

Bartłomiej Walczak

*Uniwersytet Warszawski**

Email: b.walczak@uw.edu.pl

ORCID: 0000-0002-0346-712X

Dlaczego warto uczyć nauczycielki i nauczycieli ewaluacji? Perspektywy teoretyczne i praktyczne**

Summary

WHY IS IT A GOOD IDEA TO TEACH TEACHERS ABOUT EVALUATION? THEORETICAL AND PRACTICAL PERSPECTIVES**

The paper argues for a wider presence of evaluation in the training of future teachers. The argumentation relies on an analysis of the evaluation theory and its impact on the quality of teaching. The milestones are presented, along with Donna Merten's division of evaluation theory into four types: postpositivist, constructivist, transformative and pragmatic. Instead of showing a linear vision of evaluation development, the author proposes a heterodox approach, as suggested by Henryk Mizerek. Next, the paper discusses the evaluation status of the trans-discipline. The second part proposes a review of research on the impact of evaluation on the quality of teaching and learning, including a discussion of limitations in available data and reviews. The last section focuses on evaluation at the micro level: schools, teams, and individual teachers. The author concludes that theoretical development and internal differentiation in evaluation, together with the proven influence on the quality of teaching and learning, bring arguments to strengthen the presence of evaluation in teacher education curricula. In particular, self-evaluation, with its ability to improve self-reflexivity, needs more attention.

Keywords: educational evaluation, self-evaluation, evaluation theory, impact of evaluation on the quality of school work

* Adres: Instytut Stosowanych Nauk Społecznych, Uniwersytet Warszawski, ul. Nowy Świat 69, 00-927 Warszawa

** Finansowanie publikacji: Uniwersytet Warszawski

Wstęp

Ewaluacja edukacyjna w Polsce jest dyscypliną stosunkowo młodą, ale z pewnością nie nową (Korporowicz, 2011; Mizerek, 2016; Niemierko, 2019). Po okresie do 1993 r., nazywanym przez Henryka Mizerka „erą milczenia” (Mizerek, 2017, s. 51–52), nastąpił stopniowy rozwój, związany z jednej strony z działalnością akademików i akademikzek zajmujących się tą tematyką, z drugiej – systemowymi reformami, wprowadzającymi ewaluację do prawa oświatowego i praktyki szkolnej. Pozostaje otwarte pytanie o integrację tych procesów. Jak zauważa Justyna Nowotniak, ewaluacja jako praktyka wydaje się nadal odseparowana od ewaluacji jako dyscypliny naukowej (Nowotniak, 2019, s. 17). Znajduje to wyraz w słabej obecności ewaluacji w programach kształcenia przyszłych nauczycieli, którzy uzupełniają wiedzę w ramach studiów podyplomowych lub komercyjnych szkoleń (Jaskuła, 2018, s. 77–78). Jakkolwiek uczelnie odegrały znaczącą rolę we wprowadzaniu ewaluacji do polskiej edukacji (Sury, 2020, s. 44), nie znajduje to odbicia w specjalizacji jednostek organizacyjnych ani programach studiów pedagogicznych (Sury, 2020, s. 50). W niniejszym tekście zamierzam uzasadnić potrzebę większego nacisku na rozwój kompetencji ewaluacyjnych wśród studentów i studentek kierunków nauczycielskich oraz pedagogicznych. Argumentację opieram na dwóch podstawach: (1) teoretycznej dojrzałości ewaluacji jako dystynktywnej transdyscypliny, wymagającej wyodrębnionych kursów oraz (2) dowodach na korzystny wpływ na jakość nauczania. W ostatniej części artykułu proponuję koncentrację na autoewaluacji jako rozwiązaniu najbardziej adekwatnym do potrzeb nauczycieli i nauczycielek.

Transdyscyplina na progu heterogeniczności

W pierwszej części tekstu dokonam przeglądu historii ewaluacji i jej rozdzielania się na poszczególne nurty. Bez krytycznej refleksji nad rozwojem w długiej perspektywie czasowej i zależnościami między poszczególnymi nurtami trudno wyobrazić sobie świadome stosowanie ewaluacji w praktyce nauczycielskiej. Jednocześnie bogactwo dorobku teoretyków i teoretyczek stanowi argument za wyodrębnieniem ewaluacji jako osobnego przedmiotu nauczania. Wreszcie historia ewaluacji pokazuje przejście od bardzo „technicznego” pomiaru efektów do urefleksyjnienia roli ewaluatora i uświadomienia odpowiedzialności za upodmiotowienie interesariuszy.

Choć działań o charakterze protoewaluacyjnym można szukać w wielu kulturach i okresach historycznych, początki współczesnej ewaluacji zwykło się lokować w końcu XVIII w., wraz z wprowadzeniem przez Williama Farisha ilościowych ocen prac uczniów. Jak zauważa Lance Hogan, pozwoliło to na odejście od oce-

niania retorycznego stylu wypowiedzi i dało początek psychometrii (Hogan, 2007, s. 4). W kolejnych latach w różnych krajach rozwijano systemy oceniania, traktowane nie tylko jako miara jakości pracy uczniów, ale i nauczycieli. Niemniej kolejnym krokiem milowym były prowadzone niemal równo sto lat po reformie Farisha badania Josepha Rice (Rice, 1897; Guba, Lincoln, 1989, s. 23). Rice był zainteresowany efektywnym wykorzystaniem czasu w szkole. Prowadzone od 1887 r. badania nad nauczaniem gramatyki na imponującej próbie 38 tysięcy uczniów z różnych części Stanów Zjednoczonych są uznawane za pierwszą ewaluację programu kształcenia (Guba, Lincoln, 1989, s. 23; Hogan, 2007, s. 4). Rice zakładał, że lepsza jakość procesu uczenia się może pozwolić na poszerzenie podstawy programowej poza podstawowe (*basic*) przedmioty. Kluczem do zmiany miała być praca nad kompetencjami nauczycieli (Rice, 1897, s. 419).

