

dr Maciej Jakubowski

Uniwersytet Warszawski – Wydział Nauk Ekonomicznych

Zastosowanie modeli EWD do ewaluacji polityki edukacyjnej

Artykuł prezentuje modele edukacyjnej wartości dodanej, które mogą być stosowane do ewaluacji polityki edukacyjnej w oparciu o dane z całej populacji uczniów lub reprezentatywnych prób. Modele EWD są znacznie bardziej wiarygodne niż typowe analizy statystyczne nieuwzględniające „wyjściowego” poziomu umiejętności uczniów. Przy ewaluacji polityki edukacyjnej interesuje nas bowiem wpływ, jaki mają oceniane programy edukacyjne na *przyrost* wiedzy uczniów, a nie na jej *poziom*. Poziom wiedzy związany jest bowiem przede wszystkim z pochodzeniem społecznym ucznia, na które w krótkim okresie nie możemy wpływać. Przyrost wiedzy zależy już jednak w dużej mierze od tego, co działo się w trakcie nauki na określonym poziomie, w tym przypadku w gimnazjum¹.

W artykule pokazano zastosowanie kilku modeli EWD do oceny efektywności wydatków na gimnazja oraz wpływu liczby uczniów w klasie na przyrost ich umiejętności. Przedstawione modele są na tyle ogólne, że mogą być swobodnie stosowane do oceny innych rozwiązań w systemie edukacji. Zaletą ewaluacji metodami EWD, poza wzięciem pod uwagę wejściowego i wyjściowego poziomu umiejętności, jest także możliwość ilościowej analizy danych z całej populacji. W ten sposób rezultaty ewaluacji są bardzo precyzyjne, a wnioski można uogólniać na uczniów, nauczycieli i szkoły w całej Polsce (lub odpowiednio liczne podgrupy). W ten sposób metody EWD dają podstawę do oceny polityki edukacyjnej na każdym poziomie decyzyjnym: ogólnokrajowym, regionalnym, a nawet gminy lub szkoły.

W Polsce opracowano i udokumentowano metodę EWD dla gimnazjów (patrz: Dolata [red.], 2007; Jakubowski, 2008a; www.ewd.edu.pl). Warianty tej metody wykorzystano w niniejszym artykule. Dla przykładu poddano ocenie

¹ Według Autora uzasadnionym zarzutem wobec metody EWD jako oceny efektywności pracy gimnazjów jest niemożność uwzględnienia wszystkich czynników wpływających na rozwój ucznia, w tym niezależnych od pracy szkoły (np. dodatkowych zajęć). Jednak w przypadku ewaluacji polityki edukacyjnej metodą EWD, gdzie ocenie poddawane są grupy uczniów, szkół, czy gmin, trudno jest takie zarzuty uznać za zasadne. Stanowiłyby one problem jedynie w przypadku, gdy oceniana polityka byłaby silnie skorelowana z cechami nieuwzględnionymi w modelu. Co więcej, dla modelu z efektami stałymi jedynie różnice między kohortami uczniów tych samych szkół mogą wpływać na oceny EWD. Trudno sobie wyobrazić, by w całej populacji takie różnice dotyczące dodatkowych zasobów materialnych, społecznych czy kulturowych występowały na dużą skalę, a na dodatek były skorelowane z ocenianą polityką. W każdym przypadku badacz musi jednak te kwestie osobno rozważyć w zależności od tego, jakie działania poddawane są ewaluacji.

efektywność wydatków na gimnazja oraz wpływ wielkości klas w gimnazjum na przyrost wiedzy uczniów. Przykłady te zaczerpnięte są z innych prac empirycznych Autora (por. Jakubowski, 2007; Jakubowski, 2008b; Jakubowski, Sakowski, 2006). Zainteresowani tymi badaniami czytelnicy powinni sięgnąć do wskazanych artykułów, gdzie szczegółowo opisano kontekst, metodologię, zbiory danych oraz interpretację wyników². W niniejszej pracy analizy potraktowano jako przykłady ewaluacji metodą EWD, opisując najważniejsze aspekty metodologiczne. Warto nadmienić, że zarówno wydatki na gimnazja, jak i liczebność uczniów w klasach, dotyczą finansowych nakładów na edukację gimnazjalną. Ewaluacja metodą EWD może jednak dotyczyć dowolnych aspektów polityki edukacyjnej, o ile możliwe są one do skwantyfikowania. W polskim kontekście nasuwają się liczne pomysły na podobne badania dotyczące np. wpływu wykształcenia nauczycieli, segregacji uczniów między szkołami i klasami, wpływu różnic w realizowanych programach i metodach nauczania itp. Ważnych kwestii do badania jest sporo, warto więc spojrzeć, jak można wykorzystać modele EWD do tego celu.

Opis źródeł i charakterystyka danych

W pracy wykorzystano dane z dwóch źródeł. Wyniki egzaminów pochodzą z baz danych Centralnej i Okręgowych Komisji Egzaminacyjnych³. Drugim jest Bank Danych Regionalnych Głównego Urzędu Statystycznego⁴. Z pierwszego źródła uzyskano dane dotyczące wyników egzaminów wraz ze skromnym zestawem indywidualnych charakterystyk. Z drugiego źródła uzyskano dane dotyczące wydatków oświatowych gmin oraz opieki przedszkolnej z lat 1995 – 2006. Obydwa źródła zostały opisane w artykułach wymienionych powyżej. Warto jedynie podkreślić, że połączenie wyników sprawdzianu i egzaminu gimnazjalnego nie powiodło się dla ok. 7% uczniów, głównie ze względu na zmianę miejsca zamieszkania lub podchodzenie do egzaminu z inną kohortą uczniów. Dane GUS dla 2007 roku nie były natomiast dostępne w trakcie powstawania artykułu i zastąpiono je danymi z 2006 roku, zakładając ciągłość wydatków (co odpowiada praktyce wcześniejszych lat). Wydatki zostały zdeflowane w oparciu o indeks HICP Eurostatu oraz policzone dla lat odpowiadających nauce ucznia w gimnazjum, podobnie jak wskaźniki partycypacji w opiece przedszkolnej dla lat, w których uczeń potencjalnie mógł z niej korzystać (por. Jakubowski, 2007, 2008b). Ważnym założeniem jest, że dany uczeń całą swoją wcześniejszą edukację realizował w tym samym gimnazjum, co w polskim kontekście jest zapewne

² Prace te dostępne są na stronie www.wne.uw.edu.pl/mjakubowski. Mogą zostać także przesłane przez email, kontakt: mjakubowski@uw.edu.pl.

