluation reports, a recommendations database, and the authors' long-term professional experience, the study identifies recurring mechanisms that systematically weaken the cognitive and decision-making functions of evaluation. These include, among others, the overproduction of studies and methodological inflation, cognitive superficiality, analytical fragmentation, self-censorship in the formulation of recommendations, a lack of ownership of findings, and the temporal misalignment of evaluations with decision-making processes. The authors argue that these problems are not incidental but constitute a stable institutional pattern in which evaluation primarily serves a formal and legitimising role with respect to existing policies. The article also critically reflects on claims regarding the transformative potential of artificial intelligence, indicating that without changes in the logic of research design, such technologies may reinforce existing dysfunctions.

Keywords

evaluation, cohesion policy, cognitive superficiality, recommendations, institutionalisation

Corresponding author(s):

Beata Paczek, Politechnika Warszawska, Wolański sp. z o.o. , ul. Ossowskiego 11/91, 03-542 Warszawa, beata.paczek@pw.edu.pl

Wstęp

Badania ewaluacyjne stały się dziś nieodłącznym elementem zarządzania środkami publicznymi, zwłaszcza środkami pochodzącymi z funduszy Unii Europejskiej. W nowoczesnym podejściu do polityk publicznych nie wystarcza już jedynie poprawne, zgodne z prawem wydatkowanie pieniędzy. Oczekuje się także, że decyzje będą oparte na dowodach (evidence-based policy), a więc na systematycznym sprawdzaniu, czy interwencje rzeczywiście przynoszą zamierzone efekty, dla kogo są korzystne, jakie mają efekty uboczne i jak można je ulepszyć. Ewaluacja jest jednym z głównych narzędzi dostarczania takiej wiedzy (Olejniczak 2008a).

W Polsce rozwój ewaluacji jest szczególnie silnie związany z polityką spójności i wykorzystaniem funduszy Unii Europejskiej. Od końca lat 90. i początku lat 2000. kolejne regulacje unijne wprowadzały wymóg prowadzenia badań ewaluacyjnych na różnych etapach cyklu życia programu (ex ante, on-going, ex post). Wraz z tymi wymogami powstał rozbudowany system ewaluacji: jednostki ewaluacyjne w administracji, plany ewaluacji na poziomie krajowym i regionalnym, procedury zlecania badań, bazy raportów i systemy monitorowania wdrażania rekomendacji (Olejniczak, Kozak and Ledzion 2012). Ewaluacja stała się formalnie wpisana w cykl wydatkowania środków publicznych – bez niej niemożliwe byłoby dziś programowanie i rozliczanie znacznej części środków UE.

Z perspektywy normatywnej badania ewaluacyjne powinny być prowadzone zgodnie z kilkoma kluczowymi zasadami, akcentowanymi zarówno w standardach międzynarodowych (Joint Committee on Standards for Educational Evaluation 1981; OECD 2019), jak i w polskiej literaturze przedmiotu (Korporowicz 2011, Olejniczak 2008b):

- rzetelność i przejrzystość – jasne formułowanie pytań badawczych, dobór adekwatnych metod, jawność założeń i ograniczeń, uczciwe przedstawianie wyników (także niewygodnych);
- użyteczność i orientacja na użytkownika – projektowanie badań tak, by ich wyniki mogły realnie wesprzeć decyzje (dobór zakresu, formy raportu, sposobów upowszechnienia do potrzeb konkretnych odbiorców) – w duchu *utilisation-focused evaluation* (Patton 2008);
- niezależność i integralność – minimalizowanie konfliktów interesów, unikanie nacisków na „upiększanie” wyników, dbałość o to, by ewaluacja nie była jedynie narzędziem legitymizacji wcześniej podjętych decyzji;
- etyka i odpowiedzialność społeczna – poszanowanie uczestników badań, wrażliwość na potencjalne skutki rekomendacji (dla grup wrażliwych, mniejszych beneficjentów, lokalnych społeczności), unikanie manipulowania wnioskami.

Równocześnie w literaturze od lat zwraca się uwagę na problemy i wyzwania związane z praktyką ewaluacji. Wskazuje się m.in. na:

- ryzyko, że ewaluacja zostanie sprowadzona do formalnego wymogu towarzyszącego wydatkowaniu środków, zamiast pełnić funkcję rzeczywistego uczenia się i korekty polityk (Jąkański 2024);
- rozbieżność między deklarowanymi celami (uczenie się, doskonalenie interwencji) a faktycznym wykorzystaniem wyników badań, które bywa ograniczone lub symboliczne (Kupiec 2014a);
- problemy jakościowe – uproszczone metody, ogólnikowe rekomendacje, brak dopasowania badań do realnych potrzeb decydentów (Olejniczak and Śliwowski 2014);
- napięcie między funkcją kontrolną i rozliczeniową (audyt, monitoring) a funkcją refleksyjną i krytyczną, nastawioną na ujawnianie słabości i formułowanie trudnych wniosków (Korporowicz 2008);
- strukturalne ograniczenia wpisane w sposób projektowania i zlecania ewaluacji – silne sformalizowanie procedur, standaryzacja pytań i narzędzi, koncentracja na zgodności z wytycznymi i wskaźnikami zamiast na pogłębionej diagnozie, co sprzyja „bezpiecznym” badaniom i ogranicza przestrzeń dla krytycznej refleksji oraz formułowania nieoczekiwanych, potencjalnie kontrowersyjnych wniosków (Bienias et al., 2024);
- autocenzurę instytucjonalną i profilaktyczne usuwanie odważnych wniosków – poprzez wykreślanie trudnych rekomendacji na etapie konsultacji lub „wypychanie” ich z tabeli monitorowania do mniej zobowiązującej części opisowej raportu, co eliminuje najbardziej krytyczne postulaty zmian (EGO and LBiE 2025).

Na tym tle zrozumiałe stają się coraz ostrzejsze diagnozy formułowane w literaturze na temat polskich doświadczeń z ewaluacją – dotyczące jej rytualnego charakteru, braku rzeczywistej krytyki, pozorności działań czy autocenzury w formułowaniu rekomendacji (np. Kupiec et al., 2020). W ujęciu teoretycznym ewaluacja powinna stanowić fundament procesu uczenia się instytucji publicznych, oparty na krytycznej refleksji i otwartej interakcji między badaczami a użytkownikami wyników. Jednak w polskiej praktyce administracyjnej funkcja ta bywa osłabiana przez dominację logiki rozliczania i kontroli, co sprzyja formułowaniu wniosków „bezpiecznych” kosztem ich realnej użyteczności poznawczej. Przesunięcie akcentu z doskonalenia polityk na formalną zgodność proceduralną sprawia, że proces produkcji wiedzy staje się często jedynie elementem legitymizacji działań, a nie impulsem do zmiany (Olejniczak, 2008c). W dalszej części tekstu autorki przedstawiają własne doświadczenia badawcze, które można traktować jako rozwinięcie i uszczegółowienie, w jaki sposób opisane wyżej problemy ujawniają się w praktyce konkretnych procesów ewaluacyjnych.

Celem artykułu jest identyfikacja i analiza mechanizmów systemowych osłabiających poznawczą oraz decyzyjną funkcję ewaluacji polityk publicznych w Polsce. Analiza opiera się na przeglądzie raportów ewaluacyjnych zgromadzonych w Bazie Badań Ewaluacyjnych, analizie struktury rekomendacji oraz doświadczeniach autorek wynikających z realizacji badań ewaluacyjnych. Artykuł wnosi wkład do literatury przedmiotu poprzez konceptualizację powtarzalnych „grzechów” praktyki ewaluacyjnej oraz pokazanie ich systemowego charakteru, wykraczając poza analizę pojedynczych badań. Nowością jest także operacjonalizacja zjawiska „pozorności poznawczej” oraz „bezpiecznego języka” rekomendacji poprzez analizę całej populacji rekomendacji systemowych. Tekst wzbogaca także debatę o perspektywę wykonawcy badań, pozwalającą uchwycić operacyjne uwarunkowania pracy ewaluatora oraz mechanizmy kształtujące treść raportów i rekomendacji.