Jak widać, w twierdzeniu Michaela Scrivena: „ewaluacja jest młodą dyscypliną, choć jednocześnie bardzo starą praktyką” (1996, s. 395) nie ma wiele, jeśli w ogóle, przesady. Scriven wskazuje, że rozwój praktyki ewaluacji nie był równomierny z podbudową teoretyczną. W pierwszych dekadach XX w. miało miejsce upowszechnienie psychometrycznych technik pomiaru kompetencji uczniów oraz instytucjonalizacja. Powołano m. in. znane *American Educational Research Association* (pierwotnie jako *National Association of Directors of Educational Research*, 1916). Liczba narzędzi psychometrycznych używanych jedynie w USA przekroczyła 5000 (Hildreth, 1979 [1933, 1945]). W tym okresie pomiar i ewaluacja były traktowane synonimicznie, a rozwój ewaluacji jako dyscypliny koncentrował się na metodologii.

Od wczesnych lat 30. XX w. możemy mówić o bardziej intensywnym rozwoju teorii ewaluacji. Okres ten jest nierozzerwalnie związany z postacią Ralphi Tylera. Hogan (2007, s. 5) nazywa wręcz czas do końca II wojny światowej „epoką Tylera”. Skąd tak mocne twierdzenie? Rozwój dyscypliny naukowej wyznacza kilka kryteriów – określenie specyficznego obszaru, wykształcenie pojęciowości, przyjęcie rozwiązań epistemologicznych i metodologicznych, wreszcie instytucjonalizacja. Ewaluacja edukacyjna w latach 30. miała wyodrębniony przedmiot, do pewnego stopnia była zinstytucjonalizowana. Tyler zapoczątkował wykorzystywanie kryteriów ewaluacyjnych, dając tym samym niespotykaną perspektywę epistemologiczną, w dłuższej perspektywie prowadzącą do specyficznych rozwiązań metodologicznych i analitycznych. Jednocześnie wykształcił znaczną liczbę ewaluatorów (w tym, na przykład, szeroko znanego z metody analizy rzetelności Lee Cronbacha), którzy nadawali kształt rozwojowi dyscypliny przez kolejne dekady (Madaus, 2004).

Chociaż paradygmatyczne osadzenie Tylera było ewidentnie pozytywistyczne, wychodziło poza typowe dla tego okresu utożsamianie ewaluacji z testami. Nacisk na ciągłość procesu ewaluacji, w celu maksymalizacji osiągnięć efektów uczenia się przez uczniów, to zapowiedź podziału na podejście sumatywne i formatywne,

wprowadzone przez Michaela Scrivena dobre 30 lat później. Postulat włączenia do ewaluacji różnorodnych metod – obserwacji, analizy prac uczniowskich, zastosowania wywiadów, badań ankietowych i studiów przypadku, wreszcie wykorzystanie profesjonalnej wiedzy nauczycieli – musiał poczekać przynajmniej jedno pokolenie (Madaus, 2004, s. 72–73).

Kolejna istotna cezura miała miejsce w latach 60. XX w. Dochodzi wtedy do uświadomienia *wartościowania* jako składowej ewaluacyjnej episteme. „Zarówno opis, jak i osąd [*judgment*] są kluczowe – w istocie, są dwoma podstawowymi działaniami w ewaluacji” pisze Robert Stake (1967, s. 1). Jest to moment, w którym ujawnia się dystynktywna cecha ewaluacji jako szerokiej strategii epistemologicznej w nauce: dopuszczenie do wydawania sądów wartościujących. Wpisuje się to zresztą w zachodzące w tym okresie zmiany w naukach społecznych, wiążące się z zakwestionowaniem pozytywistycznego obiektywizmu, szczególnie zwrot lingwistyczny w antropologii kultury i rozwój paradygmatu fenomenologicznego w socjologii. Jak zauważa Michael Scriven, używanie sądów wartościujących wymaga wprowadzenia standardów, do których można się odwoływać w procesie wnioskowania (Scriven, 1966). Oznacza to nie tylko odrzucenie dogmatu nauki wolnej od wartościowania – w istocie wiele nauk społecznych odkrywa w tym okresie wpływ przedsądów na proces tworzenia wiedzy – ale też poddanie ocenie samych standardów (Guba, Lincoln, 1989, s. 30).

Od tego momentu trudno jest mówić o liniowym rozwoju ewaluacji. Rozwój nauk społecznych i edukacyjnych, a zwłaszcza upowszechnienie konstruktywizmu oraz podejść krytycznych i emancypacyjnych przekładają się na znaczące zróżnicowanie nurtów w ewaluacji. W metodologii badań ewaluacyjnych nabierają znaczenia strategie jakościowe, umożliwiające upodmiotowienie interesariuszy. Stan, w którym obecnie znajduje się teoria ewaluacji można za Henrykiem Mizerkiem określić jako heterodoksja na progu heterogeniczności. „Żadna z koncepcji, szkół, żadne z podejść w ewaluacji nie ma monopolu na prawdę. Rośnie przekonanie, że nie istnieje taka teoria, która może zapełnić intelektualnie wszelkie potencjalne obszary zainteresowań ewaluacji” (Mizerek, 2017, s. 40).

Liniowe wizje ewaluacji (Guba, Lincoln, 1989; Hogan, 2007) zdają się pasować raczej do opisu rozwoju dyscypliny w aspekcie chronologicznym. Przyjrzyjmy się nieliniowym próbom klasyfikacji teorii ewaluacyjnych. Pewną próbą kompromisu jest „drzewo teorii ewaluacji”, przedstawione przez Marvinę Alkina i Christinę Christie (2004). Proponują oni spojrzenie na rozwój ewaluacji w podziale na trzy osie, odzwierciedlające trzy priorytety rozwojowe: zastosowanie, metody i wartościowanie. Miejsce poszczególnych modeli teoretycznych na osiach (gałęziach drzewa) zależy od roli w rozwoju danego nurtu i momentu historycznego, w którym zostały wprowadzone.

Na obecny stan teorii ewaluacji można wreszcie patrzeć z nieliniowej perspektywy paradygmatycznej, nawet jeśli trudno zgodzić się, że poszczególne nurty wyczerpują cechy paradygmatów w klasycznym, Kuhnowskim rozumieniu. Ostatecznie możemy obserwować koegzystencję różnych nurtów, a nie ich zastępowanie w porządku czasowym (Kuhn, 2001). Klasyfikację na duże grupy teorii proponuje Donna Mertens (2015). Podzieliła ona teorie ewaluacji na cztery grupy: postpozytywistyczne, konstruktywistyczne, pragmatyczne i transformatywne.

