

Aleksandra Jasińska

Michał Modzelewski

Instytut Badań Edukacyjnych

Międzyszkolne zróżnicowanie wyników nauczania po pierwszym etapie kształcenia

Wprowadzenie

W Polsce od 1999 r. obowiązuje dziewięcioletni jednolity i powszechny system szkolnictwa, na który składa się sześcioletnia szkoła podstawowa i trzyletnie gimnazjum. Pierwszym oficjalnym progiem selekcyjnym jest koniec gimnazjum, po którym uczniowie mogą wybrać dalszy kierunek swojej drogi edukacyjnej: ścieżkę zawodową – technikum lub zasadniczą szkołę zawodową – albo ścieżkę akademicką, realizowaną w liceach. Wybór ten dokonuje się w dużej części w oparciu o wyniki na egzaminie gimnazjalnym. Zgodnie z założeniami jednolitego systemu szkolnictwa to, jakie wyniki uczniowie osiągają na tym egzaminie, lub szerzej – szanse na sukces edukacyjny uczniów nie powinny zależeć od tego, do której szkoły uczęszczają.

Jednym ze sposobów na liczbowe ujęcie problemu jednolitości kształcenia jest wskaźnik zróżnicowania międzyszkolnego wyników kształcenia (Dolata, 2011). Wskaźnik ten pozwala na określenie, na ile szkoły różnią się między sobą osiągnięciami szkolnymi uczniów. Pierwszym krokiem do jego wyliczenia jest dokonanie dekompozycji wariancji, czyli określenia, jaka część zróżnicowania indywidualnych wyników wszystkich uczniów może zostać przypisana zróżnicowaniu średnich grupowych (tj. średnich wyników dla szkół), a jaka zróżnicowaniu wyników poszczególnych uczniów w szkołach. W drugim kroku, komponent międzyszkolny zróżnicowania odniesiony jest do całkowitej wariancji wyników uczniów, a uzyskany wynik informuje nas, jaki odsetek całkowitego zróżnicowania osiągnięć szkolnych uczniów możemy przypisać podziałowi uczniów na szkoły. W przypadku idealnie jednolitego systemu szkolnego szkoły osiągałyby te same średnie wyniki w testach, a wskaźnik zróżnicowania osiągałby wartość 0% (czyli, inaczej mówiąc, uczniowie każdej ze szkół osiągaliby średnio takie same wyniki). W przypadku odwrotnym, całe zróżnicowanie wyników indywidualnych uczniów sprowadzałoby się do różnic między szkołami – wszyscy uczniowie w danej szkole osiągaliby te same wyniki, a wskaźnik przyjąłby wartość 100% (to, do jakiej szkoły uczęszczałby uczeń, całkowicie determinowałoby jego wynik).

Zróżnicowanie międzyszkolne na danym poziomie kształcenia zdefiniowane tak jak powyżej nie pozwala rozróżnić dwóch niezależnych zjawisk, które mogą mieć wpływ na jego wielkość. Pierwszym jest zróżnicowanie uczniów na wejściu ze względu na uprzednie osiągnięcia, ich cechy indywidualne i charakterystyki ich rodzin, drugim zaś różny poziom jakości kształcenia szkół. Pierwszy przypadek oznacza, że różne szkoły pracują z uczniami o różnym

poziomie uprzednich umiejętności lub różnych konfiguracjach cech indywidualnych i społecznych, mających wpływ na osiągnięcia, tzn., że w systemie istnieją szkoły, które przyjmują głównie lepszych uczniów, i takie, do których uczęszczają uczniowie słabsi. W drugiej sytuacji mamy do czynienia ze zjawiskiem, w którym szkoły ucząją z różną efektywnością, co przekłada się na zróżnicowanie ich średnich wyników. Najbardziej prawdopodobne jest jednak to, że zachodzą oba te zjawiska, przekładając się na obserwowane na danym etapie kształcenia zróżnicowanie szkół ze względu na wyniki nauczania.

By dobrze zrozumieć tę swoistą dialektykę zróżnicowania, warto przedstawić dwa skrajne hipotetyczne scenariusze działania tych procesów. W pierwszym, maksymalizującym zróżnicowanie, oba wspomniane procesy działają w jednym kierunku. W systemie szkolnym mielibyśmy wtedy do czynienia z sytuacją, w której nie dość, że uczniowie byłiby podzieleni między szkoły w zależności od swoich uprzednich osiągnięć lub cech indywidualnych i społecznych związanych z osiągnięciami szkolnymi, to, co więcej, szkoły skupiające uczniów lepszych uczyłyby zdecydowanie lepiej niż szkoły skupiające uczniów słabszych. Proces ten przyczyniłby się do zwiększania wyjściowego zróżnicowania wyników – lepsi stawaliby się jeszcze lepsi, a słabsi jeszcze słabsi.