Metodyka

W niniejszym artykule przyjęto podejście zbliżone do metodyki stosowanej w badaniach ewaluacyjnych polityk publicznych, łącząc perspektywę praktyczną z analizą istniejącego dorobku ewaluacyjnego. Podstawą wywodu są z jednej strony wieloletnie doświadczenia autorek w projektowaniu, prowadzeniu i wykorzystywaniu badań ewaluacyjnych w kontekście polityk publicznych, z drugiej – analiza wybranych badań ewaluacyjnych dotyczących wykorzystania funduszy europejskich w Polsce.

Tekst ma charakter syntetyczno-refleksyjny: nie rekonstruuje jednego, zamkniętego projektu badawczego, lecz porządkuje i konceptualizuje powtarzające się problemy („grzechy” ewaluacji), ujawniające się w różnych ewaluacjach i w praktyce instytucjonalnej korzystania z ich wyników. Analizowane są zarówno aspekty metodologiczne (jakość badań, dobór metod, sposób formułowania wniosków i rekomendacji), jak i instytucjonalne oraz kulturowe uwarunkowania praktyk ewaluacyjnych (logika zlecenia badań, oczekiwania administracji, sposób organizacji wdrażania rekomendacji, mechanizmy autocenzury i „wygładzania” krytycznych wniosków).

Proces badawczy oparto na dwupoziomowej analizie danych o badaniach ewaluacyjnych. Podstawowym źródłem wiedzy były raporty zgromadzone w Bazie Badań Ewaluacyjnych (BBE) prowadzonej przez Krajową Jednostkę Ewaluacji dostępne na stronie internetowej (www.ewaluacja.gov.pl) KJE według stanu na 4.12.2025 roku. Baza obejmuje cały katalog badań zrealizowanych od 2004 roku. W wymiarze ilościowym analizie poddano całą populację badań zarejestrowanych w systemie, co pozwoliło na identyfikację trendów dotyczących liczby badań oraz struktury wykonawców. W wymiarze jakościowym przeprowadzono pogłębioną analizę treści w ramach celowej kwerendy 62 badań – dostępnych raportów końcowych oraz dokumentów towarzyszących badaniom (np. Opisy Przedmiotu Zamówienia).

W analizie ilościowej uwzględniono raporty pochodzące z perspektywy finansowej 2004–2006. Zdecydowano się na to, z uwagi na to, że autorki chciały uchwycić ewolucję rynku ewaluacji oraz zmiany jego charakterystyki w czasie.

Badania lat 2004–2006 zostały celowo wyłączone z analizy jakościowej, ponieważ celem tej części badania było pogłębione rozpoznanie mechanizmów funkcjonujących w relatywnie dojrzałej fazie systemu ewaluacji w Polsce. Włączenie badań realizowanych na etapie jego początkowego rozwoju

mogłoby prowadzić do nadreprezentacji zjawisk charakterystycznych dla fazy wdrożeniowej, a nie dla ustabilizowanego modelu funkcjonowania systemu.

Jak wskazano, w celu wyciągnięcia pogłębionych wniosków na temat obecnego stanu realizacji działań ewaluacyjnych i części zidentyfikowanych problemów przeanalizowano wybranych 62 raportów z badań zgromadzonych w Bazie Badań Ewaluacyjnych. Zdecydowano się na analizę 12 raportów odnoszących się do perspektywy finansowej 2007-2013, 38 raportów dotyczących perspektywy 2014-2020 oraz 12 raportów z perspektywy 2021-2027. Dobór miał charakter celowy i był ukierunkowany na uchwycenie możliwie szerokiego zróżnicowania praktyk ewaluacyjnych funkcjonujących w systemie polityki spójności w Polsce. Przy wyborze raportów uwzględniono następujące kryteria:

- zróżnicowanie instytucji zlecających badania (łącznie 28 unikatowych instytucji);
- zróżnicowanie wykonawców badań (27 różnych podmiotów);
- zróżnicowanie kosztów badań (od ok. 80 tys. zł do ok. 1 mln zł);
- reprezentację różnych typów badań ewaluacyjnych (m.in. ewaluacje tematyczne, procesowe czy o charakterze horyzontalnym);
- uwzględnienie badań zleczanych przez instytucje funkcjonujące na różnych poziomach systemu wdrażania polityki spójności (krajowym i regionalnym);
- możliwość analizy zmian w czasie pomiędzy perspektywami finansowymi (np. ewaluacje podobne tematycznie realizowane w różnych perspektywach UE).

Pozwoliło to uchwycić zarówno ciągłość pewnych dysfunkcji, jak i zmiany w czasie, w tym pojawianie się nowych zjawisk, takich jak bardziej sformalizowane mechanizmy monitorowania rekomendacji czy rosnące znaczenie języka wskaźników.

W analizowanej próbie 12,5% badań stanowiły ewaluacje realizowane przez autorki artykułu. Zostały one włączone do analizy ze względu na dobrą znajomość ich przebiegu oraz dostęp do pełnej dokumentacji badawczej. Jednocześnie uwzględnienie badań realizowanych przez autorki zwiększało możliwość porównywania obserwacji wynikających z analizy dokumentów z doświadczeniem praktycznym. Potencjalne ryzyko stronniczości zostało ograniczone poprzez analizę raportów realizowanych przez szerokie grono innych wykonawców oraz instytucji.

Uzupełniającym źródłem danych wykorzystanych w niniejszym artykule była baza rekomendacji z badań ewaluacyjnych dotyczących perspektyw finansowych 2007-2013, 2014-2020 i 2021-2027. Na jej podstawie przeprowadzono analizę częstości słów z wykorzystaniem oprogramowania MAXQDA, której celem było zidentyfikowanie dominujących wzorców językowych w formułowaniu rekomendacji ewaluacyjnych. Analiza ta pozwoliła na empiryczne uchwycenie zjawiska tzw. „bezpiecznego języka” – charakteryzującego się dominacją czasowników modalnych i asekuracyjnych oraz brakiem jednoznacznego przypisania odpowiedzialności. Kontynuacją tej analizy była jakościowa interpretacja semantyczna wybranych elementów rekomendacji, ukierunkowana na identyfikację różnic między językiem postulatywnym a językiem diagnostycznym i reformatorskim.

Artykuł ma zatem charakter analizy krytycznej typu lessons learned: nie tyle ocenia pojedyncze badania, ile rekonstruuje powtarzalne mechanizmy dysfunkcyjne, pokazując, jak kształtują one praktykę ewaluacji i ograniczają jej potencjał poznawczy oraz naprawczy.

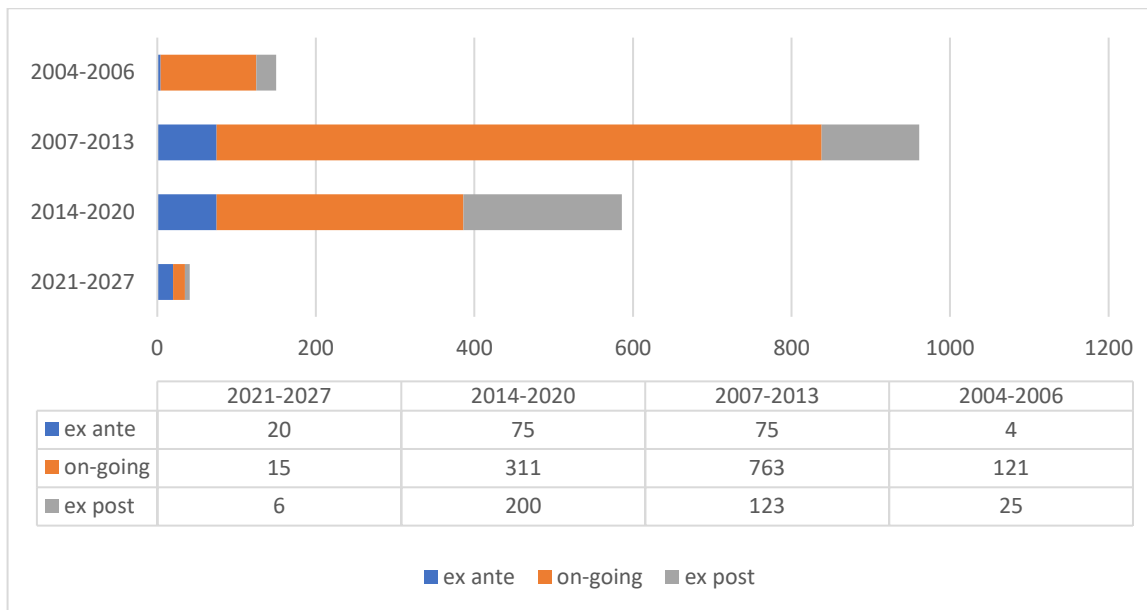
Nadprodukcja ewaluacji

Jednym z kluczowych mechanizmów osłabiających poznawczą funkcję ewaluacji jest zjawisko nadprodukcji badań ewaluacyjnych, któremu towarzyszy „inflacja metodyczna”.