³ Zbiory danych wykorzystane w artykule powstały w trakcie prac grupy eksperckiej powołanej przez Centralną Komisję Egzaminacyjną i kierowanej przez dr. Romana Dolatę, której zadaniem jest opracowanie i testowanie modeli edukacyjnej wartości dodanej. Autor chciałby szczególnie podziękować Centralnej Komisji Egzaminacyjnej za udostępnienie wyników egzaminów oraz wsparcie prac badawczych.

⁴ Więcej informacji i dane BDR można znaleźć na: www.stat.gov.pl/bdr

prawdą dla większości uczniów ze względu na stosunkowo niewielką migrację między szkołami i miejscowościami. Trzeba też dodać, że dla kilkudziesięciu szkół nie udało się dołączyć danych GUS ze względu na niezgodny kod gminy w bazach OKE/CKE i BDR GUS. Ostateczną liczbę uczniów, gimnazjów i gmin wykorzystanych w analizie podano w tabelach podsumowujących wyniki regresji. Zarówno niewielka liczba, jak i losowy charakter braków danych gwarantują, że ich nieuwzględnienie nie miało wpływu na ostateczne wyniki.

Ewaluacja metodą EWD

Zaproponowana metoda badania opiera się na dwóch podejściach. Pierwsze z nich odpowiada dominującej na świecie metodzie EWD, jeśli chodzi o ocenę efektywności pracy szkół lub nauczycieli. Jest to hierarchiczny model liniowy (inne stosowane nazwy to: model mieszany, z efektami losowymi, wielopoziomowy, z losową stałą lub współczynnikiem), gdzie dane indywidualne uczniów (w tym wyniki sprawdzianu) stanowią poziom pierwszy analizy i są podstawą do oszacowania efektów losowych gimnazjów (poziom drugi) oraz gmin (poziom trzeci). Podejście to zapobiega błędom związanym z agregacją danych lub pominięciem powiązania obserwacji wewnątrz grup. Wykorzystanie na poziomie indywidualnym danych o wynikach egzaminu kończącego szkołę podstawową powoduje, że model należy interpretować jako wyjaśniający nie tyle *poziom* wiedzy, co jej *przyrost* w trakcie trzyletniej nauki uczniów w gimnazjum. Tak więc zaletą modeli EWD jest, przynajmniej częściowe, wyłączenie wpływu na poziom osiągnięć czynników niezależnych od szkoły, także niemierzalnych, takich jak wrodzone zdolności (por. Raudenbush, Bryk, 2002; Goldstein, 1999).

Postać ogólną trzypoziomowego modelu EWD można zapisać poniższym równaniem:

$$y_{isg} = \pi_0 + X_{isg} \beta + u_g + v_{sg} + \varepsilon_{isg}$$

gdzie y_{isg} to wynik *itego* ucznia w szkole s i gminie g , a wektor X_{isg} zawiera wszystkie interesujące nas charakterystyki ucznia, szkoły i gminy. W naszym przypadku y oznacza łączny wynik egzaminu gimnazjalnego, a wektor X zawiera wynik sprawdzianu danego ucznia (ew. także jego kwadrat). Inne uwzględnione w modelach wydatków na gimnazja zmienne można odczytać z tabeli zawierającej wyniki regresji (Tabela 1.). Część losowa równania zawiera reszty na poziomie ucznia ε_{isg} oraz efekt losowy gminy v_g i szkoły u_{sg} . Przyjmujemy typowe założenia o niezależności efektów losowych oraz o normalności i skończonej wariancji ich rozkładów:

$$\varepsilon_{isg} \sim N(0, \sigma^2_\varepsilon), u_{sg} \sim N(0, \sigma^2_u), v_g \sim N(0, \sigma^2_v)$$

Model opisany powyższym równaniem (1) nazywany jest modelem wielopoziomowym z losowymi stałymi, w tym przypadku na poziomie gminy oraz szkoły. Inną grupę stanowią modele EWD z losowym nachyleniem wyników egzaminu na niższym poziomie, w naszym przypadku z losowym nachyleniem współczynnika określającego relację między wynikami sprawdzianu i egzaminu gimnazjalnego. O ile szkoły różnią się pod względem tej relacji, to model z losowym nachyleniem pozwala określić odrębne efekty dla uczniów słabszych oraz zdolniejszych. Bardziej „płaskie” nachylenie sprawdzianu oznacza większą zdolność szkoły do wyrównywania szans edukacyjnych: szkoła ta radzi sobie lepiej z uczniami słabszymi niż podobne szkoły, a gorzej z uczniami lepszymi. Nie oznacza to jednak, że taka szkoła nie może być ogólnie szkołą bardziej (lub mniej) efektywną niż inne gimnazja, a jedynie to, że jej efektywność jest zróżnicowana i zależy od poziomu uczniów. Aby analizować tego typu zależności, zakłada się niezerową kowariancję między losową stałą oraz nachyleniem. W praktyce oznacza to, że dopuszczamy możliwość np. negatywnej korelacji, a więc tego, że szkoły z wyższymi wynikami uczniów cechuje także słabsza (bardziej płaska) relacja między wynikami na wejściu i wyjściu, a więc cechuje je też wyższa zdolność do wyrównywania wyników uczniów. Są to jednak tylko przykłady, które mogą zostać potwierdzone lub obalone w konfrontacji z rzeczywistymi danymi. W praktyce takie relacje mogą nie zachodzić lub też mogą być odwrotne (np. szkoły z wyższą przeciętną EWD są równocześnie szkołami zwiększającymi zróżnicowanie wyników uczniów, a więc sprzyjającymi uczniom zdolniejszym). Współczynnik nachylenia można próbować objaśniać innymi zmiennymi, testując hipotezy o ich „wyrównawczym” wpływie. W naszym przypadku będziemy badać efekt interakcji między wydatkami a nachyleniem wyników sprawdzianu, a więc testować hipotezę o wyrównującym wpływie wydatków na osiągnięcia uczniów.