Nurt postpozytywistyczny to historycznie pierwsza szkoła, zapoczątkowana przez opisywanego powyżej Ralpha Tylera i jego uczniów, w szczególności Lee Cronbacha. Na poziomie epistemologicznym przedstawiciele tego nurtu zakładają możliwość obiektywnego poznania z wykorzystaniem (choć nie zawsze) narzędzi z warsztatu badaczy ilościowych. Współczesne modele teoretyczne to przede wszystkim konstrukt zaproponowany przez Daniela Stufflebeama, wyprowadzony z pojęć: *context, input, process i product*, a opisywany akronimem CIPP (Stufflebeam, Coryn, 2014; Stufflebeam, 2000, s. 279; Stufflebeam, Shinkfield, 1984). Stufflebeam chciał wyjść poza mierzenie efektu, włączając w stworzony przez siebie model czynniki kontekstowe, wpływające na potrzeby i możliwości oraz sposób określania celów; wkład w momencie podejmowania interwencji i wreszcie sam przebieg procesu. Ewaluacje prowadzone w nurcie postpozytywistycznym charakteryzują się silnym osadzeniem teoretycznym (*theory-driven evaluation*). W praktyce oznacza to pogłębioną wiedzę na temat problemów, które ma rozwiązać dana interwencja, jej założeń i struktury, kontekstu społecznego oraz kulturowego, w jakim jest wdrażana (Donaldson, 2022).

Nurt konstruktywistyczny w ewaluacji rozwijał się od późnych lat 70. ubiegłego wieku. Wydaje się, że jego źródła szukać należy przede wszystkim w konstruktywistycznie zorientowanej socjologii. Ewaluacja konstruktywistyczna dzieli ontologiczne i epistemologiczne założenia socjologii spod znaku Bergera i Luckmanna: przekonanie o wielości wytwarzanych społecznie rzeczywistości, znaczeniu interakcji między ewaluatorem a badanymi, czy konieczności urefleksyjniania perspektywy badacza. Od strony metodologicznej sięgano przede wszystkim po strategie jakościowe, wykorzystując ich dialogiczny charakter, dający możliwość osadzenia wniosków w szerszym kontekście (Mertens, 2015, s. 11). W ewaluacjach zorientowanych konstruktywistycznie widać fenomenologiczny postulat wnikięcia w mikroświaty badanych (Schütz, 2012). W konsekwencji ewaluator, a mówiąc ściśle jego lub jej wiedza, traci uprzywilejowany status. Wnioski są kompromisem wielu perspektyw, osiągniętym w dialogicznym procesie (Simons, 1987). Do najważniejszych przedstawicieli tego nurtu można zaliczyć Roberta Stake'a, Ernesta House'a, Jennifer Green, Egona Gubę, Yvonne Lincoln, Helen Simons czy Barry'ego MacDonalda (Mizerek, 2017, s. 43).

Nurt nazwany przez Donnę Mertens transformatywnym, a wcześniej emancypacyjnym, jest zróżnicowany wewnętrznie, gdyż czerpie z dorobku różnych teorii krytycznych: feministycznej, queer, krytyki postkolonialnej, minority studies i in. Elementem łączącym autorów i autorki z tego nurtu, a zarazem pozwalającym na jego wyodrębnienie, nie jest tym razem szczególna orientacja ontologiczna, epistemologiczna czy metodologiczna, lecz aksjologiczna. Podjęty przez konstruktivistów wątek sprawiedliwości społecznej zostaje tu ustawiony w centrum. Rolą ewaluacji jest dawanie głosu osobom i grupom niesłyszonym, marginalizowanym, a jednocześnie działanie na rzecz transformacji (stąd nazwa) ich statusu. Na pierwszy plan wysuwa się polityczny wymiar ewaluacji. Nurt ten obejmuje liczne podejścia, by wymienić za Henrykiem Mizerkiem najważniejsze: ewaluacja inkluzywna, demokratyczna, upelnocniająca, rozwojowa, feministyczna, czy transformatywna ewaluacja uczestnicząca (Mizerek, 2017, s. 45–46).

Nurt pragmatyczny, ostatnie z podejść wyróżnionych przez Mertens wiąże się, przede wszystkim, z pracami Michaela Q. Pattona i koncepcją UFE (*Utilization-Focused Evaluation*). Patton jako kluczowe wyzwanie stawia rzeczywiste wykorzystanie wniosków z ewaluacji w podejmowaniu decyzji i zmianach (Patton, 1997, s. 6). Ewaluacja, jego zdaniem, powinna być oceniana wyłącznie z perspektywy użyteczności dla precyzyjnie określonych interesariuszy (Patton, 1997, s. 20). Ewaluator wchodzi w rolę facylitatora, który pomaga interesariuszom zdefiniować swoje potrzeby i dopasować model ewaluacji w interaktywnym procesie. Jak widać, tu również możemy obserwować odejście od pozytywistycznego modelu ewaluatora – eksperta, którego wiedza ma charakter rozstrzygający. Warto jednak zauważyć, że w odróżnieniu od nurtu konstruktivistycznego „dialogiczność” dotyczy osób, które faktycznie mają korzystać z ewaluacji, nie wszystkich interesariuszy. Pragmatyzm obejmuje też rozwiązania metodologiczne i osadzenie teoretyczne. Wszystko podporządkowane jest użyteczności. Tym samym dobór strategii badawczych będzie zależał od specyfiki danego projektu. W rezultacie pragmatyki są orędownikami stosowania strategii mieszanych (Mertens, 2015, s. 39; zob. Mertens, Hesse-Biber, 2013).

Jak każda typologia, tak i propozycja Donny Mertens może budzić wątpliwości. Czy pragmatyzm jest w istocie odrębnym „paradygmatem”? Gdzie jest miejsce na modele z nurtu partycypacyjnego lub sięgające do filozofii postmodernizmu, poststrukturalizmu czy dekonstrukcji? Poza wymienionymi wątpliwościami typologia Mertens ma w mojej opinii zaletę polegającą na tym, że podkreśla różnice między najważniejszymi „paradygmatami” we współczesnej ewaluacji. Pozwala to nie tylko na zrozumienie stanu heterodoksji, ale być może przekroczenie progu heterogeniczności.