Analiza danych z Bazy Badań Ewaluacyjnych wskazuje, że perspektywy 2004-2006 dotyczyło 150 badań. W kolejnej perspektywie tj. 2007-2013 zrealizowano 961 badań ewaluacyjnych, z których blisko 80% stanowiły ewaluacje on-going, kolejnego okresu tj. 2014-2020 dotyczyło 600 badań, a obecnego – 2021-2027 53 badania (por. Wykres 1).

Tak duża liczba badań ewaluacyjnych wskazuje na instytucjonalizację ewaluacji jako stałego elementu cyklu wdrażania programów finansowanych ze środków UE, a nie jako narzędzia selektywnej refleksji poznawczej. Obserwowane zmniejszenie liczby realizowanych ewaluacji nie musi oznaczać odejścia od logiki nadprodukcji. Jak wynika z analizy raportów, można bowiem zaobserwować tendencję do konsolidowania wielu celów badawczych w ramach pojedynczych, rozbudowanych ewaluacji, charakteryzujących się wysoką liczbą pytań badawczych i wykorzystaniem wielu metod.

Wykres 1. Liczba zrealizowanych badań ewaluacyjnych dotyczących poszczególnych perspektyw finansowych UE



Źródło: opracowanie własne na podstawie danych z Bazy Badań Ewaluacyjnych 1 [stan na 4.12.2025]

W 62 badanych raportach liczba stosowanych metod badawczych w pojedynczym badaniu wahała się od 2 do 11, przy średniej bliskiej 6 metod. Dominującymi technikami pozostają desk research, ankiety ilościowe oraz wywiady pogłębione (IDI), które są często uzupełniane o panele ekspertów, benchmarking czy studia przypadków. Analiza dokumentacji zamówień publicznych – zapisów OPZ wskazuje na liczne przypadki premiowania ofert, które deklarują zastosowanie dodatkowych metod badawczych – wykraczających poza już rozbudowany, wskazany przez zamawiającego katalog metod, obejmujący nierzadko co najmniej sześć z nich. Tego typu zapisy sprzyjają konkurencyjności wykonawców liczbą i różnorodnością narzędzi badawczych, a nie spójnością koncepcji badawczej czy adekwatnością metod do charakteru problemu. W konsekwencji dobór metod staje się elementem strategii przetargowej, a nie wynikiem refleksji metodologicznej.

W analizowanych raportach identyfikowano przypadki realizacji poszczególnych metod badawczych w sposób uproszczony, bez pełnej integracji wyników. Z doświadczeń autorek wynika, że zjawisko to może być powiązane z ograniczoną liczebnością zespołów badawczych (najczęściej do 5 osób) oraz napiętymi harmonogramami realizacji ewaluacji.

Skala nadprodukcji uwidacznia się także w intensywnym wykorzystaniu wywiadów i badań ankietowych. W 92% analizowanych raportów zastosowano wywiady jakościowe, przy czym ich liczba sięgała nawet ponad 150 w jednym badaniu, a średnia dla najnowszej perspektywy (tj. 2021-2027) przekraczała 40 wywiadów. Ankiety wykorzystywano w niemal 91% badań, przy czym liczba zrealizowanych kwestionariuszy była skrajnie zróżnicowana, (choć to jest zależne od typu ewaluacji i jej skali) – od kilkudziesięciu do ponad 60 000 ankiet w jednym projekcie. Zróżnicowanie to w części przypadków wynikało z charakteru badania i jego skali, jednak analiza dokumentacji wskazuje, że wielkość prób

¹ Pominięto badania nachodzące na dwie perspektywy bądź zduplikowane w BBE.

badawczych bywała również determinowana założeniami projektu badawczego formułowanych przez Zamawiających.

Konsekwencją nadprodukcji badań i ich zakresu jest narastające zmęczenie respondentów oraz obniżenie jakości pozyskiwanych danych. Powtarzalność tematów i podobieństwo narzędzi badawczych prowadzą do schematycznych odpowiedzi, dostosowanych do oczekiwań instytucji zamawiającej, a nie do rzeczywistego doświadczenia uczestników interwencji. W skrajnych przypadkach te same jednostki samorządu terytorialnego były w krótkim okresie angażowane w kilka równoległych badań dotyczących tego samego programu lub obszaru interwencji, realizowanych w ramach odrębnych ewaluacji (o projekt CWD pytano np. w badaniu dot. bezpośrednio projektu CWD (Wolański and IDEA Instytut 2025), badaniu dot. efektywności Pomocy Technicznej (Ecorys Polska and EVALU 2025) oraz badaniu wpływu polityki spójności 2014-2020 na rozwój miast i ich obszarów funkcjonalnych (Ecorys Polska and Instytut Rozwoju Miast i Regionów 2025).

W rezultacie nadprodukcja ewaluacji i inflacja metodyczna wzmacniają model ewaluacji jako produktu raportowego, w którym miarą jakości badania staje się jego skala – liczba metod, respondentów i stron, a nie zdolność do generowania syntetycznej, użytecznej wiedzy decyzyjnej.

Pozorność poznawcza

Analiza 62 raportów ewaluacyjnych wskazuje, że jednym z podstawowych przejawów pozorności poznawczej jest sposób formułowania pytań badawczych. Ich liczba wahała się od 3 do 63, przy średniej przekraczającej 21 pytań na raport, przy czym wiele z nich miało charakter wielowątkowy i obejmowało kilka odrębnych problemów badawczych w ramach jednego pytania. Taki sposób projektowania badań wpisuje się w szersze zjawisko, w którym – jak trafnie zauważono – „ewaluacja w polskiej rzeczywistości społecznej, zarządczej i decyzyjnej niestety przyjmuje na wielu płaszczyznach cechy działań pozornych” (Jaskuła 2022: 39).

W analizowanych raportach końcowych z badań ewaluacyjnych listy pytań badawczych nie zawsze były wprost prezentowane w części końcowej dokumentu, w efekcie czego odbiorca nie miał pełnej jasności, jakich odpowiedzi poszukiwano w badaniu. Utrudnia to orientację w logice ewaluacji oraz ocenę adekwatności zastosowanych metod badawczych do postawionych pytań. Z analizy raportów wynika, że zamiast dokonywać selektywnego wyboru kluczowych problemów wymagających rozstrzygnięcia, ewaluacje obejmują szeroki katalog zagadnień, często o ograniczonej wadze decyzyjnej. W rezultacie raporty formułują liczne obserwacje cząstkowe, lecz rzadziej prowadzą do jednoznacznych wniosków dotyczących podstawowych założeń interwencji, mechanizmów jej działania ani rzeczywistych efektów. Pozorność ewaluacji nie polega więc na braku danych empirycznych – identyfikowany jest raczej ich nadmiar – lecz na braku priorytetyzacji problemów badawczych oraz selektywności poznawczej.