Drugim skrajnym scenariuszem jest ten, w którym procesy te wzajemnie się znoszą. W tym przypadku mamy do czynienia z wysokim poziomem segregacji uczniów „na wejściu”, ale system szkolny pełni funkcję kompensacyjną dla uczniów słabszych pod względem uprzednich osiągnięć. W tym scenariuszu szkoły skupiające słabszych uczniów osiągają średnio wyniki nie gorsze od szkół, do których uczęszczają uczniowie lepsi co do uprzednich osiągnięć lub o korzystniejszej konfiguracji cech indywidualnych i społecznych mających wpływ na osiągnięcia szkolne. W taki sposób można próbować interpretować przyczyny niskiego zróżnicowania międzyszkolnego, w sytuacji gdy pomiar osiągnięć podsumowuje pewien okres kształcenia.

W przypadku gdy dysponujemy tylko wynikami osiągnięć szkolnych uczniów, rozplątanie wpływu tych dwóch procesów nie jest do końca możliwe. Sytuacja zmienia się jednak, gdy mamy dostęp do informacji o uprzednich osiągnięciach uczniów lub tych indywidualnych cech ucznia, które mogą mieć wpływ na jego osiągnięcia, ale nie są zależne od szkoły. Kontrolując je, jesteśmy w stanie „wyrzącić” efekt związany z nielosowym przypisaniem uczniów o różnych charakterystykach do klas i uzyskać informację o zróżnicowaniu jakości pracy szkoły. Wskaźnik zróżnicowania międzyszkolnego w tej sytuacji pozwala określić, na ile szkoły różnią się co do średniej efektywności nauczania. Tak zdefiniowana miara jest punktem wyjścia dla badań szkolnych uwarunkowań efektywności kształcenia. Pozwala bowiem skoncentrować się na tej części zróżnicowania wyników nauczania, która związana jest z procesami dziejącymi się w szkole, wyłączając znaczenie tych czynników, na które szkoła nie ma wpływu.

Jak widać, miara zróżnicowania międzyszkolnego kryje w sobie bardzo dużo informacji o tym, co dzieje się w danym systemie kształcenia. W związku z tym jest dobrym wskaźnikiem charakteryzującym system oświaty na poziomie ogólnym. W polskim systemie szkolnym regularną analizę i monitoring zróżnicowania międzyszkolnego dla okresu jednolitego kształcenia umożliwia

wprowadzony w 2002 roku system egzaminów zewnętrznych. Ogranicza się on jednak do dwóch pomiarów: pod koniec szkoły podstawowej (sprawdzian) oraz na zakończenie gimnazjum. Analiza tych danych pozwoliła na zaobserwowanie niepokojących tendencji wzrostowych zróżnicowania międzyszkolnego ze względu na wyniki egzaminów gimnazjalnych w dużych miastach (Dolata, Jasińska i Modzelewski, 2012).

W niniejszym tekście zaprezentowane zostały wyniki analizy zróżnicowania międzyszkolnego po pierwszym etapie kształcenia (tj. po trzeciej klasie szkoły podstawowej) przeprowadzonej z wykorzystaniem danych z badania *Szkolnych uwarunkowań efektywności kształcenia* (SUEK) realizowanego przez Instytut Badań Edukacyjnych. Obraz ten skonfrontowano z wynikami z międzynarodowych badań kompetencji PIRLS i TIMSS. Dodatkowo, wykorzystując informacje o cechach uczniów i ich rodzin zebranych w badaniu SUEK, podjęto próbę oszacowania zróżnicowania międzyszkolnego efektywności nauczania, przybliżając znaczenie obu wspomnianych procesów dla wariacji międzyszkolnej wyników nauczania.

Metoda

Pytania badawcze

W ramach analizy problemu międzyszkolnego zróżnicowania wyników nauczania postawiono trzy kluczowe pytania badawcze. Po pierwsze zapytano o to, jakie jest zróżnicowanie międzyszkolne wyników nauczania po pierwszym etapie kształcenia w Polsce. Pomocniczym pytaniem w odniesieniu do podjętego problemu jest pytanie o to, jak bardzo szkoły różnią się między sobą ze względu na cechy uczniów, którzy się w nich uczą, a które mogą mieć znaczenie dla ich osiągnięć. Ostatnim było pytanie o to, jakie jest zróżnicowanie międzyszkolne efektywności nauczania po pierwszym etapie kształcenia.

