Edukacja 2012, 4(120), 120-128
ISSN 0239-6858

Polityka edukacyjna w perspektywie porównawczej

Anna Wojciuk*

Recenzja raportu: Mona Mourshed, Chinezi Chijioke i Michael Barber, *Jak najlepiej doskonalone systemy szkolne na świecie stają się jeszcze lepsze*. Warszawa 2012, Centrum Edukacji Obywatelskiej (wydanie oryginalne: *How the world's most improved school systems keep getting better*. McKinsey & Company, 2010).

Politologia w perspektywie porównawczej

Od starożytności kolejne pokolenia ludzi ciekawiły podobieństwa i różnice między systemami i procesami politycznymi, a także zastanawiały implikacje odmiennych sposobów urządzenia życia społecznego. Najsłynniejsze politologiczne studium porównawcze z czasów starożytnych pozostawił po sobie Arystoteles, który w *Polityce* dokonał klasyfikacji ustrojów na monarchię, arystokrację i demokrację oraz ich wersje skorumpowane: tyranię, oligarchię i anarchię. Dzieło to jest uważane za prekursorskie w dziejach politologicznych studiów porównawczych (*comparative politics*). Systematyczna i rygorystyczna komparatystyka rozwinęła się dopiero w II połowie XX w. W tym czasie badaczy najbardziej nurtowały zależności między rozwojem ekonomicznym i demokracją, warunki sprzyjające wybuchom rewolucji, transformacje demokratyczne, ruchy społeczne, przestrzeganie praw człowieka, a także synergia między różnymi formami instytucji państwowych a efektywnością rządów demokratycznych.

Rozwojowi tej gałęzi nauk o polityce towarzyszyła przez cały czas istotna kontrowersja. Sceptycy pytali, na ile w ogóle międzykrajowe porównania są możliwe i sensowne. Na poziomie fundamentalnych założeń teoretycznych wymagają one bowiem akceptacji zasady uniwersalizmu – czyli zgody co do tego, że wszystkie porównywane przypadki da się sprowadzić do wspólnego mianownika. Komunitarianie i inni zwolennicy relatywizmu kulturowego uważają to założenie za fałszywe. Ich zdaniem różnych państw, kultur, cywilizacji nie da się porównać, nie ma bowiem żadnych uniwersalnych norm, wartości i reguł, a zatem wspólny mianownik nie istnieje (MacIntyre, 1971). Nie wnikając w szczegóły skomplikowanej i wielowątkowej debaty między uniwersalistami a relatywistami, warto zwrócić uwagę, że ta kontrowersja filozoficzna ma poważne implikacje dla nauk empirycznych, a poszukiwanie wspólnego mianownika dla porównywanych przypadków pozostaje dla komparatystyki jednym z większych wyzwań.

Porównawczym studiom politologicznym towarzyszyć mogą różnorodne cele: po pierwsze, opis szczególnie ciekawego, wyjątko-

* Zespół Analiz Średniookresowych, Instytut Badań Edukacyjnych. Email: a.wojciuk@ibe.edu.pl

wego państwa lub grupy państw; po drugie, sklasyfikowanie różnorodnych przypadków w ramach ogólniejszych kategorii; po trzecie, testowanie hipotez dotyczących przebiegu procesów politycznych w ramach różnych kategorii państw i wreszcie po czwarte, prognozowanie przebiegu procesów politycznych na podstawie wcześniej dokonanych generalizacji (Landmann, 2008). W ostatnich dwóch dekadach, szczególnie w państwach zachodnich, rozwija się tzw. *evidence informed policy,* czyli polityka publiczna oparta na faktach. W ramach tego podejścia zakłada się, że badania naukowe mogą pomóc w tworzeniu skutecznych polityk publicznych i ich optymalnej implementacji (Davies, Nutley i Smith, 2000). Na potrzeby *evidence informed policy* zbiera się i analizuje dane ilościowe i jakościowe, poszukując takich interwencji w sferze publicznej, które przyniosłyby pożądany skutek. W przypadku implementowania jakiejś zmiany po raz pierwszy w państwie, które nie ma własnych doświadczeń lub są one negatywne, poszukuje się dobrych praktyk za granicą. Niekiedy zamawiane są w tym celu przeglądy systematyczne, które pozwalają zapoznać się z rezultatami wszelkich badań dotyczących danego zagadnienia, także zagranicznych. Międzykrajowym porównaniom służą też coraz bardziej szczegółowe bazy danych zbierane przez państwa, organizacje międzynarodowe, pozarządowe, firmy doradcze. W dziedzinie polityki edukacyjnej przełomem dla porównywania różnych systemów oświaty było badanie umiejętności uczniów PISA. Przeglądy systematyczne oraz mniej rygorystyczne *policy papers* i raporty organizacji międzynarodowych często porównują interwencje publiczne dokonane w różnych systemach, oceniają ich efektywność, identyfikują dobre praktyki i na tej podstawie przedstawiają decydentom politycznym rekomendacje.