Analiza wybranych opisów przedmiotu zamówienia potwierdza dominację pytań o charakterze opisowym i ocennym nad pytaniami analitycznymi i wyjaśniającymi. Pytania badawcze koncentrują się przede wszystkim na identyfikacji zakresu działań, ich postrzeganej trafności oraz deklarowanych efektów, natomiast rzadko prowadzą do rekonstrukcji mechanizmów oddziaływania interwencji czy identyfikacji warunków brzegowych jej skuteczności. W analizowanych dokumentach stosunkowo rzadko identyfikowano pytania problemowe umożliwiające testowanie założeń interwencji lub rozstrzyganie, dlatego w jednych kontekstach dane rozwiązania przynoszą oczekiwane rezultaty, a w innych nie.

Przykładem tej tendencji jest ewaluacja dotycząca instrumentów terytorialnych, w której analiza OPZ ujawniła brak pytania o fundamentalnym znaczeniu dla funkcji ewaluacji – pytania o sens istnienia instrumentów terytorialnych jako takich. Badanie koncentruje się na ocenie sposobu ich wdrażania, struktury zarządzania oraz postrzeganej skuteczności, nie podejmując próby odpowiedzi na pytanie, jaką rolę w systemie polityki rozwoju instrumenty te mają spełniać, w jakich warunkach generują rzeczywistą wartość dodaną oraz czy cele przypisywane ZIT i IIT wymagają istnienia odrębnego instrumentu terytorialnego. Z doświadczeń autorek wynika, że podobny schemat projektowania badań występował również w innych analizowanych przypadkach.

postulatywne formułowanie zaleceń, które nie stawia mocnych tez ani nie przypisuje odpowiedzialności – tylko zachęca do niedookreślonych działań naprawczych. Pojęcia sugerujące głęboką zmianę strukturalną pojawiają się marginalnie (po 1 razie), co wskazuje, że ewaluacje pełnią głównie funkcję legitymizującą – postulując intensyfikację i rozszerzanie istniejącego wsparcia.

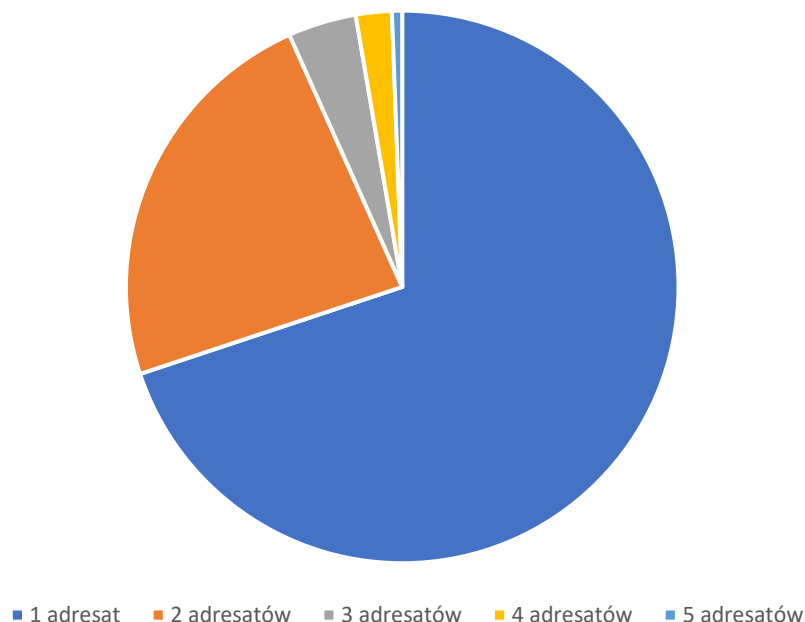
Po drugie – i co pokazują doświadczenia autorek – funkcjonuje mechanizm „wypychania” i chowania trudnych rekomendacji. Ponieważ rekomendacje wpisane do oficjalnej tabeli podlegają formalnemu monitorowaniu i rozliczeniu, pojawia się silna presja – zarówno ze strony zamawiających, jak i samych wykonawców – aby nie umieszczać w niej wniosków „trudnych”, rewolucyjnych lub wymagających znacznych nakładów sił i środków. W efekcie najbardziej wartościowe, krytyczne spostrzeżenia są „wypychane” poza tabelę: trafiają do części opisowej raportu jako luźne sugestie, są łagodzone na etapie konsultacji, a w skrajnych przypadkach – całkowicie usuwane z końcowej wersji dokumentu pod pretekstem ich „niedojrzałości” lub „braku możliwości wdrożenia” – wdrożenie rekomendacji z badania *ex post* często nie jest możliwe w perspektywie finansowej następującej po tej perspektywie, której dotyczy badanie, lecz dopiero w kolejnej, dla której proces programowania jeszcze się nie rozpoczął. Mechanizm ten prowadzi do sytuacji, w której zarówno zamawiający, jak i wykonawcy unikają rekomendacji nieszablonowych, obawiając się trudności z ich późniejszym formalnym rozliczeniem. Ten ostatni przypadek został omówiony w części artykułu dotyczącej „nierespektowania czasu i logiki badania”.

Po trzecie, doświadczenia autorek wskazują, że elementem strategii kontroli nad przekazem jest niepublikowanie analiz częściowych będących wsadem do właściwych badań ewaluacyjnych lub zlecenie analiz własnych jednostek ewaluacyjnych, niepodlegających wymaganiom ani zasadom prowadzenia badań ewaluacyjnych. Zaobserwować to można było również podczas analizy 62 raportów, w których w jedynie nielicznych opublikowano załączniki do raportów. Niepokojącym zjawiskiem jest również praktyka zlecenia pogłębionych analiz, ekspertyz czy badań terenowych, które następnie nie są publikowane ani udostępniane szerokiemu gronu odbiorców. Są one traktowane jedynie jako „wsad” do właściwej ewaluacji, co pozwala zamawiającemu na selektywny wybór tych fragmentów, które pasują do założonej narracji. Ukrywanie surowych danych lub częściowych raportów – czasami krytycznych – pod płaszczykiem „materiałów roboczych” skutecznie ogranicza transparentność procesu i uniemożliwia niezależną weryfikację wniosków końcowych. W ten sposób ewaluacja, zamiast być narzędziem debaty publicznej, staje się narzędziem zarządzania wizerunkiem instytucji.

Brak współwłasności rekomendacji

Jednym z kluczowych warunków skuteczności rekomendacji ewaluacyjnych jest poczucie ich współwłasności przez instytucje odpowiedzialne za projektowanie i wdrażanie polityk publicznych. Literatura przedmiotu wskazuje, że rekomendacje współtworzone i konsultowane z interesariuszami na kolejnych etapach badania mają większe szanse na wdrożenie oraz praktyczne wykorzystanie, ponieważ zwiększają zaangażowanie odbiorców i poczucie odpowiedzialności za ich realizację (Haber 2008). W praktyce jednak proces ewaluacyjny często przyjmuje formę sekwencyjną: ewaluator prowadzi analizę i formułuje rekomendacje, natomiast instytucja publiczna oczekuje na gotowy dokument, ograniczając swój udział do etapu formalnej akceptacji lub przekazania uwag do sformułowanych zaleceń. Analiza całego zbioru rekomendacji (N = 4089) pokazuje, że 69,9% rekomendacji posiada tylko jednego adresata (por. Wykres 2). Oznacza to, że typowa rekomendacja ewaluacyjna jest formułowana jako zalecenie kierowane do pojedynczej instytucji.

Analiza poziomu szczegółowości adresatów rekomendacji wskazuje jednak, że tylko marginalna część zaleceń kierowana jest do konkretnych wydziałów lub departamentów instytucji. Zdecydowana większość rekomendacji adresowana jest na poziomie całych instytucji, bez wskazania komórek organizacyjnych odpowiedzialnych za ich realizację. Może to wynikać zarówno z ograniczonego uwzględniania realnej struktury organizacyjnej instytucji w procesie formułowania rekomendacji, jak i z uwarunkowań systemowych, takich jak zmienność struktur organizacyjnych administracji publicznej czy koncentracja kompetencji decyzyjnych na poziomie kierownictwa instytucji. Jednocześnie niski poziom operacjonalizacji adresatów może utrudniać przypisanie odpowiedzialności za wdrożenie zaleceń oraz ograniczać ich implementacyjność.