Dane

Dane wykorzystane do analiz omawianego tu zagadnienia pochodzą z *Badania szkolnych uwarunkowań efektywności kształcenia* realizowanego przez Instytut Badań Edukacyjnych w ramach projektu „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego” współfinansowanego ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Kapitał Ludzki. Jest to studium podłużne realizowane na ogólnopolskiej, losowej próbie 172 szkół podstawowych¹, w których badaniem zostało objętych łącznie 300 oddziałów. Badanie rozpoczęło się w roku szkolnym 2010/11 w trzecich klasach szkół podstawowych. Celem projektu jest identyfikacja kluczowych czynników wpływających na efektywność kształcenia. Pierwsza faza projektu koncentrowała się na nauczaniu zintegrowanym i składała się z dwóch etapów badawczych. Podczas pierwszego etapu, który odbył się w III klasie, dokonano m.in. pomiaru cech uczniów i ich rodzin. Natomiast

¹ Z operatu losowania zostały wykluczone szkoły specjalne i integracyjne, szkoły mistrzostwa sportowego i sportowe, przyszpitalne oraz eksperymentalne (pojedyncza szkoła w skali kraju), a także szkoły o prognozowanej liczbie uczniów poniżej 11 osób. Dodatkowo wyłączono także oddziały integracyjne.

podczas drugiego etapu, zrealizowanego na początku IV klasy, dokonano pomiaru osiągnięć szkolnych. Poniżej scharakteryzowano krótko te zmienne, które zostały wykorzystane w prezentowanych analizach.

Osiągnięcia szkolne. Pomiar osiągnięć szkolnych został przeprowadzony podczas drugiego etapu badania, na początku roku szkolnego 2011/12 (początek IV klasy) za pomocą standaryzowanych testów opracowanych specjalnie na potrzeby badania (Jasińska i Modzelewski, 2012a) i objął umiejętność czytania, świadomość językową i umiejętności matematyczne. Wykorzystane w badaniu testy charakteryzują się wysoką rzetelnością; współczynniki rzetelności wyliczone z modeli IRT, będące stosunkiem wariancji oszacowań wyników EAP dla uczniów i wariancji umiejętności w populacji (patrz: Adams, 2005) wynoszą 0,86-0,88. Testy mają także udokumentowaną trafność (Jasińska i Modzelewski, 2012b). W celu wyznaczenia poziomu umiejętności uczniów dla każdej z trzech skal wyspecyfikowano model IRT (model Rascha) z regresją latentną uwzględniającą zmienne, dla których dane zostały zebrane na pierwszym etapie badania (Wu, 2005). Za wskaźnik poziomu badanych umiejętności przyjęto estymatory *plausible values* (PV) (patrz np.: Adams, 2005; Wu, 2005), które zostały wyrażone na skali o średniej równej 100 i odchyleniu standardowym równym 15 w populacji uczniów objętych badaniem.

Inteligencja. Poziom inteligencji uczniów został zmierzony na pierwszym etapie badania, w roku szkolnym 2010/11, kiedy uczniowie uczęszczali do III klasy. Do pomiaru inteligencji wykorzystano Test Matryc Ravena, w wersji Standard, formie klasycznej (Jaworowska i Szustrowa, 1991). Wyniki tego testu zostały wyskalowane wielogrupowym dwuparametrycznym modelem IRT z wykorzystaniem programu MIRT (Glas, 2010). W analizach wykorzystano estymatory EAP (*expected a posteriori*) (patrz np.: de Ayala, 2009).

Wiek i płeć. Dane o wieku i płci uczniów pochodzą z ankiety rodzicielskiej wykorzystanej w pierwszym etapie badania. Wiek uczniów wyrażono w tygodniach w czasie pisania testów osiągnięć szkolnych.

Status społeczno-ekonomiczny rodziny pochodzenia ucznia. W analizach uwzględniono trzy wskaźniki statusu rodziny ucznia. Informacje potrzebne do ich zbudowania zebrano za pomocą ankiet rodzicielskich. Pierwszym z nich jest standardowy wskaźnik najwyższego statusu zawodowego rodziców (HISEI) wykorzystywany m.in. także w badaniu PISA (por. OECD, 2010). Na podstawie odpowiedzi na serię standardowych pytań zakodowano wykonywane przez rodziców zawody zgodnie z klasyfikacją ISCO-08. Następnie poszczególnym wartościom ISCO-08 przypisano odpowiadające im wartości indeksu statusu społeczno-ekonomicznego ISEI-08. Wskaźnik HISEI zbudowano w taki sposób, że przyjmuje on wyższą wartość ze wskaźników ISEI obojga rodziców lub jedyną dostępną, jeśli dla jednego z rodziców nie posiadano danych.