Przy szacowaniu modeli z losowymi współczynnikami celowe jest „wycentrowanie” zmiennej objaśniającej tak, aby uzyskane oceny parametrów mogły być interpretowane w interesujący z praktycznego punktu widzenia sposób. W naszym przypadku oryginalna zmienna zawierająca wyniki sprawdzianu ma wartość minimalną 0, przez co oszacowane efekty nie mają bezpośredniej interpretacji. Stąd w modelu z losowym nachyleniem wyniki sprawdzianu zostały przekształcone tak, aby wartość 0 odpowiadała dolnemu decylovi (10 percentylowi) wyników sprawdzianu uczniów gimnazjów danej gminy (tzw. „group centering”). O ile wycentrowanie wokół średniej dla próby (tzw. „grand mean centering”) daje przy pewnych założeniach model ekwiwalentny modelowi bez przekształceń zmiennej objaśniającej, to już centrowanie wokół statystyki dla grupy zmienia interpretację modelu (de Leeuw, 2005). Model ten ma teraz ciekawą interpretację, mierząc wpływ interesujących nas czynników na sytuację uczniów o relatywnie niskim poziomie wiedzy na progu gimnazjum (w stosunku do poziomu wiedzy uczniów danej gminy).

Model z losowym nachyleniem zapiszemy zestawem równań, osobno definiując poziom ucznia oraz stałą i nachylenie dla gminy⁵:

$$\text{poziom indywidualny:} \quad y_{isg} = \pi_0 + D_{isg} \beta + \pi_1 spr_{isg}^* + \varepsilon_{isg}$$

$$\text{poziom gminy:} \quad \pi_0 = \gamma_{00} + G_0 \gamma_{01} + u_{0g}$$

$$\pi_1 = \gamma_{10} + G_1 \gamma_{11} + u_{1g}$$

gdzie na poziomie indywidualnym gdzie y_{isg} , oraz ε_{isg} zdefiniowano jak powyżej, D_{isg} to wektor zerojedynkowych charakterystyk ucznia (płeć, dysleksja, laureat), natomiast spr_{isg}^* to przekształcony w opisany powyżej sposób wynik sprawdzianu (wycentrowany wokół dolnego decyla). Równania na poziomie gminy objaśniają przeciętny wynik egzaminu (parametr π_0) oraz nachylenie indywidualnych wyników sprawdzianu (parametr π_1) za pomocą zestawu zmiennych na poziomie szkoły oraz gminy, w tym wydatków na ucznia podczas jego nauki w gimnazjum. Przez u_{0g} oraz u_{1g} oznaczono efekty losowe na poziomie gminy. Zakładamy, że $\varepsilon_{isg} \sim N(0, \sigma^2 \varepsilon)$, $u_{0g} \sim N(0, \sigma^2 u_0)$, $u_{1g} \sim N(0, \sigma^2 u_1)$, oraz $cov(u_{0g}, u_{1g}) \neq 0$.

Wszystkie powyższe modele oszacowano metodą największej wiarygodności za pomocą procedury `-xtmixed-` w programie Stata (por. RabeHesketh, Skrondal, 2005), łącznie dla trzech roczników uczniów. Przez to wariancja wykorzystana do oszacowania interesujących nas efektów dotyczy zarówno zróżnicowania wewnątrz gmin lub szkół na przestrzeni lat, jak i zróżnicowania między gminami i szkołami w każdym roku (tzw. „within-” oraz „between-variance”). Dla każdego roku dopuszczono zmiany w średnim nachyleniu wyników sprawdzianu oraz różnice w stałych (średnich wynikach egzaminów w całej populacji), co było niezbędne, biorąc pod uwagę brak porównywalności wyników egzaminów między latami i poziomami nauczania.

Nieco inną metodę zastosowano w badaniu wpływu wielkości klasy. W tym przypadku model dotyczy jedynie poziomu szkół i celowo nie wykorzystano w nim wariancji wyników między gimnazjami. Wcześniejsze badania (por. Jakubowski, Sakowski, 2006) dobitnie wykazały, że wielkość klasy jest zmienną silnie skorelowaną z wielkością szkoły i miejscowości, a przez to z takimi cechami jak wykształcenie rodziców (czy ogólnie poziom kapitału ludzkiego) czy zasoby kulturowe, społeczne i finansowe dostępne uczniom i szkołom. Przez to badania wykorzystujące zróżnicowanie liczby uczniów w klasach między szkołami lub gminami do oceny wpływu wielkości klasy wykazują pozorne zależności i są niewiarygodne. Jedynym sposobem określenia rzeczywistego wpływu wielkości klasy, zakładając brak możliwości przeprowadzenia eksperymentu terenowego, jest korzystanie z tzw. naturalnych eksperymentów. W niniejszym badaniu chcemy spojrzeć na to, jak w dużej mierze losowe zmiany populacji uczniów szkoły

⁵ Pominęto poziom szkoły, aby ułatwić trudne obliczenia numeryczne. Wprowadzenie efektów na poziomie szkół nie zmieniło jednak rezultatów.

między latami wpływają na osiągnięcia jej uczniów. Zakładając, że zmiany te wynikają głównie z różnej liczebności kohort uczniów związanej z liczbą urodzeń w danym roczniku w danej miejscowości lub rejonie, to mamy tu do czynienia z naturalnym eksperymentem. Oczywiście założenie to nie zawsze jest spełnione. Zmiany między latami mogą wynikać np. z rosnącej popularności danej placówki czy migracji między miejscowościami. Jednak dla większości szkół mają charakter egzogeniczny. Żeby zwiększyć pewność, że mamy do czynienia ze zmianami losowymi, badanie powtórzone zostanie na gimnazjach, gdzie zmiany liczby uczniów między latami są niewielkie, a klasy mają typowe rozmiary. Pominiemy też szkoły niepubliczne, które mają mniejsze klasy, ale i zupełnie odmienne zasady funkcjonowania i finansowania. Prezentowane tu podejście można uznać za dyskusyjne, ale chcemy je zaprezentować jako przykład wykorzystania naturalnego eksperymentu w metodzie EWD. W innych wariantach można skupić się jedynie na gminach z jednym gimnazjum, czy też użyć zmiennych instrumentalnych celem pominięcia nielosowej wariancji wielkości klas. Szukanie losowej wariancji w wielkości klas jest zadaniem wymagającym dużej kreatywności badacza, podobnie jak przy większości „naturalnych eksperymentów”. Stąd wyniki takich badań, o ile przekonujące, publikowane są w najlepszych pismach naukowych.