Tak popularne w ostatnich latach raporty międzynarodowe, *policy papers* i przeglądy systematyczne często są de facto przykładami studiów porównawczych polityk publicznych, realizują też zazwyczaj jeden lub kilka z tradycyjnych celów komparatystyki, takich jak opis przypadków, klasyfikacja państw i interwencji, testowanie hipotez na temat tego, jakie interwencje działają. Najciekawsze i szczególnie poszukiwane są jednak prognozy oraz rekomendacje. Omawiane dokumenty rzadko jednak odwołują się do metodologii *comparative politics* i niekiedy zdają się nie dostrzegać jej dorobku. Zjawisko to jest nie tylko interesujące poznawczo, ale także może mieć poważne konsekwencje praktyczne, ponieważ nieświadomość lub niezastosowanie się do fundamentalnych zasad metodologicznych może prowadzić do błędnych wniosków na poziomie badawczym, a także do decyzji politycznych, które są daleko słabiej uzasadnione, niż wskazują na to przywoływane na ich poparcie badania.

Interesującym studium przypadku w tej dziedzinie jest raport firmy doradczej McKinsey z 2010 r. *Jak najlepiej doskonalone systemy szkolne na świecie stają się jeszcze lepsze* (Mourshed, Chijioke i Barber, 2012). Warto wziąć ten tekst na warsztat z uwagi na duże międzynarodowe zainteresowanie dokumentem wśród decydentów politycznych, urzędników, liderów edukacyjnych, badaczy i praktyków. Przeszukiwanie internetu na pomocą Google daje prawie 21 800 odniesień do angielskiej wersji tej publikacji, natomiast wyszukiwanie wcześniejszego edukacyjnego raportu tej samej firmy z 2007 r. *How the world's best-performing schools* come out on *top* (Barber i Mourshed, 2007) daje aż 30 200 odniesień (14 czerwca 2012). Wśród nich jest duża grupa artykułów i raportów publikowanych przez agencje rządowe, instytuty naukowe, indywidualnych badaczy oraz organizacje pozarządowe. Frank Coffield stwierdza, że obydwa studia porównawcze firmy McKinsey, firmowane nazwiskiem Michaela Barbera, edukacyjnego lidera i au-

torytetu w tej dziedzinie, są tak cenne dla decydentów politycznych, ponieważ oferują jasne recepty wypowiadane z przekonaniem oraz redukują złożoność, a tym samym niepewność (Coffield, 2011). Analizując politykę edukacyjną Stanów Zjednoczonych i Wielkiej Brytanii, Coffield podaje liczne przykłady dokumentów państwowych, które się na te raporty powołują oraz decyzji o interwencjach publicznych inspirowanych omawianymi analizami. W Unii Europejskiej tekst ten przyjęto z większą ostrożnością (Coffield, 2011).

W 2012 r. Centrum Edukacji Obywatelskiej opublikowało polskie tłumaczenie raportu. Jest to dobra okazja do zastanowienia się nad międzykrajowymi porównaniami systemów edukacyjnych. Ich zalety są znane: stanowią ważne źródło inspiracji dla liderów zmian, a dostarczając informacji, umożliwiają konkurencję między systemami i tym samym dają impuls do reform. Międzykrajowe porównania są jednak dużym wyzwaniem metodologicznym i aby wnioski z nich wywiedzione były słuszne, a rekomendacje trafne, badania muszą spełniać szereg kryteriów warsztatowych.