Wykres 2. Rozkład liczby adresatów w rekomendacjach

Źródło: opracowanie własne na podstawie bazy rekomendacji z 27 lipca 2024 roku

Pomimo teoretycznie zwiększających się wymogów dotyczących konsultowania rekomendacji z badań ewaluacyjnych (formułowanych jako oczekiwanie Zamawiających w OPZ) doświadczenia autorek wskazują, że wiele instytucji nie odpowiada aktywnie na zaproszenie do wyrażenia swojej opinii czy propozycji zmian. Część podmiotów przesyła uwagi do raportów o charakterze niemerytorycznym (np. stylistyczne, językowe), przesyła je po terminie zgłaszania uwag czy nie przesyła ich w ogóle. Podczas spotkań konsultacyjnych część osób, mimo zalogowania się, nie zabiera głosu wcale. Znane badaczkom są również przypadki, gdy na etapie raportu końcowego identyfikowane są uwagi do założeń metodycznych badań zgłaszane przez osoby, które wcześniej nie były włączone przez Zamawiającego w proces projektowania badania. W rezultacie proces konsultacji rekomendacji może w części przypadków przyjmować charakter formalny i pozorny, który nie prowadzi do rzeczywistej wymiany wiedzy ani współtworzenia wniosków.

Niedostateczna wiedza

W literaturze dotyczącej ewaluacji wielokrotnie podkreślano deficyty kompetencyjne po stronie zamawiających – słabą znajomość standardów badawczych, nieumiejętność formułowania pytań ewaluacyjnych czy zbyt wąskie rozumienie roli ewaluacji (np. Kupiec 2014b). Analiza zebranych raportów oraz dokumentacji towarzyszącej pokazuje jednak, że problem niedostatecznej wiedzy w równym stopniu może dotyczyć także wykonawców badań.

Przede wszystkim skład zespołów badawczych ich kompetencje członków tych zespołów nie zawsze są jasno udokumentowane w raportach końcowych. W ofertach i raportach wymieniane są czasami rozbudowane zespoły ekspertów, konsultantów i analityków, jednak zakres faktycznego udziału poszczególnych osób w realizacji badania bywa trudny do ustalenia na podstawie dostępnej dokumentacji. Opisy ról są skrajnie ogólne („udział w realizacji badania”, „wsparcie merytoryczne”), a w dalszej części raportu nazwiska członków zespołu już się nie pojawiają. Mimo opracowania ramowego harmonogramu badania i przypisania ról do poszczególnych zadań nie jest wiadomym, kto odpowiadał za konstrukcję narzędzi, kto analizował dane, a kto przygotowywał wnioski i rekomendacje. Taki sposób

prezentowania składu zespołu sugeruje, że jest on raczej elementem oferty marketingowej niż rzeczywistym zasobem wiedzy wykorzystanym w procesie badawczym.

Co więcej, analiza raportów oraz danych z Bazy Badań Ewaluacyjnych wskazuje, że te same podmioty realizują ewaluacje w bardzo różnych dziedzinach polityk publicznych, od edukacji, przez pomoc społeczną, po infrastrukturę czy zdrowie. W analizowanych raportach rzadko widać ślady pogłębionej znajomości specyfiki danej branży: opisy kontekstu są powierzchowne i szablonowe, powtarzają te same ogólniki („obszar o kluczowym znaczeniu dla rozwoju społeczno-gospodarczego”, „istotne wyzwania stojące przed beneficjentami”), a niekiedy pojawiają się błędne nazwy instytucji, mylone są typy działań czy wskaźniki charakterystyczne dla danej dziedziny. Z doświadczeń autorek, pełniących role kierownicze i eksperckie w licznych projektach, wynika, że podobne schematy opisu pojawiały się w badaniach dotyczących różnych sektorów polityk publicznych. Może to wskazywać na ograniczoną specjalizację tematyczną części podmiotów realizujących ewaluacje.

Tabela 1. Liczba badań i obszarów tematycznych dla 10 najważniejszych wykonawców badań ewaluacyjnych

Lp.	Wykonawca	Liczba badań	Liczba obszarów
1	EVALU Sp. z o.o.	180	21
2	EU-CONSULT Sp. z o.o.	115	18
3	ECORYS Polska Sp. z o.o.	88	17
4	PSDB Sp. z o.o.	110	14
5	Openfield Sp. z o.o.	53	14
6	IDEA Instytut Sp. z o.o.	26	13
7	PAG Uniconsult Sp. z o.o.	93	12
8	Analiz Rynku Sp. z o.o.	64	12
9	ASM - Centrum Badań i Analiz Rynku Sp. z o.o.	52	12
10	EGO s.c.	61	11

Źródło: opracowanie własne na podstawie danych z Bazy Badań Ewaluacyjnych [stan na 4.12.2025]

Szczegółowa analiza informacji z BBE o wszystkich zrealizowanych badaniach ewaluacyjnych prowadzi do wniosku, że rynek ewaluacji w Polsce jest silnie skoncentrowany wokół niewielkiej grupy podmiotów. Dziesięciu największych wykonawców – realizujących badania samodzielnie lub w konsorcjach – odpowiada łącznie za ponad 800 badań, przy czym lider – EVALU Sp. z o.o. – realizował ewaluacje w 21 różnych obszarach tematycznych (od edukacji, przez zdrowie, transport, energetykę, po system realizacji polityki spójności). Podobnie EU-CONSULT Sp. z o.o. (115 badań, 18 obszarów) czy ECORYS Polska (88 badań, 17 obszarów) czy PSDB (110 badań, 14 obszarów) pojawiają się jako wykonawcy w niemal wszystkich kluczowych dziedzinach polityk publicznych (por. Tabela 1). W praktyce oznacza to, że te same firmy ewaluują programy dotyczące rynku pracy, zdrowia, środowiska, infrastruktury czy edukacji, co stawia pod znakiem zapytania ich rzeczywistą znajomość specyfiki poszczególnych branż i zdolność do pogłębionej diagnozy mechanizmów działania badanych interwencji. W przypadku badań realizowanych w konsorcjach udział dodatkowych podmiotów często wiąże się z dostarczeniem unikalnych kompetencji metodycznych lub potencjału organizacyjnego, co może wpływać pozytywnie na standardy techniczne badania. Niemniej jednak, skala obecności liderów rynku w niemal wszystkich dziedzinach polityk publicznych (od 11 do 21 obszarów tematycznych na podmiot) wskazuje na istotne wyzwanie strukturalne. Zjawisko to wynika w dużej mierze ze specyfiki polskiego rynku ewaluacji, który ze względu na swoją ograniczoną wielkość i cykliczność zamówień (powiązaną z rytmem funduszy UE), utrudnia firmom wąską specjalizację sektorową przy jednoczesnym zachowaniu stabilności zatrudnienia własnych kadr. Wykonawcy zmuszeni są do dywersyfikacji portfela zleceń, co sprzyja budowaniu kompetencji uniwersalnych, ale może utrudniać kumulację wiedzy branżowej wewnątrz organizacji. W efekcie, mimo wysokiej sprawności procesowej, w raportach końcowych nadal identyfikuje się ryzyko

powierzchności diagnoz wynikające z braku osadzenia wniosków w specyficznym kontekście danej branży.