Ostatecznie w każdym z nurtów znajdziemy twierdzenia, z którymi (zapewne) każdy ewaluator-praktyk się zgodzi:

- Należy dążyć do rzetelnego i obiektywnego zbierania danych i wnioskowania;
- Warto uwzględniać perspektywy różnych interesariuszy, szczególnie wykluczonych;
- Warto przeprowadzać tylko takie ewaluacje, które mają znaczenie dla interesariuszy i przekładają się na działanie organizacji.

Czego uczy historia rozwoju dyscypliny z punktu widzenia nauczyciela-praktyka? Po pierwsze pokazuje, że wartościowe badanie nie musi obejmować skomplikowanych, czasochłonnych i trudnych do urzeczywistnienia w szkole strategii badawczych, choć konieczny jest metodologiczny rygor. Po drugie umożliwia włączenie w procesy decyzyjne w szkole osób z grup, których głos często bywa niesłyszalny – uczniów, w szczególności pochodzących z grup defaworyzowanych lub z niepełnosprawnościami, pracowników niepedagogicznych czy technicznych. Po trzecie historia dyscypliny poszerza obszary oddziaływania ewaluacji. Już nie tylko jakość procesu nauczania czy zarządzania tym procesem, na dowolnym poziomie, ale też upodmiotowienie aktorów społecznych obecnych w życiu szkoły, a nawet – czas na mocne sformułowanie – zmiana społeczna. Po czwarte wreszcie, złożona historia dyscypliny pozwala zaobserwować, że wybory dokonywane podczas projektowania ewaluacji to nie tylko kwestia względnie prostych decyzji metodologicznych, lecz przede wszystkim, mówiąc językiem hermeneutyki, ontologicznych, a często również aksjologicznych. Ich uświadomienie jest, by pozostać przy pojęciach Hansa-Georga Gadamera (2004), sposobem na ograniczenie przesądów w procesie poznania.

Przyjrzyjmy się kwestii wyodrębnienia ewaluacji jako samodzielnej dyscypliny akademickiej. „Jestem naukowcem-ewaluatorem, zajmuję się nauką ewaluacją” brzmiało credo Micheala Q. Pattona w głośnym tekście z 2018 r. (Patton, 2018, s. 183). W tym samym tekście Patton definiuje nauką ewaluację jako:

[...] systematyczne badanie zalet, wartości, użyteczności i znaczenia przedmiotu ewaluacji z wykorzystaniem procedur dowodzenia naukowego, które obejmują stosowanie logiki, transparentność metod, poddawanie wyników weryfikacji oraz dostarczanie dowodów i wyraźnych uzasadnień wspierających oparte na empirii interpretację, wartościowanie i sądy (Patton, 2018, s. 187).

Czy w istocie można mówić o ewaluacji jako odrębnej dyscyplinie naukowej?

Michael Scriven (2004, s. 186) zwraca uwagę na cztery kryteria, pozwalające na mówieniu o interdyscyplinarnej ewaluacji jako odrębnej nauce. Pierwszym z nich jest specyficzny obszar zainteresowań – określanie znaczenia lub wartości działań, logicznie odrębny od innych dyscyplin, podejmujących próby opisu, klasyfikacji, wyjaśnienia itd. Drugim – aparat metodologiczny obejmujący wartościowanie. Trzecim – szerokość podejmowanych zagadnień. Po czwarte wreszcie „produkty” ewaluacji mają istotne znaczenie społeczne i intelektualne. Ewaluacja, niezależnie

od, opisanego wyżej, rozwoju teorii charakteryzuje się wyodrębnioną pojęciowością. Wyszła tym samym poza kategorię narzędzia wspierającego podejmowanie decyzji (Scriven, 2004, s. 188).

Scriven zawiesza poprzeczkę jeszcze wyżej, zaliczając ewaluację do grupy transdyscyplin. Nie chodzi mu o interdyscyplinarność, która dla ewaluacji – wyrosłej z pedagogiki, psychologii, socjologii, ekonomii i innych nauk społecznych – jest cechą immanentną. Nauki transdyscyplinarne są uprawiane w ramach innych dyscyplin (pedagogiki, socjologii i in.), służą ich rozwojowi, ale jednocześnie zachowują swoją odrębność. Do najbardziej oczywistych przykładów transdyscyplin można zaliczyć logikę czy statystykę (Scriven, 2004, s. 186; zob. Mizerek, 2017, s. 25). Transdyscypliny, jak pokazuje Scriven na przykładzie biostatystyki i technologii komunikacyjnych, mają istotny wkład w rozwój poszczególnych dyscyplin nauki, od nauk ścisłych czy inżynierskich po społeczne i humanistyczne (Scriven, 2013, s. 167–168).

Jak widać droga, którą przeszła ewaluacja jest dość długa. Można wyodrębnić kilka składowych tego procesu: rozwój, często stojących wobec siebie w mniejszej lub większej opozycji, nurtów teoretycznych; odejście od koncentracji na efektach i uwzględnienie z jednej strony politycznego wymiaru procesu ewaluacji i różnych pozycji zajmowanych przez ewaluatorów, zaś z drugiej odpowiedzialności za upodmiotowienie interesariuszy. Złożoność teoretyczna i użyteczność w różnych obszarach wiedzy naukowej i działalności praktycznej uzasadnia przypisanie ewaluacji statusu transdyscypliny. Dla dzisiejszych ewaluatorów historia może być przede wszystkim doskonałym ćwiczeniem intelektualnym, pokazującym arbitralność niektórych decyzji metodologicznych i teoretycznych, pozwalającym na znaczącą redukcję przesądów w procesie poznania.

Czy ewaluacja może zmienić szkołę?