Zastosowana metoda EWD opiera się na modelu regresji z efektami stałymi, opisanym poniższym równaniem:

$$y_{is} = \beta_0 + X_{is}\beta + \eta k_s + u_s + \varepsilon_{is}$$

gdzie y_{is} to wynik egzaminu gimnazjalnego i -tego ucznia w szkole s , wektor X zawiera cechy ucznia, w tym dysleksję, płeć, status laureata oraz wynik sprawdzianu (wraz z kwadratem), a u_s to efekt stały na poziomie szkoły (co jest równoważne wprowadzeniu zmiennej zerowej oznaczającej każde gimnazjum). Kluczową zmienną jest tu k_s , czyli średnia wielkość klasy w szkole s . Pozytywna ocena parametru η oznaczać będzie, że większa liczba uczniów w klasie sprzyja przyrostowi wiedzy, a negatywna, że w liczniejszych klasach uczniowie przeciętnie osiągają mniej podczas nauki w gimnazjum.

Ten stosunkowo prosty model można oczywiście rozszerzyć o dodatkowe zmienne lub np. kontrolując selekcję między klasami⁶. Tutaj ma on jednak głównie cel dydaktyczny: pokazuje, jak oszacować metodą EWD czynniki wpływające na zmiany w przyrostach umiejętności w tych samych szkołach między latami. Model ten pomija różnice między szkołami, przez co ma inną interpretację niż modele EWD z efektami losowymi. Warto porównać uzyskane przez niego

⁶ Wykorzystanie średniej wielkości klasy w szkole, a nie faktycznej wielkości dla każdej klasy ogranicza wpływ na uzyskane oszacowania segregacji między klasami. Zastosowanie średniej wielkości klasy jako instrumentu dla faktycznej wielkości klasy byłoby jednak bardziej wskazane metodą (por. Jakubowski, Sakowski, 2006 jako przykład zastosowania zmiennych instrumentalnych w tym kontekście). Wyniki podejścia ze zmiennymi instrumentalnymi oraz innych specyfikacji i restrykcji nakładanych na próby dostępne są w: Jakubowski, 2008c.

wyniki także z oszacowaniami dla zwykłej regresji, bez efektów na poziomie szkół, która wykazuje pozorny związek wielkości klas z przyrostem wiedzy wynikający z powiązania tej zmiennej z innymi, nieuwzględnionymi w modelu cechami. W ten sposób widać, że umiejętne zastosowanie metod EWD może prowadzić do zupełnie różnych wniosków niż zastosowanie metod nieuwzględniających zarówno wiedzy uczniów na wejściu do gimnazjum, jak i ukrytych różnic między szkołami.

W modelu EWD z efektami stałymi niezbędne jest skorygowanie błędów standardowych o pogrupowanie uczniów w szkołach (taką korektę automatycznie daje model z efektami losowymi, ale nie ze stałymi). Wykorzystano tu opcję `-vce(cluster)-` w procedurze `-xtreg-` w programie Stata (por. Rogers, 1993 oraz dokumentację Stata dla procedur `-xt-`). Ponadto, podobnie jak poprzednio, dodano do modelu różnice w stałych oraz nachyleniu sprawdzianu między latami odzwierciedlające brak porównywalności wyników egzaminów zewnętrznych między rocznikami i poziomami nauczania.

Opis wyników: model EWD wydatków na gimnazja

Głównym celem artykułu jest zaprezentowanie możliwości ewaluacji metodą EWD, a nie dyskusja dotycząca polityki edukacyjnej, przez co w opisie wyników skupimy się na ich poprawnym odczytaniu, a nie na interpretacji. Zainteresowanych czytelników odsyłamy do oryginalnych artykułów, gdzie przeprowadzono szerszą dyskusję.

Tabela 1. poniżej przedstawia wyniki oszacowań dla modelu EWD z losowymi stałymi oceniającego efektywność wydatków na gimnazja. W tabeli pominięto oceny parametrów dla zmiennych zerowyjedynekowych, takich jak płeć, dysleksja i status laureata. Kolumna (1) pokazuje wyniki dla modelu jedynie z tymi zmiennymi. W dolnej części przedstawiono oszacowania odchylenia standardowego stałych na poziomie szkoły oraz gminy i reszt na poziomie ucznia. Oceny te są istotnie statystycznie różne od zera, wskazując znaczne, niewyjaśnione zmiennymi w modelu, zróżnicowanie wyników egzaminu gimnazjalnego między gminami i szkołami (nie jest to jeszcze model EWD, więc objaśnia surowe wyniki). Kolumna (2) zawiera już model EWD. Wyniki sprawdzianu są wysoko istotne, a ich wprowadzenie znacząco zmniejsza niewyjaśnioną wariancję na każdym poziomie analizy: ucznia, szkoły oraz gminy (patrz zmiany w ocenach odchylenia standardowego efektów losowych). Model (3) wprowadza wydatki na ucznia gimnazjum oraz oznaczenie szkół niepublicznych, a modele (4) i (5) dodatkowe zmienne na poziomie szkoły (logarytm naturalny liczby uczniów, rozstęp ćwiartkowy wyników sprawdzianu) oraz gminy (logarytm naturalny liczby uczniów, logarytm naturalny dochodów na mieszkańca, wydatki na ucznia w trakcie nauki w szkole podstawowej, procent dzieci 3 – 5 letnich, które korzystały z przedszkoli w kohorcie ucznia, oznaczenie typu miejscowości: wieś, miasto do 20 tys., miasto 20 – 100 tys., miasto pow. 100 tys. oraz oznaczenie OKE).

Większość wprowadzonych zmiennych ma jedynie zmniejszać niewyjaśnioną wariację i nie powinna być bezpośrednio interpretowana. Zmienne te stanowią tzw. „proxy” dla ukrytych cech szkół i gmin, a oceny parametrów nie odzwierciedlają tu ich rzeczywistego wpływu, a raczej pośredni efekt zmiennych ukrytych lub pominiętych (por. kwestia wielkości klasy analizowana poniżej).