Po trzecie, niedostateczna wiedza i doświadczenie badawcze ujawniają się w sferze metodycznej. W części raportów brakuje podstawowych informacji o sposobie doboru próby, liczebności zrealizowanych wywiadów czy ankiet, terminach i warunkach prowadzenia badań terenowych. Zdarzają się sytuacje, w których deklarowane liczby respondentów budzą wątpliwości (np. wyjątkowo wysokie wskaźniki zwrotu ankiet, niespójne sumy w tabelach bądź też wyciąganie wniosków z badań ilościowych realizowanych na nienaturalnie małej liczbie osób np. sześciu mieszkańców miasta), a opis procedury terenowej jest na tyle lakoniczny, że nie sposób zweryfikować, czy badanie w ogóle przeprowadzono zgodnie z deklarowanym scenariuszem. Niekiedy obiecywana w raporcie anonimowość respondentów pozostaje w oczywistej sprzeczności z bardzo szczegółowymi opisami jednostek czy stanowisk, pozwalającymi na ich łatwą identyfikację. Takie praktyki świadczą nie tylko o braku znajomości standardów etycznych, lecz także o elementarnych brakach w warsztacie badawczym. Te obserwacje nie znajdują odzwierciedlenia w komentarzach zamawiających zamieszczanych w bazie badań ewaluacyjnych. Jedynie w przypadku 8 badań zamieszczonych w bazie badań ewaluacyjnych (według stanu na 4.12.2025 roku) dodano komentarz w kolumnie „Informacje nt. problemów z realizacją badania” – w dwóch przypadkach była to nierzetelna realizacja badania, a w jednym – naliczenie kar umownych. Może to sugerować ograniczoną skalę formalnego raportowania problemów jakościowych, choć interpretacja tego zjawiska wymaga ostrożności.

W rezultacie „grzech niedostatecznej wiedzy” ma charakter dwojaki. Z jednej strony prowadzi do powierzchownych, schematycznych diagnoz, które nie uwzględniają rzeczywistej logiki działania badanych programów i instytucji. Z drugiej – osłabia wiarygodność samych wyników, skoro nie można mieć pewności co do jakości zebranych danych ani kwalifikacji osób odpowiedzialnych za ich interpretację. Ewaluacja staje się wówczas nie tyle narzędziem uczenia się systemu, ile techniczną usługą świadczoną przez podmioty, które nie dysponują ani pogłębioną wiedzą branżową, ani solidnym zapleczem metodycznym.

Fragmentaryczność ewaluacji

Kolejnym mechanizmem osłabiającym poznawczą funkcję ewaluacji jest jej fragmentaryczność, rozumiana jako koncentracja na wąskich wycinkach interwencji publicznej przy jednoczesnym braku perspektywy systemowej.

Analiza zakresu tematycznego badań transportowych wskazuje, że interwencje oceniane są w silnym powiązaniu z konkretnym programem lub działaniem np. konkretnymi osiami priorytetowymi. W rezultacie wpływ inwestycji transportowych oceniany jest w wielu raportach fragmentarycznie, bez odniesienia do równoległych działań realizowanych np. w danej przestrzeni geograficznej. Jednocześnie należy wskazać, że w badaniach tych stawiane są jednostkowe pytania o wpływ inwestycji z danej osi priorytetowej na rozwój społeczno-gospodarczy czy spójność terytorialną. Analizowany jest on jednak fragmentarycznie, bez odniesienia do równoległych działań realizowanych w obszarach tematycznie powiązanych, jak i innych takich jak planowanie przestrzenne czy rozwój usług publicznych. W analizowanych raportach rzadko identyfikowano podejścia umożliwiające ocenę łącznego wpływu wielu interwencji realizowanych równolegle w tym samym obszarze. Brakuje analiz, które pozwalałyby odpowiedzieć na pytanie, w jakim stopniu obserwowane zmiany są rezultatem pojedynczej interwencji, a w jakim – efektem splotu wielu działań realizowanych równolegle w tym samym obszarze.

Fragmentaryczność ewaluacji wzmacniana jest także przez brak integracji wyników badań realizowanych równolegle. Nadprodukcja ewaluacji prowadzi do sytuacji, w której liczne raporty dotyczące zbliżonych zagadnień funkcjonują obok siebie, bez mechanizmów syntezy i porównania wniosków. Z doświadczeń autorek wynika, że jedną z przyczyn może być ograniczona koordynacja pomiędzy instytucjami zlecającymi badania oraz brak systemowych mechanizmów porównywania wyników. W perspektywie 2014-2020 identyfikowano również przypadki, w ramach których, zgodnie z założeniami badania, planowano wykorzystać część wyników czy wniosków z adekwatnych badań regionalnych. Po przystąpieniu do realizacji badania okazywało się, że badania regionalne nie były

jeszcze ukończone i zatwierdzone, co utrudniało realizację badania podsumowującego efekty interwencji.

Analiza struktury rekomendacji potwierdza dominację perspektywy programowej w badaniach ewaluacyjnych. Spośród 4089 zidentyfikowanych rekomendacji³ aż 78% stanowiły rekomendacje programowe operacyjne, podczas gdy rekomendacje horyzontalne – odnoszące się do mechanizmów systemowych i międzysektorowych – stanowiły mniej niż 4% całości, a rekomendacje pozasystemowe ok. 3%. Taki rozkład może wskazywać, że ewaluacje koncentrują się przede wszystkim na optymalizacji funkcjonowania pojedynczych programów lub działań, w ograniczonym stopniu podejmując refleksję nad interakcjami pomiędzy różnymi politykami publicznymi oraz efektami skumulowanymi interwencji realizowanych w tej samej przestrzeni.

Nierespektowanie czasu i zakresu badania

Jednym z istotnych problemów identyfikowanych w ewaluacjach jest także niedopasowanie czasowe pomiędzy momentem realizacji badania, formułowania rekomendacji a realnym cyklem decyzyjnym oraz instytucjonalnym, w którym rekomendacje te miałyby zostać wdrożone. Problem ten dotyczy zarówno zbyt wczesnego, jak i zbyt późnego lokowania badań wobec procesów programowania, wdrażania i modyfikacji interwencji publicznych. Z wieloletnich doświadczeń autorek wynika, że wiele zaleceń formułowanych jest w odniesieniu do kolejnej perspektywy finansowej, mimo że badania realizowane są w warunkach regulacyjnych perspektywy poprzedniej. Prowadzi to do strukturalnego niedopasowania: ewaluatorzy diagnozują problemy, na których rozwiązanie w danym cyklu programowania nie ma już przestrzeni decyzyjnej. Obserwacje poczynione w ramach praktyki wykonawczej autorek wskazują, że szczególnie ewaluacje realizowane na etapie *ex post* charakteryzują się niską skutecznością wdrożeń. Kluczowe rozstrzygnięcia instytucjonalne dla nowego okresu programowania zapadają często przed zakończeniem pogłębionej analizy efektów okresu poprzedniego. W efekcie ewaluacja, zamiast dostarczać wiedzy do programowania, staje się często jedynie formalnym domknięciem cyklu wydatkowania środków, co czyni jej wnioski spóźnionymi wobec realnych potrzeb decydentów.

Innym problemem jest formułowanie wymagań analitycznych nieadekwatnych do stanu dostępnych danych. W praktyce oznacza to oczekiwanie, że ewaluatorzy przeanalizują zjawiska, które nie zostały w żaden sposób udokumentowane pierwotnie. Przykładowo wyzwaniem jest ocenianie efektów interwencji bez przeprowadzenia pomiaru początkowego. Dotyczy to zwłaszcza ewaluacji miękkich rezultatów programów doradczych skierowanych do przedstawicieli JST, takich jak wzrost wiedzy, kompetencji czy zdolności instytucjonalnych. Identyfikowane są badania, w których formułowano wnioski dotyczące poprawy kompetencji czy poziomu wiedzy, mimo że na początku interwencji nie przeprowadzono żadnego badania bazowego, które pozwalałoby określić stan wyjściowy. Formułując wnioski ewaluator odnosi się więc do subiektywnych odczuć badanych, z których część mogła nawet nie brać udziału w danym wsparciu szkoleniowo-doradczym. Prowadzi to do trudności w sformułowaniu właściwych rekomendacji odpowiadających na zdiagnozowane problemy.