W tej części ukazę problem ewaluacji i zmiany w szkole, odwołując się do badań, które pokazują oddziaływanie ewaluacji na jakość nauczania. Badania te skupiają się na ogół na ewaluacjach zewnętrznych, prowadzonych przez jednostki administracji publicznej, których polskim odpowiednikiem są kuratoria oświaty. Wynika to zapewne z wysokiego poziomu instytucjonalizacji, oczekiwania rozliczalności scentralizowanych systemów ewaluacji oraz ich międzynarodowego usieciowienia (Grek i in., 2013). Kluczowym wyzwaniem w analizach jest określenie obszaru zmiany. W literaturze część analiz skupia się na wykorzystaniu wyników badania ewaluacyjnego, co jest dobrym wskaźnikiem użyteczności, ale nie pokazuje bezpośrednio zmian w jakości pracy szkół. Ponadto samo pojęcie „wykorzystania” wyników ewaluacji jest polisemiczne (Mark, Henry, 2004, s. 39). Przykładowo

w przeglądzie systematycznym, zbierającym publikacje z lat 1986–2005, wykazano zaangażowanie interesariuszy jako czynnik facylitujący wykorzystanie wyników ewaluacji. Innymi słowy, im większe włączenie zamawiających w proces, im lepszy kontakt między nimi a zespołem badawczym, tym większa użyteczność (Johnson i in., 2009). Wniosek ten wydaje się zgodny ze zdrowym rozsądkiem, warto jednak zauważyć, że przegląd systematyczny bez uwzględnienia tzw. *grey literature* może dać – w przypadku analizy projektów ewaluacyjnych – niepełne rezultaty. Po pierwsze z uwagi na publikowanie zdecydowanej większości raportów poza recenzowanymi czasopismami. Po drugie proces selekcji i recenzowania artykułów może ograniczać reprezentację publikacji pokazujących nieudane aplikacje. Po trzecie istotna wydaje się także ograniczona liczba krajów uwzględnionych w przeglądach.

Marielle Klerks (2013) w przeglądzie systematycznym dotyczącym wpływu ewaluacji zewnętrznej (*school inspection*) na jakość pracy szkół zaproponowała trzy obszary: (1) funkcjonowanie szkoły; (2) działania nauczycieli; oraz (3) osiągnięcia szkolne uczniów. Analiza wykazała poprawę jakości funkcjonowania szkół i – w przeważającej części badań – zmianę praktyk nauczycieli, a także niewielki lub zerowy związek z poprawą wyników (Klerks, 2013, s. 9–15). Autorka podkreśla, że „nie ma dowodów na to, że ewaluacja zewnętrzna automatycznie prowadzi do poprawy jakości edukacji” (Klerks, 2013, s. 16). Nelson i Ehren (2014) uzupełnili klasyfikację zaproponowaną przez Klerks o wpływ na uruchomienie/ doskonalenie ewaluacji wewnętrznej.

Do autoewaluacji wrócimy za chwilę; wnioski dla trzech obszarów zidentyfikowanych przez Klerks są w zasadzie tożsame. Nelson i Ehren przedstawiają liczne czynniki moderujące wpływ ewaluacji, leżące zarówno wewnątrz szkoły (np. model liderstwa czy postawy wobec ewaluacji), ewaluatorów (np. jakość rekomendacji, sposób przekazywania wyników, tworzenie rankingów) czy władz lokalnych (np. wsparcie w rozwoju szkół). Warto dodać, że ten przegląd systematyczny wyszedł poza bazy recenzowanych tekstów.

Wniosek o nieznacznym pozytywnym wpływie ewaluacji na wyniki egzaminów zewnętrznych został potwierdzony w metaanalizie badań z 67 amerykańskich szkół (Schueler i in., 2022). Schueler i inni zaobserwowali wzrost wyników z matematyki i brak zmiany dla języka angielskiego dla całej próby; po zawężeniu do grup defaworyzowanych siła wpływu wzrosła. Badania europejskie nie pokazują tego zróżnicowania, wskazują za to na silny wpływ na wyniki z matematyki i, nieco słabszy, z języka ojczystego (Hofer i in., 2020, s. 12).

Niestety brakuje długoterminowych, longitudinalnych badań. Jak zauważają Ehren i Shackleton (2016) zmiana nie musi mieć charakteru linearnego. Ich badanie pokazało jej dyfuzyjny i cykliczny charakter. Co więcej, pewną wadą tekstów syntetyzujących wpływ ewaluacji na jakość pracy szkół w różnych krajach jest brak

uwzględnienia zróżnicowania systemów ewaluacyjnych: ich formuły, częstotliwości, konsekwencji i związków z ewaluacją wewnętrzną (por. Ehren i in., 2015). Jeśli chodzi o efekty ewaluacji zewnętrznej w polskim systemie oświaty, dostępne dane są ograniczone. W badaniu z 2016 r. na próbie losowej dyrektorów szkół i przedszkoli uzyskano 86% wskazań na zmiany wprowadzone na skutek ewaluacji zewnętrznej, jednak jedynie jedna czwarta respondentów potrafiła podać konkretne przykłady. Najczęściej wskazywano na zmiany w sposobie prowadzenia dokumentacji (sic!), dalej współpracy z otoczeniem, promocji placówki, śledzenia losów absolwentów, analizy wyników egzaminów zewnętrznych i współpracy z rodzicami. Zmiany wykraczające poza kwestie biurokratyczne zostały wskazane przez 17% badanych dyrektorów. Nietrudno jednocześnie zauważyć, że obszary te są niemal literalnym powtórzeniem nazw standardów ewaluacji, opisanych w obowiązującym wówczas prawie oświatowym, co skądinąd pokazuje modelujący potencjał ewaluacji zewnętrznej. Zmiany dotyczące bezpośrednio nauczania – nowe metody, indywidualizacja, praca projektami itd. – zostały opisane przez 3% badanych (Kołodziejczyk, Walczak, 2016, s. 743). Późniejsze badanie na nielosowej próbie niemal 700 dyrektorów (Sury i in., 2022) pokazało bardziej stonowane opinie: 49% dostrzegło poprawę przebiegu procesów dydaktycznych, 43% badanych wskazało na poprawę efektów nauczania, zaś 43,5% badanych podkreśliło poprawę efektów działalności wychowawczej. Ponad połowa dyrektorów wskazała na zwiększenie zaangażowania nauczycieli, ponad jedna trzecia – rodziców w procesy decyzyjne, zaś niespełna jedna trzecia – włączenia uczniów w procesy dydaktyczne i wychowawcze.