Oceny parametrów dla kluczowej zmiennej określającej poziom wydatków w przeliczeniu na ucznia w trakcie całej jego nauki w gimnazjum są spójne i bardzo precyzyjne. Wyniki na pierwszy rzut oka sugerują, że większe wydatki wpływają negatywnie na przyrost wiedzy uczniów. Trzeba jednak podkreślić, że z praktycznego punktu widzenia nie są one odróżnialne od zera. Wydatki mierzone są w tysiącach złotych, a więc efekt ok. -0.07 punktów na tysiąc złotych należy uznać za pomijalny. Ogromna liczba obserwacji i precyzja modelu EWD powodują, że efekt ten jest istotny statystycznie, ale nie ma on znaczenia dla polityki edukacyjnej. Przy takich wielkościach oceny parametrów mogłyby być zarówno dodatnie, jak i ujemne, w każdym przypadku oznaczałoby to jednak, że wydatki nie wpływają na przyrost wiedzy w gimnazjum.

Kolejna tabela przedstawia rezultaty dla modelu EWD z losowym nachyleniem wyników sprawdzianu. Tym razem efektywność wydatków oceniana jest przez ich wpływ na przyrost wiedzy uczniów relatywnie słabych (z wynikiem ze sprawdzianu ok. 10 percentyla wyników w danej gminie). Chociaż negatywny efekt jest ponownie istotny statystycznie, to także tu powinien być interpretowany jako nierozróżnialny od zera z praktycznego punktu widzenia. Podobnie wpływ wydatków na nachylenie sprawdzianu, choć ujemny i istotny statystycznie, to jednak jest tak niewielki, że nie może mieć znaczenia dla polityki edukacyjnej. Warto zwrócić uwagę, że nachylenie wyników sprawdzianu jest tylko nieznacznie zróżnicowanie w polskich gimnazjach, choć istotnie różne od zera (patrz odchylenie standardowe efektów losowych). Korelacja między EWD a nachyleniem jest ujemna, co w tym przypadku oznacza, że szkoły podnoszące wyniki uczniom słabym cechuje przeciętnie zdolność do wyrównywania wyników uczniów, a przez to są one nieco mniej efektywne w odniesieniu do uczniów zdolniejszych. Nie oznacza to jednak, że nie są to szkoły z ogólnie wyższą EWD. Trzeba też wziąć pod uwagę, że korelacja ta jest bardzo słaba.

Tabela 1. Model edukacyjnej wartości dodanej z efektami losowymi

Zmienna zależna: łączny wynik egzaminu gimnazjalnego	(1)	(2)	(3)	(4)	(5)
Sprawdzian		1.254*** (0.007)	1.256*** (0.007)	1.252*** (0.007)	1.253*** (0.007)
Sprawdzian * rok = 2006		-0.012*** (0.003)	-0.012*** (0.003)	-0.012*** (0.003)	-0.013*** (0.003)
Sprawdzian * rok = 2007		0.042*** (0.003)	0.042*** (0.003)	0.042*** (0.003)	0.043*** (0.003)
Sprawdzian ^2		0.012*** (0.000)	0.012*** (0.000)	0.012*** (0.000)	0.012*** (0.000)
rednie wydatki na ucznia w trakcie nauki w gimnazjum			-0.067*** (0.020)	-0.076*** (0.020)	-0.066*** (0.020)
Szkoła niepubliczna			2.569*** (0.167)	3.179*** (0.178)	2.934*** (0.220)
Logarytm naturalny liczby uczniów w szkole				0.544*** (0.044)	-0.173** (0.058)
Rozstęp wiatkowy wyników sprawdzianu w gimnazjum				-0.035*** (0.007)	-0.045*** (0.007)
Logarytm naturalny liczby uczniów w gminie				-0.046 (0.047)	0.259*** (0.059)
Logarytm naturalny dochodów na mieszkańca gminy				1.422*** (0.209)	1.555*** (0.213)
Wydatki na ucznia w trakcie nauki w szkole podstawowej					-0.337*** (0.035)
Procent dzieci 3-5 letnich, które korzystały z przedszkola					0.008*** (0.002)
Oznaczenie typu miejscowości					Tak
Oznaczenie OKE			Tak	Tak	Tak
Stałe dla każdego roku	Nie	Tak	Tak	Tak	Tak
Stała	54.509*** (0.105)	15.354*** (0.107)	14.941*** (0.188)	2.719 (1.549)	4.268** (1.588)
Oszacowania efektów losowych					
Odch. std. efektów gmin	2.12	1.22	0.90	0.96	0.95
Odch. std. efektów szkół	6.53	3.03	2.90	2.85	2.84
Odch. std. reszt	16.86	10.34	10.34	10.34	10.34
Log restricted-likelihood	-5633162	-4983968	-4979920	-4979816	-4972839
Liczba uczniów	1325059	1325059	1324076	1324076	1322255
Liczba gimnazjów	6212	6212	6211	6211	6211
Liczba gmin	2467	2467	2466	2466	2466

1) W nawiasach podano błędy standardowe skorygowane o pogrupowanie uczniów na poziomie szkół i gmin.

2) ***, **, * oznaczają odpowiednio istotność na poziomie 1%, 5%, 10%.

3) W tabeli pominięto zmienne na poziomie ucznia: płeć, dysleksję, status laureata olimpiady, dla których współczynniki były identyczne z wcześniejszymi badaniami (por. Jakubowski, 2006, 2008a).

Tabela 2. Model edukacyjnej wartości dodanej z losowym nachyleniem wyników sprawdzianu

Zmienna zależna: średni wynik egzaminu gimnazjalnego		
Scentrowany wynik sprawdzianu (0 = 10百分yl oryginalnych wyników)	1.240*** (0.010)	1.213*** (0.010)
średnie wydatki na ucznia w trakcie nauki w gimnazjum	-0.055** (0.024)	-0.105*** (0.024)
średnie wydatki * sprawdzian	-0.001 (0.002)	-0.002* (0.002)
Logarytm naturalny liczby uczniów w szkole		0.279*** (0.022)
Logarytm naturalny liczby uczniów w gminie		-0.172*** (0.055)
Logarytm naturalny dochodów na mieszkańca gminy		1.648*** (0.221)
Różnica procentowa wyników sprawdzianu wśród uczniów gimnazjum		-0.128*** (0.006)
średni wynik sprawdzianu wśród uczniów gimnazjum		0.180*** (0.006)
Stała	4.372** (1.654)	-0.215 (1.67)
Oszacowania efektów losowych		
Odchylenie standardowe nachylenia wyników sprawdzianu	0.10	0.11
Odchylenie standardowe efektów gmin	2.32	2.36
Korelacja nachylenia sprawdzianu i efektów gmin	-0.17	-0.18
Odchylenie standardowe reszt na poziomie ucznia	10.47	10.46
Log restricted-likelihood	-4993013	-4991958
Liczba uczniów	1324076	1324076

1) W nawiasach podano błędy standardowe skorygowane o pogrupowanie uczniów na poziomie szkół i gmin.