Dyskusja i wnioski

Przeprowadzona analiza wskazuje, że problemy jakości i użyteczności ewaluacji polityk publicznych w Polsce mają charakter nie tylko incydentalny, lecz w znacznym stopniu systemowy. Zidentyfikowane mechanizmy – obejmujące m.in. nadprodukcję badań i inflację metodyczną, pozorność poznawczą, fragmentaryczność analiz, autocenzurę, brak współwłasności rekomendacji czy nieadekwatność czasu badania – wzajemnie się wzmacniają, prowadząc do osłabienia funkcji poznawczej i decyzyjnej ewaluacji.

Uzyskane wyniki pozostają spójne z wcześniejszymi diagnozami dotyczącymi funkcjonowania systemu ewaluacji w Polsce. W analizowanej literaturze przedmiotu wskazywano na napięcie pomiędzy

³ Baza rekomendacji według stanu z 27 lipca 2024 roku udostępniona na potrzeby realizacji badania pn. „Badanie problemów związanych z wdrażaniem rekomendacji opracowanych w ramach ewaluacji działań Polityki Spójności”.

formalnym obowiązkiem prowadzenia ewaluacji a rzeczywistą instytucjonalizacją jej funkcji zarządczej (Żuber & Bienias 2008). Podobnie obserwacje dotyczące ograniczonej użyteczności rekomendacji oraz braku odpowiedzi na kluczowe pytania badawcze znajdują potwierdzenie w badaniach nad regionalnymi programami operacyjnymi (Kupiec 2014b). W niniejszym artykule zidentyfikowano dodatkowo zjawisko nadmiaru pytań badawczych oraz ich wielowątkowości, co można interpretować jako rozwinięcie wcześniejszych obserwacji dotyczących trudności w koncentracji ewaluacji na kluczowych problemach decyzyjnych.

Dominacja perspektywy programowej i operacyjnej nad analizą horyzontalną sprzyja fragmentaryzacji obrazu rzeczywistości. Ewaluacje koncentrują się na pojedynczych instrumentach, osiach priorytetowych czy działaniach, rzadziej podejmując próbę rekonstrukcji efektów skumulowanych i interakcji pomiędzy równolegle realizowanymi interwencjami. W konsekwencji system polityk publicznych oceniany jest w częściach, bez pełnego uchwycenia mechanizmów zmian zachodzących w danej przestrzeni społeczno-gospodarczej. Zjawisko to można interpretować jako rozwinięcie wcześniejszych obserwacji dotyczących formalizacji i segmentacji procesu ewaluacyjnego (Żuber & Bienias 2008).

Analiza procesu formułowania i konsultowania rekomendacji wskazuje dodatkowo na ograniczony poziom rzeczywistej współwłasności wniosków ewaluacyjnych, co przejawia się m.in. w ich relatywnie słabym osadzeniu w realnych strukturach organizacyjnych i decyzyjnych instytucji publicznych oraz dominacji logiki zewnętrznej analizy eksperckiej. Wyniki niniejszego badania można również interpretować w świetle perspektywy relacji pomiędzy zlecającym i wykonawcą ewaluacji. W literaturze podkreślano znaczenie kompetencji zamawiających, sposobu zarządzania kontraktem oraz jakości współpracy pomiędzy stronami (Haber 2008; Ledzion 2008), natomiast przeprowadzone analizy wskazują, że ograniczenia jakości ewaluacji mogą dotyczyć zarówno instytucji zlecających, jak i wykonawców badań. Sugeruje to złożony, relacyjny charakter problemu wykraczający poza jednostronne przypisywanie odpowiedzialności jednej ze stron procesu ewaluacyjnego.

Istotnym elementem ewaluacji jest opieranie się na pozornym „bezpieczeństwie”, które sprzyja autocenzurze i „wygładzaniu” wniosków. Rekomendacje formułowane są często w sposób asekuracyjny, pozbawiony jednoznacznych tez i przypisania odpowiedzialności, a najbardziej krytyczne obserwacje są marginalizowane lub wypychane poza formalne mechanizmy monitorowania. Ewaluacja traci w ten sposób zdolność do pełnienia funkcji krytycznej, stając się narzędziem legitymizującym wcześniej podjęte decyzje, a nie przestrzenią refleksji nad ich zasadnością. Zidentyfikowany w badaniu mechanizm operowania „bezpiecznym językiem” oraz unikania pytań fundamentalnych pozwala na interpretację polskiej ewaluacji w kategoriach rytuału instytucjonalnego (Jaskuła 2022). Wyniki analizy rekomendacji dowodzą, że system premiuje powtarzalność i przewidywalność wniosków kosztem ich wartości poznawczej. Zjawisko to koresponduje z diagnozami dotyczącymi lęku przed innowacyjnością (Leszczyńska 2016) – zarówno po stronie zamawiających, jak i wykonawców. W tym ujęciu cenzura i autocenzura nie są jedynie błędami warsztatowymi, lecz racjonalną strategią przetrwania w systemie, który rygorystycznie rozlicza z wdrożenia zaleceń, nie oferując jednocześnie przestrzeni na ryzyko badawcze czy błąd pilotażowy.

Rytualny charakter ewaluacji znajduje swoje dopełnienie w nierespektowaniu czasu i zakresu badania. Jak wskazują dane źródłowe dotyczące niskiej skuteczności wdrożeń rekomendacji (EGO and LBiE 2025), mamy do czynienia z systemowym problemem marginalizacji zaleceń, który pogłębia się wraz z cyklem życia programu – najniższe wskaźniki implementacji odnotowuje się właśnie w badaniach typu *ex post*. Synteza tych danych z autorską analizą rekomendacji w ramach niniejszego badania prowadzi do wniosku o strukturalnym charakterze dysfunkcji: ewaluacja w Polsce często przestaje być narzędziem *evidence-based policy*, a staje się narzędziem *policy-based evidence*, służącym do legitymizacji *post-hoc* przyjętych już kierunków interwencji, a nie wspierania uczenia się instytucji i ulepszania przyszłych interwencji.

W perspektywie kolejnych okresów programowania szczególnego znaczenia nabiera kwestia jakości wiedzy ewaluacyjnej. Analiza rynku wykonawców wskazuje, że bez wzmocnienia realnych kompetencji branżowych i metodycznych ewaluatorów, jak również zamawiających, trudno oczekiwać pogłębionej diagnozy złożonych interwencji i wskazywania potrzebnych zmian. Standaryzacja produktów ewaluacyjnych i koncentracja rynku sprzyjają efektywności proceduralnej, lecz jednocześnie ograniczają

zdolność do krytycznej analizy mechanizmów działania polityk publicznych. W dłuższej perspektywie grozi to utrwaleniem modelu ewaluacji jako technicznej usługi, a nie procesu generowania wiedzy eksperckiej.

Na tym tle coraz częściej pojawiają się głosy ekspertów, że technologiczna innowacja może stanowić impuls do przełamania opisanych ograniczeń. Pojawienie się narzędzi opartych na sztucznej inteligencji (ang. artificial intelligence, AI) bywa przedstawiane jako potencjalny przełom w praktyce ewaluacyjnej (Olejniczak et al., 2025). AI oferuje możliwość szybkiej analizy dużych zbiorów danych, automatyzacji syntez oraz redukcji kosztów operacyjnych. Jednak, zdaniem autorek, analiza przedstawionych mechanizmów prowadzi do wniosku, że zastosowanie AI nie rozwiązuje kluczowych dysfunkcji systemu, a w pewnych warunkach może je wręcz wzmocnić.

AI doskonale wpisuje się w logikę nadprodukcji i inflacji metodycznej. Automatyzacja analizy: dokumentów, transkrypcji czy danych ilościowych obniża koszty generowania kolejnych raportów, co może prowadzić do dalszego rozszerzania zakresu badań, bez równoległej refleksji nad ich użytecznością. Ryzyko polega na tym, że AI stanie się narzędziem produkcji jeszcze bardziej rozbudowanych, lecz poznawczo płytkich opracowań – dłuższych, bardziej „nasyconych danymi”, ale niekoniecznie bardziej trafnych analitycznie.