Raport *Jak najlepiej doskonalone systemy szkolne na świecie stają się jeszcze lepsze* jako międzynarodowe studium porównawcze

Analiza przeprowadzona w omawianym raporcie ma charakter indukcyjny i można ją scharakteryzować według typowych elementów obecnych w politologicznych studiach porównawczych:

- wyjaśniane zróżnicowania: postęp w umiejętnościach uczniów mierzony wynikami badań międzynarodowych
- jednostki obserwacji: systemy szkolne
- miary: dla zmiennej zależnej (wyników uczniów) – wskaźniki ilościowe, dla zmiennych niezależnych (interwencje publiczne) – wskaźniki jakościowe
- teoria: podejście systemowe z założeniem linearnej, uniwersalnej dla wszystkich systemów logiki zmian
- metoda porównawcza: porównanie niewielkiej liczby przypadków i zastosowanie w tym celu *Most Different Systems Design* (MDSD)
- analiza: klastrowanie jako metoda poradzenia sobie z problemem zbyt dużej liczby zmiennych przy zbyt małej liczbie przypadków, pominięto zagadnienie ekwiwalencji przypadków, doboru próby dokonano po zmiennej zależnej, nie zabezpieczono się przed błędem pominiętej zmiennej
- obserwowalne implikacje studium: rekomendacje i prognoza, że inne państwa na podobnym etapie rozwoju: jeśli wdrożą podobne działania, osiągną analogiczne rezultaty (na podstawie: Mourshed, Chijioke i Barber, 2012, s. 19–20).

Niewątpliwie mocną stronę raportu stanowią zastosowane miary. Dla zmiennej zależnej zastosowano skalę wypracowaną już wcześniej przez znanego ekonomistę edukacji Ericka Hanusheka. Aby zmierzyć poprawę wyników uczniów, stworzono uniwersalną skalę przez znormalizowanie skal zastosowanych w kilku różnych międzynarodowych badaniach osiągnięć uczniów. Stosując metodologię Hanusheka, wyciągnięto porównywalne dane z różnych okresów i różnych testów, a następnie znormalizowano różne skale oceny w jedną, uniwersalną skalę. Dużą zasługą autorów raportu jest stworzenie indeksu dla zmiennych niezależnych, czyli interwencji publicznych w obszarze edukacji. W tym celu zidentyfikowano w sumie 575 interwencji w badanej grupie 20 państw.

Jedyną wspólną dla wszystkich przypadków zmienną niezależną jest istnienie lidera świadomego, że wyzwania mają charakter systemowy, a nie ograniczają się do ko-

rekty jednego elementu systemu. Informacje o samych interwencjach miały charakter jakościowy i były zbierane w trybie wywiadów indywidualnych oraz wniosków z obserwacji dokonanych podczas wizyt studyjnych we wszystkich badanych krajach. W trakcie tych wizyt przeprowadzono ponad 200 wywiadów z liderami i pracownikami systemów edukacyjnych. Interwencje uporządkowano i podzielono na 10 obszarów działania (np. rozwój zawodowy, odpowiedzialność, model uczenia się), a następnie wyobrębniono 60 podkategorii. Dane jakościowe na temat interwencji publicznych zbierano nad wyraz starannie. Należy docenić pracę wykonaną w trakcie wizyt studyjnych odbywanych we wszystkich analizowanych krajach, dotarcie do ciekawych rozmówców oraz przeprowadzenie ponad 200 wywiadów ze zróżnicowaną reprezentacją liderów kluczowych dla analizowanych reform. W efekcie tych działań studium ma charakter intensywny i zorientowany na przypadki, a nie na zmienne, co jest jego mocną stroną.

Mocną stroną raportu jest także teoria, a konkretnie podejście systemowe. Autorzy podjęli ambitną próbę spojrzenia na edukację holistycznie, starali się umieścić interwencje publiczne w kontekście konkretnego poziomu zaawansowania danego systemu. Holistyczne podejście nie jest tu jednak konsekwentne. W kilku miejscach zasygnalizowano, że inne, obok poziomu zaawansowania systemu, elementy kontekstu wprowadzonych reform są w badanych przypadkach drugorzędne lub trzeciorzędne[1]. Uzasadniając swoje wybory, autorzy piszą:

> [...] w przeszłości wiele zamieszania do debaty poświęconej doskonaleniu systemu edukacyjnego wprowadzało przekonanie o odmienności drogi do sukcesu każdego systemu: mają one różne punkty wyjścia, stają przed innymi oczekiwaniami i funkcjonują w innym kontekście społecznym i politycznym. (Mourshed, Chijioke i Barber, 2012, s. 24).