Warto wreszcie wspomnieć o efekcie odwrócenia, by posłużyć się terminem Raymonda Boudon (2008). De Wolf i Janssens zaproponowali trójdziałną typologię: (1) świadome strategiczne działania i gry („malowanie trawników”, pomijanie prawdy, wpływ na dobór badanych itp.); (2) nieświadome działania strategiczne (nacisk na formalizację i procedury, uczenie „pod testy” lub oczekiwania ewaluatorów, zawężenie perspektyw, koncentracja na wskaźnikach itp.); (3) inne efekty (stres, spoczywanie na laurach, zmiany w otoczeniu szkoły) (de Wolf, Janssens, 2007, s. 383). Znajduje to potwierdzenie w badaniach porównawczych: Ehren i współpracownicy (2015) na podstawie badań w sześciu europejskich krajach wskazują na pojawiającą się wśród dyrektorów tendencję do zawężania obszarów nauczania, zaś wśród nauczycieli – na unikanie eksperymentowania z nowymi metodami.

W stronę mikroewaluacji nauczycielskiej

W ostatniej części chciałbym zaproponować możliwe wykorzystanie tradycji ewaluacji w codziennej praktyce nauczycielskiej. Patrząc na obszar badań ewaluacyjnych, możemy wyróżnić kilka poziomów: poczynając od systemów obejmu-

jących całe kraje lub regiony, przez projekty wielostanowiskowe, dalej lokalne, dla których obszarem badania będzie konkretna miejscowość, dzielnica lub grupa szkół, po działania ewaluacyjne realizowane na poziomie pojedynczej szkoły lub placówki i wreszcie klasy bądź grupy.

O ile realizacja bardziej rozbudowanych projektów pozostaje (w większości) domeną wyspecjalizowanych ewaluatorów, badania na poziomie mikro – szkoły lub klasy – mają zazwyczaj charakter ewaluacji wewnętrznej, tj. prowadzonej przez osoby pracujące w danej organizacji (Scriven, 1991, s. 197–198). W literaturze proces taki zwykle się określać autoewaluacją (*selfevaluation*), wcześniej samoewaluacją (Mizerek, 1997). Definicje autoewaluacji są dość intuicyjne: jest to podejmowana przez pracowników szkoły działalność, mająca na celu opis i wartościowanie procesów zachodzących w szkole, (zazwyczaj) w obszarach i zgodnie z kryteriami przyjętymi przez tych pracowników, wykorzystywane dalej w podejmowaniu decyzji i podnoszeniu jakości pracy placówki, zespołów nauczycielskich lub pojedynczych nauczycieli (zob. np. Airasian, Gullickson, 1997; Schildkamp i in., 2012; Scriven, 1991, s. 327; Sury, 2020, s. 39). Zastrzeżenie co do wyboru obszarów i kryteriów wynika z powiązania w niektórych krajach systemowej ewaluacji zewnętrznej z autoewaluacją. W takich sytuacjach autoewaluacja może być projektowana zgodnie z wymaganiami instytucji prowadzącej zewnętrzną ewaluację. Przykładem jest wprowadzony w 2005 r. przez brytyjski OFSTED Self Evaluation Form. Samo powiązanie systemów niesie ze sobą ryzyko modelowania wyników pod oczekiwania wizytatorów (Nelson, Ehren, 2014, s. 5). Między obydwojema systemami istnieją liczne napięcia, łącznie z formułowaniem rozbieżnych wniosków (Vanhoof, van Petegem, 2010). Jednocześnie funkcjonujący system ewaluacji zewnętrznej przekłada się na większą intensywność działań ewaluacyjnych podejmowanych przez szkoły (Hofer i in., 2020, s. 12), a przy zapewnieniu odpowiedniego wsparcia szkołom – na poprawę jakości autoewaluacji (Blok i in., 2008). W istocie powodzenie takich powiązanych systemów zależy od rozkładu priorytetów i oczekiwań w całym systemie edukacyjnym. Jak zauważa John MacBeath, rozkład ten można uporządkować w trzech wymiarach: odgórne vs oddolne wprowadzanie zmiany w edukacji, wsparcie szkół vs formułowanie oczekiwań, nacisk na ewaluację zewnętrzną vs autoewaluację (MacBeath, 1999, s. 2)

Co wiemy o efektach autoewaluacji? Przegląd literatury przygotowany przez Nelsona i współautorów (2015) wskazuje na pozytywne oddziaływanie na doświadczenia uczenia się i osiągnięcia uczniów, co zostało zweryfikowane w oparciu o wyniki testów i zewnętrzne pomiary. Ewaluacja wewnętrzna uwrażliwia nauczycieli na warunki uczenia się, co przekłada się na działania na rzecz poprawy jakości funkcjonowania szkoły. Ponadto warsztat nauczycieli poszerza się o nowe metody zbierania danych, takie jak obserwacje lekcji, analiza prac uczniowskich czy

zbieranie informacji zwrotnych od uczniów i ich rodziców. Jest też istotną składową rozwoju zawodowego. Osobną wartością dodaną jest uwspólnienie planów i działań w przypadku projektów zespołowych (Kołodziejczyk, Kołodziejczyk, 2015, s. 117; Nelson i in., 2015, s. 24). Pozytywne efekty przeważają, niemniej w sytuacjach powiązania ewaluacji wewnętrznej z zewnętrzną może dojść do pojawienia się niekorzystnych procesów, takich jak zawężenie celów edukacyjnych. Kluczowym czynnikiem jest model przywództwa w szkole (Nelson i in., 2015, s. 25; zob. Mazurkiewicz, 2011, s. 199). Patrząc na nieco innym poziomie: autoewaluacja jest ważną składową autonomii szkół i ich uprawomocnienia (*empowerment*) w zdecentralizowanych systemach oświaty (Nevo, 2001, s. 97).