2) ***, **, * oznaczają odpowiednio istotność na poziomie 1%, 5%, 10%.

3) W tabeli pominięto niektóre zmienne (m.in. płeć, dysleksja, status laureata, kwadrat sprawdzianu, efekty dla lat oraz OKE).

Opis wyników: model EWD wpływu wielkości klasy

Przejdźmy teraz do rezultatów modelu EWD z efektami stałymi oceniającego wpływ wielkości klasy na osiągnięcia uczniów. Tabela 3. przedstawia porównanie między wynikami zwykłej regresji (kolumny 1 – 3) oraz regresji z efektami stałymi (kolumny 4 – 6). Kolumny 1. oraz 4. zawierają model nie uwzględniający wyników sprawdzianu, a więc nie są to modele EWD. Kolumny 2 – 3 oraz 5 – 6 to modele EWD uwzględniające różny zestaw zmiennych kontrolnych. Dla przejrzystości

w tabelach pominięto współczynniki dla płci, dysleksji oraz laureatów olimpiad. We wszystkich regresjach błędy standardowe skorygowano ze względu na pogrupowanie uczniów w szkołach. Bez tej korekty byłyby one znacznie niższe, błędnie sugerując większą istotność statystyczną uzyskanych wyników⁷.

Oszacowania w Tabeli 3. zostały oparte na niemal całej próbie, jedynie po wykluczeniu szkół niepublicznych (1,7% uczniów), szkół niewielkich (mniej niż 15 uczniów zdających egzamin – 0,5% uczniów) oraz szkół z klasami skrajnie małymi (poniżej 8 uczniów – 0,1% uczniów) i skrajnie dużymi (powyżej 32 uczniów – 1,2% uczniów). W ten sposób wykluczono 3,6% uczniów oraz 14% gimnazjów z dostępnej do badania próby⁸. Powstaje jednak wątpliwość, czy przy rozpatrywaniu kwestii wpływu wielkości klasy należy brać pod uwagę większość szkół, w tym szkoły z najmniejszymi i największymi klasami. Ponadto, czy szkoły, dla których zmiany wielkości klas między latami są wyjątkowo duże, mogą stanowić podstawę do quaside-ksperymentalnego wnioskowania o wpływie wielkości klas? Można przypuszczać, że takie szkoły są dość wyjątkowe, zachodziły w nich procesy decydujące o nielosowych zmianach wielkości klas (np. migracja mieszkańców do nowych miejsc pracy, przenoszenie uczniów z likwidowanych szkół, łączenie szkół itp.). Stąd w Tabeli 4. podano dodatkowe oszacowania dla prób ograniczonych do:

- a) szkół, dla których zmiany średniej wielkości klas między latami były mniejsze niż 25%;
- b) szkół spełniających warunek powyżej i z przeciętną wielkością klas poniżej dolnego i górnego percentyla w populacji szkół (w tym przypadku ze średnimi klasami między 15 a 30 uczniami).

Stosowanie takich ograniczeń ma na celu sprawdzenie, czy nasze założenia są uzasadnione. W tym przypadku chcemy się upewnić, czy rezultaty są podobne dla bardziej restrykcyjnie ograniczonej próby, gdzie prawdopodobieństwo nielosowych zmian jest mniejsze. W przypadku metod odwołujących się do naturalnych eksperymentów takie testy są niezbędne, ponieważ nie możemy bezpośrednio sprawdzić losowości zmian w wielkościach klas między latami⁹.

⁷ Intuicyjnie brak korekty ze względu na pogrupowanie uczniów w szkołach zawyża liczbę dostępnych obserwacji. W skrajnym przypadku, o ile uczniowie każdej szkoły byłiby identyczni, mielibyśmy do dyspozycji liczbę obserwacji równą liczbie szkół a nie uczniów. Oczywiście, w praktyce mamy do czynienia ze zróżnicowaniem osiągnięć wewnątrz szkół, lecz już z tabel przedstawionych powyżej widać, że spora część ogólnej wariancji wyników uczniów wyjaśniana jest przynależnością do gimnazjum i korekta błędów standardowych jest niezbędna. Jeszcze raz należy podkreślić, że niestosowanie takiej korekty w badaniach edukacyjnych (niestety, wciąż częste) prowadzi do wysuwania nieuzasadnionych wniosków.

⁸ Należy pamiętać, że łączenie danych nie powiodło się dla ok. 7% uczniów, a z drugiej strony wiele gimnazjów to szkoły nietypowe, które nie powinny być brane pod uwagę w tym badaniu (np. w bazie jest 214 gimnazjów jednoosobowych). Niestety nie posiadamy nazw szkół a tym bardziej zmiennych identyfikujących ich rodzaje, stąd wykluczanie po wielkości klas i szkół wydaje się najlepszym sposobem pozostawienia jedynie szkół "typowych".

⁹ Gdybyśmy byli w stanie określić, jaka część wariancji jest czysto losowa, to zmienną definiującą tę część powinniśmy wykorzystać jako instrument dla faktycznej wielkości klas, pomijając całą resztę wariancji. Znalezienie odpowiedniego instrumentu i sprawdzenie jego zasadności jest jednak trudne a często niemożliwe (por. Jakubowski, Sakowski, 2006).

Tabela 3. Modele do oceny wpływu wielkości klasy. Pełna próba.