Wykorzystanie AI nie eliminuje także problemu pozorności poznawczej, ponieważ algorytmy działają w granicach zadanych pytań badawczych. Jeżeli pytania są zachowawcze, opisowe i nie wskazujące na problem, AI jedynie zoptymalizuje dostarczanie odpowiedzi na źle postawione pytania badawcze. Istnieje ryzyko, że generowane przez AI syntezy będą dodatkowo wzmacniały pozór obiektywności i kompletności, maskując fakt, że kluczowe kwestie zostały wyłączone z pola analizy już na etapie projektowania badania.

Z perspektywy dyskusji zasadniczy problem polskiej ewaluacji nie dotyczy więc braku danych, metod ani technologii, lecz braku zapotrzebowania na wiedzę, która mogłaby realnie zakwestionować istniejące rozwiązania. Dopóki ewaluacja pozostanie elementem rytuału rozliczeniowego, a nie przestrzenią debaty o sensie i skutkach polityk publicznych, jej potencjał poznawczy – niezależnie od zastosowanych narzędzi, w tym sztucznej inteligencji – będzie systemowo ograniczany. Wprowadzenie do analizy perspektywy wykonawcy badań pozwoliło na odsłonięcie zakulisowych mechanizmów, które rzadko znajdują odzwierciedlenie w oficjalnych dokumentach systemowych. Proponowana w artykule konceptualizacja „grzechów” ewaluacji stanowi zatem zaproszenie do debaty nad zmianą paradygmatu: z ewaluacji jako technicznej usługi rozliczeniowej na ewaluację jako autentyczny proces krytycznej refleksji nad sensem interwencji publicznych.

Deklaracja dotycząca konfliktu interesów: Autorki oświadczają, że w związku z niniejszym badaniem nie występuje konflikt interesów.

Deklaracja dotycząca wykorzystania narzędzi generatywnej sztucznej inteligencji: Wykorzystano narzędzie AI (ChatGPT 5.3 Instant) do korekty językowej tekstu.

Finansowanie: Badanie zrealizowano bez zewnętrznego wsparcia finansowego.

Bibliografia

- Bienias S., Kupiec T., Strzęboszewski P. (red.) (2024) *Poradnik ewaluacji dla administracji publicznej*, Warszawa, Ministerstwo Funduszy i Polityki Regionalnej.
- Ecorys Polska sp. z o.o. and EVALU sp. z o.o. (2025). *Ewaluacja ex post Programu Operacyjnego Pomoc Techniczna 2014–2020*.
- Ecorys Polska sp. z o.o. and Instytut Rozwoju Miast i Regionów (2025). *Wpływ polityki spójności 2014–2020 na rozwój miast i ich obszarów funkcjonalnych*.

- EGO s.c. and LBiE (2025). Badanie problemów związanych z wdrażaniem rekomendacji opracowanych w ramach ewaluacji działań Polityki Spójności. Raport dla Ministerstwa Funduszy i Polityki Regionalnej. Warszawa.
- European Commission (2014). Better Regulation Guidelines. Brussels.
- Haber, A. (2008). Zarządzanie badaniem ewaluacyjnym – perspektywa zlecającego. In K. Olejniczak, M. Kozak, & B. Ledzion (Red.), *Teoria i praktyka ewaluacji: Podręcznik akademicki* (pp. 212–230). Wydawnictwa Akademickie i Profesjonalne.
- Jąkański, J. (2024). Analiza przyczyn niedostatecznego wykorzystania wniosków z ewaluacji: perspektywa socjologiczna. *Polski Przegląd Ewaluacyjny*, 1(3), pp. 39–41.
- Jaskuła, S. (2022). Społeczna użyteczność ewaluacji. *Polski Przegląd Ewaluacyjny*, 1, pp. 27–43.
- Joint Committee on Standards for Educational Evaluation (1981). *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw-Hill.
- Korporowicz, L. (2008). Badania ewaluacyjne. In: Kawalec, P. and Lipski, P. (eds.) *Zarządzanie badaniami naukowymi i pracami rozwojowymi w jednostkach naukowych*, Vol. 2. Lublin: Wydawnictwo LBS, p. 65.
- Korporowicz, L. (2011). Zmienne losy polskiej ewaluacji: pomiędzy nadzieją, animacją i konfuzją. In: Krajowa Jednostka Ewaluacji, *Baza Badań Ewaluacyjnych* <https://www.ewaluacja.gov.pl> (dostęp: 4.12.2025).
- Kupiec, T. (2014a). Ewaluacja jako narzędzie zarządzania w sektorze publicznym. *Studia z Polityki Publicznej*, 1(2), pp. 1–22.
- Kupiec, T. (2014b). Użyteczność ewaluacji jako narzędzia zarządzania regionalnymi programami operacyjnymi. *Studia Regionalne i Lokalne*, 15(56), 52-67.
- Kupiec, T., Wojtowicz, D., & Olejniczak, K. (2020). Between compliance and systemic change: Evaluation practice in eight CEE countries. In *Public Administration in Central Europe* (pp. 129-146). Routledge.
- Ledzion, B. (2008). Zarządzanie kontraktem ewaluacyjnym – perspektywa wykonawcy. In K. Olejniczak, M. Kozak, & B. Ledzion (Red.), *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki* (pp. 231–248). Wydawnictwa Akademickie i Profesjonalne.
- Leszczyńska, B. (2016). W poszukiwaniu lepszych rekomendacji... Jak wykorzystać podejście design thinking w pracy ewaluatora. In *Nowe wyzwania dla ewaluacji programów operacyjnych w perspektywie 2014–2020*. Urząd Marszałkowski Województwa Kujawsko-Pomorskiego.
- Ministerstwo Funduszy i Polityki Regionalnej (2024). Baza rekomendacji z badań ewaluacyjnych (stan na 27 lipca 2024). Udostępniona na potrzeby badania: Badanie problemów związanych z wdrażaniem rekomendacji opracowanych w ramach ewaluacji działań Polityki Spójności.
- Niemierko, B. and Szmigiel, M.K. (eds.) (2011). *Ewaluacja w edukacji: koncepcje, metody, perspektywy*. Kraków: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- OECD (2019). *Better Criteria for Better Evaluation*. Paris: OECD Publishing.
- OECD Development Assistance Committee (1991). *Principles for Evaluation of Development Assistance*. Paris: OECD.
- Olejniczak, K. (2008a). *Ewaluacja w administracji publicznej: teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.
- Olejniczak, K. (2008b). *Mechanizmy wykorzystania ewaluacji*. Warszawa: Wydawnictwo Naukowe Scholar.
- Olejniczak, K. (2008c). Wprowadzenie do zagadnień ewaluacji. W: K. Olejniczak, M. Kozak, B. Ledzion (red.), *Teoria i praktyka ewaluacji Podręcznik akademicki*, Warszawa: Wydawnictwa Akademickie i Profesjonalne, pp. 17–43.
- Olejniczak, K. and Śliwowski, P. (2014). Nadchodzi rewolucja? Analizy behawioralne w interwencjach publicznych. In: Haber, A. and Olejniczak, K. (eds.) *(R)ewaluacja 2. Wiedza w działaniu*. Warszawa: PARP.
- Olejniczak, K., Kozak, M. and Ledzion, B. (eds.) (2012). *Ewaluacja: teoria i praktyka w kontekście funduszy strukturalnych*. Warszawa: Wydawnictwo Naukowe Scholar.
- Olejniczak, K., Batorski, D., Pokorski, J., (eds.) (2025). *Generatywna AI w badaniach. Praktyczne zastosowania w ewaluacji polityk publicznych*. Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.
- Patton, M.Q. (2008). *Utilization-Focused Evaluation*. 4th ed. Thousand Oaks, CA: Sage.
- Wolański sp. z o.o. and IDEA Instytut sp. z o.o. (2025). Ocena realizacji projektów Centrum Wsparcia Doradczego (CWD pilotaż, CWD Plus, CWD 2023) w celu wypracowania modelu systemowego wsparcia kompetencji jednostek samorządu terytorialnego w ramach polityki regionalnej.