Podsumowanie

W artykule podjęto próbę uzasadnienia obecności ewaluacji edukacyjnej w kształceniu przyszłych nauczycieli i nauczycielek opierając się na dwóch wiązках argumentów: dojrzałości dyscypliny do stania się autonomicznym przedmiotem nauczania na studiach wyższych oraz korzyści z wdrożenia proponowanych przez nią rozwiązań w nauczycielskiej praktyce. Wydaje się, że dorobek teoretyczny ewaluacji i proces jej wyodrębnienia jako zróżnicowanej wewnętrznie transdyscypliny naukowej uzasadnia jej obecność w dydaktyce akademickiej (Maciejewska, 2014). Korpus wiedzy i jej złożoność wymagają uczenia ewaluacji w ramach osobnych kursów, nie jako składowej przedmiotów metodologicznych. Znajomość teorii ewaluacji i historii jej rozwoju pozwala na podejmowanie świadomych wyborów metodologicznych, a także na autorefleksję dotyczącą pozycji zajmowanej przez ewaluatora, jej osadzenia w relacjach władzy i wreszcie – odpowiedzialności za upodmiotowienie poszczególnych grup interesariuszy. Patrząc od strony praktycznej, dostępne badania pokazują, że wdrożenie ewaluacji w praktykę instytucjonalną i indywidualną, przy zapewnieniu odpowiednich warunków brzegowych, przekłada się na poprawę jakości nauczania. Szczególnie interesujące wydają się tu rozwiązania z poziomu mikro, łatwiejsze w zastosowaniu w szkolnej rzeczywistości, akcentujące autonomię oraz refleksyjność nauczycieli i nauczycielek (zob. Maciejewska, 2018, s. 162).

Autoewaluacja może być postrzegana jako narzędzie refleksji na indywidualnym, zespołowym lub organizacyjnym poziomie. Wymaga to odpowiednio ukształtowanej kultury szkoły (Schildkamp i in., 2012, s. 21) i wpisuje się w ideę nauczyciela jako refleksyjnego praktyka, ze wszystkimi wyzwaniami wiążącymi się z tą koncepcją (por. Mizerek, 2021; Sury, 2020, s. 11). Autoewaluacja jest jedynie narzędziem, ale jak pokazują zestawione powyżej badania, może przyczynić się do uruchomienia korzystnych procesów w szkole i podniesienia jakości pracy nauczycieli.

Jednocześnie jest w mniejszym stopniu zależna od zmian systemowych i prawnych. Nie będzie prawdopodobnie miała wpływu na decyzje podejmowane ogólnie, nie pozwoli na wygenerowanie dużych zbiorów danych, pozwalających na kreowanie polityk oświatowych na poziomie makro. Dzieje się „tu i teraz” i przy sprzyjających warunkach może przyczynić się do zmiany na poziomie mikro. Ostatecznie, jak zauważa Helen Simons, nawet jeśli priorytety, programy czy struktura nauczania są zarządzane centralnie, „edukacja rozgrywa się lokalnie” (Simons, 1997, s. 67).

References

- Airasian, P. W., Gullickson, A. R. (1997). *Teacher Self-Evaluation Tool Kit*. Thousand Oaks: Corwin Press.
- Alkin, M., Christie, C. (2004). An evaluation theory tree. W: M. Alkin (red.), *Evaluation Roots. Tracing Theorists' Views and Influences*. London: SAGE Publications.
- Blok, H., Slegers, P., Karsten, S. (2008). Looking for a balance between internal and external evaluation of school quality: evaluation of the SVI model. *Journal of Education Policy*, 23(4), 379–395.
- de Wolf, I. F., Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education*, 33(3), 379–396.
- Donaldson, S. (2022). *Introduction to theory-driven program evaluation : culturally responsive and strengths-focused applications*. New York: Routledge, Taylor & Francis Group.
- Boudon, R. (2008). *Efekt odwrócenia*. Warszawa: Wydawnictwo Oficyna Naukowa.
- Ehren, M. C. M., Gustafsson, J. E., Altrichter, H., Skedsmo, G., Kemethofer, D. i Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, 51(3), 375–400.
- Ehren, M. C. M., Shackleton, N. (2016). Mechanisms of change in dutch inspected schools: Comparing schools in different inspection treatments. *British Journal of Educational Studies*, 64(2), 185–213.
- Gadamer, H.-G. (2004). *Prawda i metoda*. Warszawa: Wydawnictwo Naukowe PWN.
- Grek, S., Lawn, M., Ozga, J., i Segerholm, C. (2013). Governing by inspection? European inspectorates and the creation of a European education policy space. *Comparative Education*, 49(4), 486–502.
- Guba, E., Lincoln, Y. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: SAGE Publications.
- Hildreth, G. (1979). *Bibliography of mental tests and rating scales*. New York: The Psychological Corp.
- Hofer, S. I., Holzberger, D. i Reiss, K. (2020). Evaluating school inspection effectiveness: A systematic research synthesis on 30 years of international research. *Studies in Educational Evaluation*, 65, 100864.
- Hogan, L. (2007). The Historical Development of Program Evaluation: Exploring Past and Present. *Online Journal for Workforce Education and Development*, 2(4), 1–14.
- Jaskuła, S. (2018). *Ewaluacja społeczna w edukacji*. Kraków: Oficyna Wydawnicza Impuls.
- Johnson, K., Greenesid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on Evaluation Use. *American Journal of Evaluation*, 30(3), 377–410.