Zdaniem autorów raportu najważniejszym punktem odniesienia dla liderów edukacyjnych w trakcie projektowania reform powinny być inne systemy na tym samym poziomie rozwoju, w których udało się osiągnąć postęp. A zatem poziom zaawansowania systemu ma znaczenie pierwszorzędne dla porównań, na drugim miejscu znajduje się pakiet interwencji, a na trzecim dopiero „historia, kultura, polityka i struktura systemu szkolnego oraz narodu" (Mourshed, Chijioke i Barber, 2012, s. 24). W praktyce jednak autorzy w swojej analizie nie biorą nigdzie pod uwagę innych, poza poziomem rozwoju, czynników różnicujących systemy. Choć w tego rodzaju studiach uproszczenia są konieczne, ponieważ liczba charakterystyk systemów, które można byłoby uwzględnić, jest zbyt duża, taka argumentacja jest kontrowersyjna. Nasuwa się pytanie, czy dostatecznie dobrze uzasadniono małe zainteresowanie innymi częściami kontekstu i czy nie lepiej byłoby porównywać systemy, które cechuje podobny kontekst. Byłaby to alternatywna, bardziej przekonująca strategia poradzenia sobie z największym wyzwaniem międzykrajowych porównań – problemem zbyt dużej liczby zmiennych przy zbyt małej liczbie przypadków. Jednak musiałaby jej towarzyszyć zapewne niechętnie widziana przez autorów redukcja celów.

Na poziomie teoretycznym rzuca się też w oczy założenie o istnieniu pewnej uniwersalnej dla wszystkich systemów, linearnej logiki zmian, w ramach której poszczególne systemy przechodzą przez te same etapy w identycznej sekwencji i te same interwencje publiczne są u nich źródłem awansu na wyższy poziom. Autorzy twierdzą, że teza ta ma charakter empiryczny, wypowiadają ją jako

[1] Choć w kilku miejscach zwracają uwagę, że interwencja powinna być dopasowana do kontekstu: historii, kultury, polityki, struktury systemu szkolnego oraz narodu (Mourshed Chijioke i Barber, 2012, s. 24).

wniosek rzekomo udowodniony na gruncie prowadzonych przez nich analiz 20 systemów (Mourshed, Chijioke i Barber, 2012, s. 24). Jest to jednak teza niedowiedziona w omawianym raporcie. Autorzy nawet nie próbują jej udowadniać. De facto teza ta stanowi założenie raportu, jednak autorzy nie powołują się na żadną teorię, aby ją ugruntować. To kluczowa kwestia, bez przyjęcia powyższego założenia nie sposób bowiem ekstrapolować wyników generalizacji dokonanej na podstawie badania 20 systemów na wszystkie systemy edukacyjne, co czynią autorzy. Zakłada się w sposób dorozumiany, że badania nad interwencjami publicznymi w obszarze edukacji mają wysoką trafność zewnętrzną dla innych systemów, będących na tym samym etapie rozwoju, chociaż wyniki badań eksperymentalnych jak dotąd nie potwierdziły tej tezy (por. Sims, 2010). Podsumowując, na poziomie teoretycznym na uznanie zasługuje systemowe podejście, jednak poważną słabością jest nieuwzględnianie licznych elementów kontekstu oraz niewypowiedziane, nieugruntowane założenie o uniwersalności i linearności zmian w systemach edukacyjnych.