Drugim wskaźnikiem jest indeks zasobów materialnych rodziny dziecka. Został on opracowany w oparciu o odpowiedzi na pytania ułożone na podstawie analogicznych pytań w kwestionariuszach uczniowskich i rodzicielskich badania PISA (OECD, 2010). Dostosowanie koncepcji tego wskaźnika do potrzeb badania SUEK polegało na uwzględnieniu w nim większej liczby pozycji

odnoszących się do dóbr mogących wspierać dziecko w nauce, a zrezygnowaniu z kilku powszechnie posiadanych w Polsce ogólnych dóbr materialnych. Strukturę danych zbadano za pomocą eksploracyjnej analizy czynnikowej dla zmiennych porządkowych. Do indeksu włączono te pozycje, które istotnie były związane z wymiarem posiadania ogólnych zasobów. Następnie wartości indeksu wyliczane były przy pomocy estymatorów EAP w skalowaniu IRT modelem 2PL/GRM, metodą MML (*Marginal Maximum Likelihood*).

Trzecim wskaźnikiem jest wykształcenie rodziców wyrażone w latach nauki potrzebnych przeciętnie do osiągnięcia danego poziomu wykształcenia. Przyjmuje on wyższą wartość ze wskaźników obojga rodziców lub jedyną dostępną, jeśli dla jednego z rodziców nie posiadano danych.

Schemat analizy danych

Oszacowanie zróżnicowania międzyszkolnego wyników uczniów zostało wyliczone z użyciem dwóch rodzajów dwupoziomowych hierarchicznych modeli liniowych: analizy wariancji z efektami losowymi oraz analizy kowariancji z efektami losowymi (Raudenbush i Bryk, 2002; Snijders i Bosker, 2012). Modele te różnią się od klasycznej regresji liniowej tym, że podczas estymacji wykorzystywana jest informacja o podziale uczniów na grupy (nie ma potrzeby włączania do modelu zmiennych typu „dummy” informujących o przynależności ucznia do danej szkoły). Informacja o wpływie tego pogrupowania na zmienną zależną znajduje swe odbicie w komponencie losowym modelu. Tym samym, modele te są efektywną i adekwatną metodą statystyczną służącą do analizy danych uzyskanych w próbie wylosowanej z wykorzystaniem złożonych schematów doboru próby oraz przejawiających hierarchiczną strukturę.

W hierarchicznych modelach liniowych efekty losowe na poziomie grup (w tym przypadku szkół) nie są szacowane podczas procesu estymacji. Estymowana jest ich wariancja (τ^2) i ewentualnie kowariancja z innymi efektami losowymi. W analizie wariancji z efektami losowymi, wariancja efektów losowych dla konkretnych poziomów mówi o nam o szacowanej wielkości zróżnicowania wyników, które możemy przypisać poszczególnym poziomom. Analiza kowariancji z efektami losowymi wylicza te same parametry, jednakże po uwzględnieniu związku zmiennej zależnej z tymi zmiennymi niezależnymi, dla których wyliczono efekty stałe lub losowe. W przypadku analiz wykonanych na potrzeby niniejszego artykułu, w ramach analizy kowariancji z efektami losowymi wpływ zmiennych niezależnych był estymowany jako efekty stałe bez komponentu losowego.

Na potrzeby przygotowania wykresów ze średnimi wynikami szkół wyznaczono bayesowskie predykcje *a posteriori* (*empirical Bayes estimation*). Metoda ta pozwala wyznaczyć wartości efektów losowych dla szkół na podstawie dwóch informacji: danych o jednostkach przynależących do konkretnej grupy (wynikach uczniów z danej szkoły) i założenia modelu o normalnym rozkładzie efektów losowych na poziomie szkół o średniej 0 i wariancji równej τ^2 (Snijders, Bosker, 2012, s. 62–67). Błędy standardowe zostały wyznaczone jako pierwiastek z wariancji rozkładu *a posteriori*.