	Regresja liniowa			Regresja z efektami stałymi		
Przeciętna liczba uczniów w klasie	0.476*** (0.025)	0.106*** (0.011)	0.051*** (0.011)	-0.001 (0.012)	-0.006 (0.011)	-0.006 (0.011)
Wynik sprawdzianu		1.221*** (0.010)	1.198*** (0.009)		1.261*** (0.009)	1.240*** (0.009)
Mediana wyników sprawdzianu			0.322*** (0.010)			0.291*** (0.009)
Różnica w punktach wyników sprawdzianu			-0.061*** (0.009)			-0.053*** (0.007)
R ²	0.02	0.64	0.64	0.01	0.63	0.63
R ² „within”				0.01	0.63	0.63
R ² „between”				0.13	0.75	0.75
Liczba gimnazjów				5280	5280	5280
Liczba uczniów	1339022	1339022	1339022	1339022	1339022	1339022

1) W nawiasach podano błędy standardowe skorygowane o pogrupowanie uczniów na poziomie szkół.

2) ***, **, * oznaczają odpowiednio istotność na poziomie 1%, 5%, 10%.

3) W tabeli pominięto płeć, dysleksję, status laureata, kwadrat sprawdzianu i efekty dla lat.

Tabela 4. Modele do oceny wpływu wielkości klasy. Próba z ograniczeniami.

Regresja z efektami stałymi	Zmiana mniejsza niż 25%			Zmiana mniejsza niż 25% i rednie klasy od 15 do 30 uczniów		
Przeciętna liczba uczniów w klasie	0.016 (0.018)	-0.013 (0.016)	-0.016 (0.016)	0.017 (0.018)	-0.015 (0.016)	-0.019 (0.017)
Sprawdzian		1.268*** (0.010)	1.244*** (0.010)		1.266*** (0.010)	1.242*** (0.010)
Mediana wyników sprawdzianu			0.324*** (0.010)			0.325*** (0.010)
Różnica w punktach wyników sprawdzianu			-0.059*** (0.007)			-0.060*** (0.007)
R ²	0.01	0.63	0.63	0.01	0.63	0.63
R ² „within”	0.01	0.63	0.63	0.01	0.63	0.63
R ² „between”	0.18	0.77	0.77	0.17	0.77	0.77
Liczba gimnazjów	3894	3894	3894	3882	3882	3882
Liczba uczniów	1140163	1140163	1140163	1128522	1128522	1128522

1) W nawiasach podano błędy standardowe skorygowane o pogrupowanie uczniów na poziomie szkół.

2) ***, **, * oznaczają odpowiednio istotność na poziomie 1%, 5%, 10%.

3) W tabeli pominięto płeć, dysleksję, status laureata, kwadrat sprawdzianu i efekty dla lat.

Powyższe rezultaty pokazują dwie rzeczy. Po pierwsze, modele bez wyników sprawdzianu mają bardzo niską moc wyjaśniającą (ok. 1 – 2% wariancji wyników egzaminu gimnazjalnego). Wprowadzenie wyników sprawdzianu, a więc zastosowanie modelu EWD, zwiększa procent wyjaśnionej wariancji do ponad 60%. Wysoka moc objaśniająca widoczna jest zarówno wewnątrz szkół (R^2 „within”), jak i między szkołami (R^2 „between”)

Co do interesującego nas wpływu wielkości klasy, to zgodnie z oczekiwaniami zwykła regresja wykazuje pozorną, pozytywną (i istotną statystycznie) zależność. Wynikałoby z tego, że wbrew intuicji większe klasy sprzyjają podnoszeniu osiągnięć uczniów¹⁰. Co więcej, oceny parametrów są relatywnie duże, sugerując, że zwiększanie klas może mieć istotny z praktycznego punktu widzenia efekt (zwiększenie klasy o 10 uczniów podnosi EWD szkoły o 0,5 – 1 punkt egzaminacyjny, co należy uznać za dużo). Jednak oszacowania regresji z efektami stałymi pokazują, że jest to korelacja pozorna. W istocie, o ile bierzemy jedynie pod uwagę zmiany w wielkości klas między latami w tych samych szkołach, to efekt ten jest nieistotny statystycznie i staje się negatywny po wzięciu pod uwagę wyników sprawdzianu. Ograniczenie próby powoduje, że wpływ wielkości klasy jest bardziej negatywny, choć nadal nieistotny. Duże błędy standardowe są nie tylko efektem korekty ze względu na pogrupowanie (i podobieństwo) uczniów w szkołach, ale i zapewne pokazują, że wykorzystanie wielkości klasy zależy od kontekstu edukacyjnego. Nie każdy nauczyciel umie pracować z małą klasą w sposób inny niż z klasą dużą. Programy nauczania powinny być dostosowane do wielkości klas, nie mówiąc o metodach pracy. Tak więc można te wyniki interpretować jako sygnał, że kosztowne zmniejszanie klas powinno być związane ze zmianami w sposobie nauczania, tak aby wzrost wydatków został odpowiednio wykorzystany. Bez tego rodzaju zabiegów utrzymywanie mniejszych klas jedynie podnosi koszty oświaty bez widocznych efektów w jakości nauczania. Oczywiście, wskazane przyczyny braku wpływu wielkości klasy w Polsce to bardziej hipotezy niż sprawdzone rezultaty i kwestia ta wymaga bardziej pogłębionych badań.

Ciekawych oszacowań dostarczają też regresje uwzględniające medianę oraz rozstęp ćwiartkowy wyników sprawdzianu uczniów w klasie gimnazjalnej. Obie miary są odporne na obserwacje odstające, a także na zmianę skali pomiaru między latami, co w przypadku polskich egzaminów jest ważne (ze względu na brak zrównywania a przez to porównywalności). Oceny parametrów dla obydwu zmiennych są zgodne z intuicją, intryguje przy tym ich stabilność i wysoka istotność statystyczna. Sugerują one, że ogólny poziom klasy ma wysoce pozytywny wpływ na osiągnięcia ucznia, a większe zróżnicowanie wyników w klasie ma wpływ negatywny. Trzeba jednak podkreślić, że są to zmienne

¹⁰ Trzeba jednak zauważyć, że istnieją oparte w materiale empirycznym teorie uzasadniające pozytywny wpływ większej liczby uczniów w klasie. Znaczenie ma tu np. możliwość znalezienia innych uczniów na podobnym poziomie umiejętności, co sprzyja nauce, a jest bardziej prawdopodobne w większych klasach (por. Dobbela i in., 2002).

zdefiniowane na poziomie klas i służą tu raczej jako dodatkowe cechy kontrolne. Sam model został opracowany celem określenia wpływu wielkości klasy przez średnią liczebność klas na poziomie szkoły. Obserwacje coroczne dla klas nie są dostępne, przez co badanie quasisperymentalne nie jest tu możliwe. Dlatego też należy te wyniki traktować ostrożnie, raczej jako zależność empiryczną, a nie przyczynowo-skutkową, lub jako źródło hipotez do kolejnych badań.