Politologia w perspektywie porównawczej dysponuje komparatystykami dużej liczby przypadków (powyżej 50), małej liczby przypadków (około 20), a także studiami pojedynczych przypadków. Autorzy analizowanego raportu wybrali metodę porównania małej liczby przypadków, co wydaje się optymalne z punktu widzenia celów raportu (Landmann, 2008). Metoda ta ma wiele zalet, ponieważ jest intensywna, operuje na bardziej konkretnym poziomie niż abstrakcyjne porównanie wielu państw, a przy starannym doborze przypadków pozwala wnioskować o większej populacji, czego nie umożliwiają z kolei pojedyncze studia przypadków. Odmiennie od ilościowego porównania w skali globalnej, ograniczonego zwykle do operacji statystycznych, pozwala na pogłębioną analizę procesów, łańcuchów zdarzeń, mechanizmów przyczynowych. Ten walor porównania małej liczby państw autorzy raportu w pełni wykorzystali. Dzięki solidnej analizie jakościowej udało im się wyjść poza powierzchowne rozumienie zmian, uchwycić charakterystyczne mechanizmy.

Porównywanie małej liczby przypadków cechują dwie główne strategie: można porównywać różne wyniki (*outcomes*) w grupie podobnych przypadków, kontrolowane są tu podobieństwa między przypadkami, a analizowane różnice, aby dostarczyć wyjaśnienia wariancji (tzw. *Most Similar Systems Design,* MSSD), lub można porównywać podobne wyniki w grupie różniących się do siebie przypadków (tzw. *Most Different Systems Design,* MDSD). W przypadku analizowanego raportu wybrano MDSD, porównywano najróżniejsze systemy edukacyjne, które osiągnęły podobny efekt – wyniki ich uczniów mierzone międzynarodowymi testami uległy poprawie.

Metodologiczną słabość MDSD w porównaniu do MSSD obrazuje Tabela 1, która pokazuje, jak mało przekonujące jest porównywanie całkowicie odmiennych od siebie systemów, w których zastosowano podobną interwencję publiczną, osiągając podobne rezultaty. Jeśli analizujemy odmienne systemy, które łączy tylko występowanie podobnej interwencji i podobnego efektu, to nie możemy powiedzieć, co spowodowało ten efekt. W szczególności nie możemy twierdzić, że to wybrana interwencja publiczna wywołała obserwowany efekt.

Autorzy omawianego raportu starali się przybliżyć do modelu MSSD, tworząc klastry systemów podobnych do siebie. Jednak klastry te zostały utworzone tylko na podstawie jednej zmiennej – wyników uczniów mierzonych wynikami badań międzynarodowych. Nie można zatem uznać, że

Tabela 1
Porównanie MSSD i MDSD

	MSSD			MDSD		
Systemy	**1**	**2**	**n**	**1**	**2**	**n**
Ich cechy	A	A	A	A	D	G
	B	B	B	B	E	H
	C	C	C	C	F	I
Zmienna objaśniająca (interwencje publiczne)	x	x	brak x	x	x	x
Zmienna objaśniana (poprawa wyników uczniów)	y	brak y		y	y	y

w ramach poszczególnych klastrów mamy rzeczywiście do czynienia z systemami podobnymi do siebie. Aby zbliżyć się do MSSD, systemy powinny zostać pogrupowane w klastry z punktu widzenia jak największej liczby podobieństw w ich charakterystykach, następnie powinno się zbadać w tych podobnych systemach, jakie zastosowano interwencje publiczne i jakie dały one skutki dla wyników uczniów. Byłaby to już jednak zupełnie inna analiza.

Wybór MDSD jest metodologicznie najsłabszym punktem tego opracowania. Nie pozwala bowiem wykazać, że w systemach podobnych, gdzie nie zaszła poprawa w wynikach uczniów, nie zastosowano omawianych interwencji publicznych. Autorzy w ogóle nie zajmują się systemami, które nie odniosły sukcesu, dowiadujemy się jedynie, że jest ich zdecydowana większość. Jednak jest możliwe, że w tych systemach zastosowano podobne interwencje, nie odnosząc pozytywnych rezultatów. Wówczas uogólnienie całego wnioskowania na większą liczbę przypadków niż omówione w raporcie staje się niemożliwe i nie należy na jego podstawie przygotowywać rekomendacji dla pozostałych państw. Wybór porównania małej liczby przypadków przy zastosowaniu MDSD naraża to studium na zarzut nieuprawnionego przenoszenia wyniku uzyskanego dla próby na całą populację (*overclaiming)*. Co więcej, nawet rekomendacje dla samych analizowanych państw mogą być wątpliwe, ponieważ nie zbadano przykładów nieudanych interwencji. Obserwowany sukces może być dziełem przypadku lub zmiennych pozostających poza obserwacją.