- Klerks, M. (2013). *The effect of school inspections: a systematic review*. <http://schoolinspections.eu/impact/wp-content/uploads/downloads/2013/12/ORD-paper-2012-Review-Effect-School-Inspections-MKLERKS.pdf>
- Kołodziejczyk, J., Kołodziejczyk, J. (2015). Ewaluacja własnej pracy nauczycieli – obszary, rozumienie, współpraca. *Forum Oświatowe*, 27(2), 115–130.
- Kołodziejczyk, J., Walczak, B. (2016). Meta-evaluation of the progress and effects of external evaluation – experiences after six years of the pedagogical supervision reform. W: L. Gómez Chova, A. López Martínez i I. Candel Torres (red.), *EDULEARN16 Proceedings* (s. 882–892). Barcelona, Spain: IATED Academy.
- Korporowicz, L. (2011). Zmienne losy polskiej ewaluacji. Pomiędzy nadzieją, animacją i konfuzją. W: B. Niemierko, M. K. Szmigiel (red.), *Ewaluacja w edukacji. Koncepcje. Metody. Perspektywy* (s. 21–28). Kraków: Grupa Tomami.
- Kuhn, T. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.
- MacBeath, J. (1999). *Schools must speak for themselves. The case for school self-evaluation*. London: Routledge.
- Maciejewska, M. (2014). Badania ewaluacyjne w edukacji i ich konteksty w akademickim kształceniu nauczycieli i pedagogów. *Opuscula Sociologica*, 4(10), 31–43.
- Maciejewska, M. (2018). Badania ewaluacyjne w edukacji jako źródło wiedzy pedagogicznej. *Forum Pedagogiczne*, 8(2), 153–165.
- Madaus, G. (2004). Ralph W. Tyler's contribution to program evaluation. W: M. Alkin (red.), *Evaluation Roots. Tracing Theorists' Views and Influences* (s. 69–79). Thousand Oaks, CA: SAGE Publications.
- Mark, M. M., Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35–57.
- Mazurkiewicz, G. (2011). *Przywództwo edukacyjne. Odpowiedzialne zarządzanie edukacją wobec wyzwań współczesności*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Mertens, D. (2015). *Research and Evaluation in Education and Psychology. Integrating Diversity With Quantitative, Qualitative, and Mixed Methods*. London: SAGE.
- Mertens, D. M., Hesse-Biber, S. (2013). *Mixed Methods and Credibility of Evidence in Evaluation: New Directions for Evaluation*. New York: Wiley.
- Mizerek, H. (1997). *Ewaluacja w szkole. Wybór tekstów*. Olsztyn: Wydawnictwo MG.
- Mizerek, H. (2016). Ewaluacja edukacyjna w Polsce. Trajektorie, perspektywy i dylematy rozwoju. *Zarządzanie Publiczne*, 1(13), 1–16.
- Mizerek, H. (2017). *Ewaluacja edukacyjna. Interdyskursywne dialogi i konfrontacje*. Kraków: Oficyna Wydawnicza Impuls.
- Mizerek, H. (2021). *Refleksja krytyczna w edukacji i pedagogice. Misja (nie)wykonalna?* Kraków: Oficyna Wydawnicza Impuls.
- Nelson, R. i Ehren, M. (2014). *Review and synthesis of evidence on the (mechanisms of) impact of school inspection*. <http://schoolinspections.eu/impact/wp-content/plugins/download-monitor/download.php?id=29>
- Nelson, R., Ehren, M. i Godfrey, D. (2015). *Literature Review on Internal Education*. <http://schoolinspections.eu/wp-content/uploads/downloads/2015/09/Literature-review-internal-evaluation.pdf>
- Nevo, D. (2001). School evaluation: Internal or external? *Studies in Educational Evaluation*, 27, 95–106.

- Niemierko, B. (2019). Czy diagnostyka edukacyjna i ewaluacja w edukacji zachowują odrębność? *Teraźniejszość – Człowiek – Edukacja*, 4(64), 1–13.
- Nowotniak, J. (2019). *Ewaluacja edukacyjna. Społeczne światy ewaluatorów*. Kraków: Oficyna Wydawnicza Impuls.
- Patton, M. Q. (1997). *Utilization-focused evaluation :the new century text*. London: SAGE Publications.
- Patton, M. Q. (2018). Evaluation Science. *American Journal of Evaluation*, 39(2), 183–200.
- Rice, J. M. (1897). The futility of the spelling grind. *The Forum*, 409–419.
- Schildkamp, K., Vanhoof, J., van Petegem, P., & Visscher, A. (2012). The use of school self-evaluation results in the Netherlands and Flanders. *British Educational Research Journal*, 38(1), 125–152. <https://doi.org/10.1080/01411926.2010.528556>
- Schueler, B. E., Asher, C. A., Larned, K. E., Mehrotra, S. i Pollard, C. (2022). Improving low-performing schools: A meta-analysis of impact evaluation studies. *American Educational Research Journal*, 59(5), 975–1010.
- Schutz, A. (2012). *O wielości światów: szkice z socjologii fenomenologicznej*. Kraków: Zakład Wydawniczy Nomos.
- Scriven, M. (1966). *The methodology of evaluation*. Social Science Education Consortium.
- Scriven, M. (1991). *Evaluation Thesaurus*. London: SAGE Publications.
- Scriven, M. (1996). The theory behind practical evaluation. *Evaluation*, 2(4), 393–404.
- Scriven, M. (2004). Reflections. W: M. Alkin (red.), *Evaluation Roots. Tracing Theorists' Views and Influences* (s. 183–195). London: SAGE Publications.
- Scriven, M. (2013). Conceptual revolutions in evaluation. Past, present, and future. W: M. Alkin (red.), *Evaluation Roots. A Wider Perspective of Theorists' Views and Influences* (s. 167–179). Chicago: Rand McNally & Co, Chicago.
- Simons, H. (1987). *Getting to know schools in a democracy: the politics and process of evaluation*. New York: The Falmer Press.
- Simons, H. (1997). Samoewaluacja szkoły. W: H. Mizerek (red.), *Ewaluacja w szkole. Wybór tekstów* (s. 59–71). Olsztyn: Wydawnictwo MG.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record: The Voice of Scholarship in Education*, 68(7), 1–15. <https://doi.org/10.1177/016146816706800707>
- Stufflebeam, D., & Coryn, C. (2014). *Evaluation Theory, Models, and Applications*. San Francisco: Jossey-Bass & Pfeiffer Imprints, Wiley.
- Stufflebeam, D. L. (2000). The CIPP model for evaluation. W: D. Stufflebeam, G. Madaus i T. Kelleghan (red.), *Evaluation Models* (s. 279–317). Warszawa: Kluwer Academic Publishers.
- Stufflebeam, D., Shinkfield, A. (1984). *Systematic Evaluation. A Self-Instructional Guide to Theory and Practice*. Warszawa: Kluwer Academic Publishers.
- Sury, Z. (2020). *Ewaluacja w pracy nauczycieli: konteksty zawodowe i tożsamościowe*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Sury, Z., Smoter, K., Kołodziejczyk, J. (2022). *Dyrektorzy i dyrektorki szkół podstawowych i ponadpodstawowych wobec ewaluacji i kontroli w nadzorze pedagogicznym. Raport z badania*. Kraków: Instytut Spraw Publicznych Uniwersytetu Jagiellońskiego.
- Vanhoof, J., van Petegem, P. (2010). Evaluating the quality of self-evaluations: The (mis)match between internal and external meta-evaluation. *Studies in Educational Evaluation*, 36(1–2), 20–26.