Zastosowane modele mają szczególną zaletę w świetle przeprowadzanych analiz zróżnicowania średnich wyników dla szkół, są bowiem dosyć odporne na artefaktualne zwiększanie wariancji międzygrupowej, wynikającej z wahań w liczebnościach poszczególnych grup (szkół). Dzieje się tak dlatego, że średnie dla szkół są ściągane do średniej populacyjnej tym bardziej, im mniej uczniów stanowiło podstawę do jej wyliczenia. Wydaje się to rozsądnym założeniem – średnia otrzymana z wyników dla małej liczby uczniów jest mniej rzetelna niż ta, do której obliczenia wzięto wyniki kilkudziesięciu uczniów.

Analizy wykonano za pomocą oprogramowania HLM6.06 (Raudenbush, Bryk i Congdon, 2008). Posłużono się metodą estymacji *full maximum likelihood* oraz wykorzystano warunkowe wagi próbkowania z poprawką na *response rate*.

Wyniki

Zróżnicowanie międzyszkolne wyników nauczania po pierwszym etapie kształcenia

Zacznijmy od przyjrzenia się poziomowi zróżnicowania międzyszkolnego ze względu na osiągnięcia szkolne po pierwszym etapie kształcenia, czyli na zakończenie nauczania zintegrowanego. W tabeli 1 przedstawiono oszacowania wariancji efektów losowych dla poziomu uczniów i szkół uzyskane za pomocą dwupoziomowych modeli pustych z losową stałą. Wariancja efektów szkół dla każdego testu osiągnięć jest statystycznie istotnie różna od zera. Wskaźnik zróżnicowania międzyszkolnego uzyskano, dzieląc wariancję efektów losowych na poziomie szkół przez wariancję całkowitą, a wynik wyrażono w procentach.

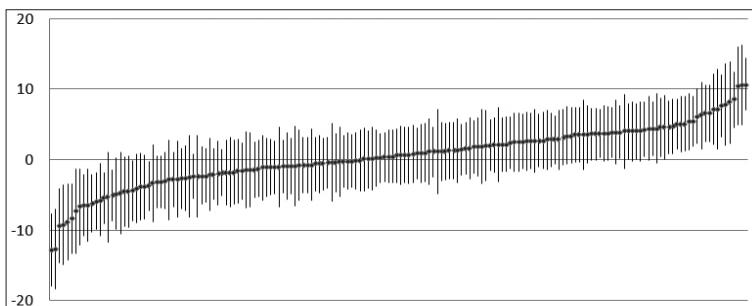
Tabela 1. Zróżnicowanie międzyszkolne wyników nauczania po I etapie kształcenia – oszacowania na podstawie dwupoziomowych modeli z losową stałą (analiza wariancji z efektami losowymi)

	Test umiejętności czytania	Test świadomości językowej	Test umiejętności matematycznych
<i>Oszacowanie efektów losowych</i>			
Wariancja efektów szkół	21,77	28,97	24,85
Wariancja na poziomie uczniów	199,62	197,63	199,61
Wskaźnik zróżnicowania międzyszkolnego	9,84%	12,79%	11,07%
Liczba uczniów	4739	4739	4634
Liczba szkół	172	172	172

Źródło: opracowanie własne

Analizy te pokazują, że podział uczniów na szkoły wyjaśnia dosyć dużą część zróżnicowania wyników nauczania po pierwszym etapie kształcenia. Oszacowane zróżnicowanie międzyszkolne waha się w zależności od testu – dla testu czytania wynosi niecałe 10%, a dla testu świadomości językowej prawie 13%, w teście matematycznym podział uczniów ze względu na szkoły odpowiada za ok. 11% całkowitego zróżnicowania wyników. Żeby dopowiedzieć

na pytanie o to, jak interpretować te wielkości, warto przedstawić oszacowane zróżnicowanie graficznie. Na rysunku 1 przedstawiono oszacowania średnich wyników szkół dla przykładowego testu – testu umiejętności czytania – wraz z 95% przedziałem ufności dla oszacowanej średniej. Wyniki szkół uporządkowano rosnąco ze względu na uzyskany średni wynik. Na osi pionowej przedstawiono skalę umiejętności czytania, w której średnia wyników w populacji uczniów wynosi 0, a odchylenie standardowe wyników indywidualnych 15. Z oszczędności miejsca ograniczono się do prezentacji wykresu tylko dla jednego testu. Dla pozostałych testów wykresy wyglądają bardzo podobnie.