Przejdźmy do oceny wykonanych analiz. Zastanówmy się, na ile autorom udało się ominąć cztery metodologiczne pułapki występujące w zasadzie we wszystkich studiach porównawczych: zbyt dużą liczbę zmiennych przy zbyt małej liczbie przypadków, problem ekwiwalencji przypadków, błąd doboru próby, błąd pominiętej zmiennej (King, Keohane i Verba, 1994, s. 118–122).

Z pierwszą pułapką autorzy raportu radzą sobie, redukując złożoność badanej rzeczywistości. Stwierdzają, że najważniejszą charakterystyką systemów edukacyjnych jest poziom wyników uzyskiwanych przez uczniów. Pozostałe elementy kontekstu mają co prawda ich zdaniem pewne znaczenie, ale nie są decydujące, dlatego uprawniony jest uniwersalizm w myśleniu o reformach systemów edukacyjnych (Mourshed, Chijioke i Barber, 2012, s. 12). Kierując się tym niedowiedzionym założeniem, tworzą klastry systemów, by następnie traktować je jako słabe, przeciętne, dobre i bardzo dobre. W ramach klastrów autorzy szczycą się jednak uwzględnieniem bardzo odmiennych systemów, co ma ich zdaniem umożliwiać ekstrapolację wyników na całą populację. Uderzające jest zróżnicowanie porównywanych systemów: wybrano 12 systemów o zasięgu krajowym, 5 systemów funkcjonujących w ramach prowincji, 2 systemy obejmujące dystryk-

ty i 1 – sieć szkół. Wiąże się z tym rozmiar tych systemów: 3 obejmują poniżej 200 szkół, 3 między 200 a 1000 szkół, 7 między 1000 a 10 000 szkół, 3 od 10 000 do 20 000 szkół i 4 większe niż 20 000 szkół. Duże są różnice w zamożności ludności: 6 systemów z PKB per capita poniżej 10 000 USD, 4 między 10 000 i 20 000 USD, 6 między 20 000 i 45 000 USD i 4 z PKB per capita między 45 000 i 60 000 USD. Pominięcie tego zróżnicowania w fazie klastrowania budzi poważne zastrzeżenia. Warto zwrócić uwagę, że gdyby nie ten zabieg autorzy raportu nie poradziliby sobie z problemem zbyt dużej liczby zmiennych przy zbyt małej liczbie przypadków albo musieliby ograniczyć swój niezwykle ambitny cel analityczny – znalezienie skutecznych recept dla wszystkich systemów edukacyjnych na świecie.

Problem ekwiwalencji przypadków nurtujący badaczy porównawczych nie został wcale poruszony przez autorów raportu. W różnych warunkach czymś innym są partycypacja polityczna, władza wykonawcza, służba cywilna (Landmann, 2008). Nie da się wykluczyć, że także systemy edukacyjne przybierają różnorodne, z trudem porównywalne formy. Na obronę autorów raportu można powiedzieć, że systemy edukacyjne, jakkolwiek odmiennie urządzone, odpowiadają sobie funkcjonalnie, a zatem spełniają kryterium ekwiwalencji. Jednak poważny niepokój może budzić w tym kontekście włączenie do wspólnej kategorii systemów z poziomu ogólnopaństwowego, ze szczebla prowincji, dystryktów, a nawet sieci szkół. Powstaje wątpliwość, na ile spełnione jest kryterium ekwiwalencji dla systemu ogólnopaństwowego liczącego ponad 200 000 szkół i systemu – sieci liczącej poniżej 200 szkół.