Podsumowanie

W artykule przedstawiono trzy modele edukacyjnej wartości dodanej, które mogą być wykorzystane do ewaluacji polityki edukacyjnej w oparciu o wyniki egzaminów zewnętrznych dostępne dla niemal pełnej populacji uczniów. Pierwszy z nich to model z efektami losowymi na poziomie gmin oraz szkół, który posłużył do oceny efektywności wydatków na gimnazja. Wyniki sugerują, że wydatki nie wpływają na przyrost wiedzy uczniów. Drugi model zastosowano do oceny tej samej kwestii, jednak w stosunku do uczniów słabszych, o niskich wynikach sprawdzianu. Jest to model z losowym nachyleniem wyników sprawdzianu, który można wykorzystać do ewaluacji dowolnych rozwiązań względem wpływu na grupy uczniów definiowane przez ich poziom umiejętności. Rezultaty sugerują, że wydatki nie podnoszą także efektywności pracy z uczniami słabszymi i nie mają wpływu na wyrównywanie poziomu umiejętności w gimnazjum.

Kolejny model EWD opiera się na regresji z efektami stałymi. Poddano w nim analizie zmiany wielkości klas w tych samych szkołach na przestrzeni lat. Metoda ta odwołuje się do tzw. naturalnego eksperymentu, gdzie brana pod uwagę jest jedynie losowa wariancja w interesujących nas czynnikach. Rezultaty sugerują, że zmiany liczby uczniów w klasach nie mają wpływu na zmiany w przyrostach wiedzy uczniów. Porównanie ze zwykłą regresją pokazuje, że pozytywna zależność między wielkością klas a wynikami uczniów ma charakter pozorny. Zastosowanie podstawowego modelu EWD nie daje tu wiarygodnych rezultatów, potrzebne jest zastosowanie bardziej przemyślanych metod, takich jak model EWD z efektami stałymi szkół, które zmniejszają ryzyko błędnego wnioskowania wynikającego z nielosowego przypisania uczniów do instrumentów polityki oświatowej.

Podsumowując, należy stwierdzić, że metody EWD to potężne narzędzie ewaluacji polityki edukacyjnej państwa. W oparciu o nie może być prowadzona bieżąca ocena działań oświatowych państwa, szkół i samorządów, ugruntowana w dużych zbiorach informacji na temat osiągnięć uczniów, a przez to precyzyjna i możliwa do uogólnienia na całą populację i podgrupy. Zastosowanie metod EWD wymaga jednak świadomości ich ograniczeń. Nie każdy istotny statystycznie wynik w regresji metodą EWD należy interpretować jako zależność przyczynowo-skutkową. Określenie takich zależności wymaga przemyślanej strategii badawczej, przetestowania leżących u jej podstaw założeń, a także sprawdzenie odporności wyników na zmiany w specyfikacji modelu i próbie.

Jakość ewaluacji metodami EWD zależy też od jakości informacji dotyczące umiejętności uczniów. Tutaj jednak dyskusja dotyczy nie tyle metody edukacyjnej wartości dodanej, co rzetelności egzaminów zewnętrznych. Im wyższa ich wiarygodność, tym bardziej obiektywne będą oceny uzyskane metodami EWD. Należy liczyć, że zarówno jakość pomiaru dydaktycznego, jak i umiejętne wykorzystanie metod EWD do ewaluacji polityki edukacyjnej, będą w najbliższych latach rosły. Tylko w ten sposób może zwiększyć się nasza wiedza o procesach zachodzących w systemie oświaty, a przez to i racjonalność polityki oświatowej w Polsce.

Bibliografia:

1. Dobbela S., Levin, J., & Oosterbeek, H., *The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition*, Oxford Bulletin of Economics and Statistics, 2002.
2. Dolata R., *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania na podstawie egzaminów zewnętrznych*, Centralna Komisja Egzaminacyjna 2007.
3. Goldstein H., *Multilevel Statistical Models* (poprawione wydanie II dostępne na: www.ats.ucla.edu/stat/examples/msm_goldstein/goldstein.pdf), 1999.
4. Jakubowski M., *Wpływ czynników ekonomicznych na wyniki egzaminów zewnętrznych*, Biuletyn Badawczy "Egzamin" 11/2007, Centralna Komisja Egzaminacyjna.
5. Jakubowski M., *Czy wydatki na gimnazja są efektywne?*, Gospodarka Narodowa 11/12/2007.
6. Jakubowski M., *Implementing valueadded models of school assessment*, European University Institute RSCAS 2008/06 working paper, 2008a.
7. Jakubowski M., *Decentralization and teaching quality*, wersja robocza dostępna na: www.wne.uw.edu.pl/mjakubowski, 2008b.
8. Jakubowski M., *Value added quasi experimental model of class size effects*, wersja robocza dostępna na: www.wne.uw.edu.pl/mjakubowski, 2008c.
9. Jakubowski M., Sakowski P., *Quasiexperimental estimates of class size effect in primary schools in Poland*, International Journal of Educational Research, 2006, Vol. 45, Issue 3.
10. de Leeuw J., *Centering in Multilevel Models*, [w:] Brian Everitt, David Howell (red.), "Encyclopedia of Statistics in Behavioral Science", Wiley & Sons, 2005.
11. Rabe-Hesketh S., Skrondal A., *Multilevel and Longitudinal Modeling using Stata*. College Station, TX: Stata Press, 2005.
12. Raudenbush S. W., Bryk A., *Hierarchical Linear Models*, wydanie II, Sage Publications 2002.
13. Rogers, W., *Regression standard errors in clustered samples*. Stata Technical Bulletin 13, 1993, (www.stata.com/support/faqs/stat/stb13_rogers.pdf).