Ryzyko błędu doboru próby jest poważnym wyzwaniem dla porównań małej liczby przypadków, zaniedbanie reguł w tym zakresie prowadzić może do niepewnych, ograniczonych wniosków, a w szczególności do nieprawidłowych konkluzji. Dzieje się tak dlatego, że rzadko dobór próby ma w takim przypadku charakter losowy, zazwyczaj odbywa się w sposób intencjonalny. W sposób celowy dobrano też próbę w analizowanym raporcie. Najbardziej problematyczny byłby dobór wyłącznie takich przypadków, które potwierdzają teorię. Autorzy raportu nie kierowali się tym kryterium, dokonali jednak drugiego w kolejności nadużycia – dobrali przypadki, kierując się wartością zmiennej zależnej (Geddes, 1990; King, Keohane i Verba, 1994). Jedną z najczęstszych pułapek, w którą wpadają komparatyści, jest wybór do analizy jedynie przypadków obejmujących czynnik, który chce się wyjaśnić. Prowadzić to może do przeszacowania wpływu wyodrębnionych zmiennych wyjaśniających lub pominięcia innych ważnych zmiennych niezależnych. Na potrzeby raportu wyselekcjonowano 20 systemów szkolnych z całego świata, które zanotowały znaczące, trwałe i powszechne osiągnięcia mierzone krajowymi lub międzynarodowymi testami począwszy od 1980 r. Odrzucono 43 systemy, które co prawda brały udział w pomiarze umiejętności uczniów umożliwiającym umieszczenie ich na uniwersalnej skali, jednak nie osiągnęły poprawy wyników. W drugim kroku podzielono wybrane systemy na dwie kategorie, aby reprezentowane były kraje rozwinięte i rozwijające się. Odrzucenie systemów, które nie osiągnęły poprawy wyników daje duże ryzyko popełnienia błędu pominiętej zmiennej.

I wreszcie ostatnia kwestia – obserwowalne implikacje studium. Autorzy uogólniają rezultaty swoich analiz i przekonują, że ich twierdzenia są słuszne dla wszystkich systemów, a na ich podstawie można przygotowywać rekomendacje dla interwencji publicznych.

> Świat musi wiedzieć, jaką naukę powinny odebrać systemy będące na różnych

> etapach rozwoju, jeśli chodzi o to, *co* i *jak* reformować. To nie jest zwykły raport. Zostały w nim uchwycone działania realizowane w czasie rzeczywistym. Dzięki swojej przejrzystości i istotnym spostrzeżeniom doprowadzi do błyskawicznego rozwoju całego obszaru związanego z reformami systemowymi (Mourshed, Chijioke i Barber, 2012, s. 11).

Jest to bardzo ambitny cel, jednak systematyczna analiza raportu pod kątem metodologicznym prowadzi do wniosku, że zastosowana metodologia nie pozwala z przekonaniem uogólnić wniosków z raportu na ogół systemów. A zatem deklarowane przez autorów cele nie zostały osiągnięte.

Podsumowanie

Autorzy międzynarodowych raportów porównawczych nie po raz pierwszy popełniają podobne błędy jak te zidentyfikowane w analizowanym opracowaniu. Znany jest przykład analizy gospodarek tzw. azjatyckich tygrysów (Landmann, 2008). W poszukiwaniu recept opartych na faktach zastanawiano się, dlaczego odniosły one sukces. Analogicznie do omawianego tu studium wzięto pod uwagę tylko gospodarki sukcesu, zamiast porównać próbkę możliwie podobnych państw z regionu, którym się powiodło, z tymi, które poniosły porażkę. Przyjęta metodologia prowadziła do wniosków na temat skutecznych strategii rozwojowych, które empirycznie okazały się błędne. Na podstawie tego rodzaju studiów m.in. Bank Światowy oraz Międzynarodowy Fundusz Walutowy prowadziły poważne i mające istotne konsekwencje polityki ekonomiczne, które później okazały się nieskuteczne (Landmann, 2008). Autorzy raportu piszą:

> niemal każdy kraj w ciągu ostatnich dwóch dekad podjął reformę systemu oświaty, jednak tylko nielicznym udało się poprawić system [...]. Przyglądamy się bliżej tym, którym udało się zanotować znaczącą, trwałą i powszechną poprawę wyników uczniów. Zastanawiamy się, co takiego zrobiły, że odniosły sukces, podczas gdy wielu się to nie udało [...]. W latach 1970–1994 większość państw OECD podwoiło, a nawet potroiło realne wydatki na edukację, jednak w większości państw wyniki uczniów pozostały na tym samym poziomie lub pogorszyły się (Mourshed, Chijioke i Barber, 2012, s. 20).

Interesujące jest, na co wydała pieniądze większość, która nie odniosła sukcesu lub której wysiłki zakończyły się porażką? Nie badając tej grupy, autorzy raportu nie mogą twierdzić, że zidentyfikowali źródła sukcesu. Jest możliwe, że wiele z państw, które nie mogą poszczycić się zadowalającymi rezultatami, zastosowało podobne interwencje publiczne, nie osiągając jednak analogicznych rezultatów. Teoretycznie możliwe jest nawet, że większość państw, które zastosowały rekomendowane przez autorów raportu zabiegi, nie odniosła sukcesu, a wybrane przypadki stanowią wyjątki od reguły (*outliers*). O tym jednak nie dowiemy się z lektury raportu McKinseya, bowiem głębszy wgląd w mechanizmy zmiany uniemożliwia już na wstępie metodologia MDSD i dobór próby według zmiennej zależnej.

Aby wiedzieć, jakie interwencje są skuteczne, trzeba dokonać porównania z systemami, którym nie udało się osiągnąć sukcesu. Czy i jakie przeprowadzono w nich reformy? Czy przeprowadzone reformy zostały skutecznie wdrożone? Jeśli zostały skutecznie wdrożone, dlaczego nie przyniosły poprawy wyników uczniów? Jeśli nie zostały skutecznie wdrożone, co stało na przeszkodzie interwencji publicznej? Z tą serią poważnych wątpliwości autorzy raportu rozprawiają się w jednym zdaniu, stwierdzając, że:

> systemy, którym nie udało się dokonać poprawy, mogły podejmować takie same działania naprawcze jak systemy, którym udało się udoskonalić, ale była między nimi jedna

> zasadnicza różnica – albo nie były wytrwałe w przeprowadzaniu tych interwencji, albo nie wprowadziły dostatecznego rygoru i dyscypliny przy ich realizacji (Mourshed, Chijioke i Barber, 2012, s. 26).

Nie wiadomo, na jakiej podstawie autorzy stawiają ten mocny wniosek, na którym opierają się wszystkie ambitne konkluzje raportu i płynące z nich rekomendacje. Nie podają żadnych wyników analiz na jego poparcie. Dopiero systematyczna analiza próbki państw, które odniosły porażkę, mogłaby uwiarygodnić wnioski w tym zakresie. Nieobecność lub nieskuteczne wdrożenie analogicznych reform w systemach na podobnym etapie rozwoju, które nie zrobiły postępów, pozwoliłyby z większym prawdopodobieństwem wnioskować, że wymienione interwencje są rzeczywiście uniwersalnie przydatne dla systemów na konkretnym poziomie rozwoju.

Jeśli natomiast okazałoby się, że pozostali zastosowali podobne interwencje, analiza państw sukcesu byłaby tym ciekawsza, pozwalałaby bowiem zrozumieć, jakie są warunki możliwości sukcesu w implementacji tych reform. Jeśli by natomiast wykazała, że w pozostałych systemach nie zastosowano analogicznych interwencji, zwiększałoby to prawdopodobieństwo, że to właśnie wymienione w raporcie *Jak najlepiej doskonalone systemy szkolne na świecie stają się jeszcze lepsze* reformy są czynnikami sprawczymi pozytywnych przemian.

Literatura

Barber, M. i Mourshed, M. (2007). *How the world's best performing schools come out on top.* McKinsey & Company.

Coffield, F. (2011). Why the McKinsey Reports will not improve school systems. Pobrano z: http://susanohanian.org/show_research.php?id=446

Davies, H. Nutley, S. i Smith, P. (2000). *What works? Evidence-based policy and practice in public services.* Bristol: Policy Press.

Geddes, B. (1990). How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis 2*(1), 131–150.

King, G., Keohane, R. i Verba, S. (1994). *Designing Social Inquiry. Scientific Inference in Qualitative Research.* Princeton: University Press.

Landmann, T. (2008). *Issues and methods in comparative politics.* New York: Routledge: 2008.

MacIntyre, A. (1971). *Against the self-images of the age: essays on ideology and philosophy*. London: Duckworth.

Mourshed, M., Chijioke, Ch. i Barber, M. (2012). *Jak najlepiej doskonalone systemy szkolne na świecie stają się jeszcze lepsze*. Warszawa: Centrum Edukacji Obywatelskiej.

Sims, Ch. A. (2010). But economics is not an experimental science. *Journal of Economic Perspectives. 24*(2), 59–68.