Źródło: opracowanie własne

Rysunek 1. Oszacowania średnich wyników szkół wraz z przedziałami ufności dla testu umiejętności czytania (średni wynik w populacji = 0, odchylenie standardowe = 15)

Wykres ten pozwala graficznie zobrazować zaobserwowane w próbie zróżnicowanie średnich wyników dla szkół. Im bardziej stromo układałyby się średnie wyniki dla szkół, z tym bardziej zróżnicowanym systemem szkolnym mielibyśmy do czynienia. Na wykresie możemy dostrzec, że większość szkół nie różni się znacząco wynikami nauczania, ponieważ przedziały ufności dla ich średnich wyników nachodzą na siebie. Jedynie na skrajach możemy dostrzec szkoły, których średnie wyniki są istotnie statystycznie różne od średnich wyników w populacji, a także różne od wyników szkół z drugiego krańca. Szkół takich jest ok. 27%. Jednak mimo tego, że większość szkół osiąga średnio podobne wyniki, jeśli spojrzymy na wielkość różnicy między szkołami o najniższych i najwyższych wynikach, to zobaczymy, że wynosi ona trochę ponad 20 punktów, czyli jeden i jedna trzecia odchylenia standardowego wyników indywidualnych! Wyniki te pokazują, że choć oszacowane wartości wskaźnika zróżnicowania międzyszkolnego dla testów osiągnięć szkolnych SUEK mogą wydawać się niewysokie, to nie oznacza to, że w przypadku każdej pary szkół różnice te są niewielkie.

Zróżnicowanie międzyszkolne wyników nauczania po pierwszym etapie kształcenia w badaniu PIRLS i TIMSS

W roku 2011 w 69 krajach odbyły się kolejne edycje międzynarodowych badań poświęconych analizie osiągnięć szkolnych uczniów TIMSS (*Trends in International Mathematics and Science Study*) i PIRLS (*Progress in International Reading Literacy Study*). Inicjatorem i organizatorem obu badań jest założone

w 1967 r. Międzynarodowe Towarzystwo Mierzenia Osiągnięć Szkolnych (IEA), a jednym z głównych celów badań jest dostarczenie rzetelnych i porównywalnych międzynarodowo danych dotyczących osiągnięć szkolnych – z zakresu matematyki i nauk przyrodniczych w przypadku badania TIMSS oraz umiejętności czytania, gdy mówimy o badaniu PIRLS. W Polsce w roku 2011 badanie PIRLS odbyło się po raz drugi (pierwszą edycją badania, w której wzięła udział Polska było badanie w 2006), w badaniu TIMSS nasz kraj uczestniczył po raz pierwszy.

Badanie PIRLS i TIMSS objęło w 2011 roku kohortę uczniów rok starszą od tej biorącej udział w badaniu SUEK. W badaniu wzięło udział 150 szkół podstawowych, w których przebadano ponad 5 tysięcy uczniów klas trzecich (Konarzewski, 2012). Zbieżność zarówno daty badań, jak i badanych populacji, a także sprawdzanego zakresu umiejętności pozwala na porównanie wyników analizy zróżnicowania międzyszkolnego osiągnięć uczniów pomiędzy badaniami SUEK a badaniami organizowanymi przez IEA. W raporcie z badań załączone są wyniki analiz modeli trzypoziomowych, które pozwalają na przyjrzenie się rozkładowi wariancji wyników pomiędzy poziomami szkół, klas i uczniów. W związku z tym, że w niniejszym artykule nasze zainteresowanie skoncentrowane jest na różnicach międzyszkolnych jako pewnej ogólnej charakterystyce systemu szkolnego, dostępne wyniki badania PIRLS i TIMSS przeliczono w ramach modeli dwupoziomowych w sposób analogiczny do analiz wyników badania SUEK. Tabela 2 zawiera zestawienie oszacowań wariancji wyników dla części matematycznej badania TIMSS oraz umiejętności czytania mierzonej w badaniu PIRLS, a także oszacowania wskaźników zróżnicowania międzyszkolnego wyników tych testów.

Tabela 2. Zróżnicowanie międzyszkolne wyników nauczania pod koniec I etapu kształcenia – oszacowania na podstawie dwupoziomowych modeli z losową stałą (analiza wariancji z efektami losowymi) na podstawie wyników badań PIRLS i TIMSS z roku 2011

	Test umiejętności czytania	Test umiejętności matematycznych
Wariancja efektów szkół	643,13	730,67
Wariancja na poziomie uczniów	4637,48	4608,53
Wskaźnik zróżnicowania międzyszkolnego	12,18%	13,69%
Liczba uczniów	5005	5027
Liczba szkół	150	150

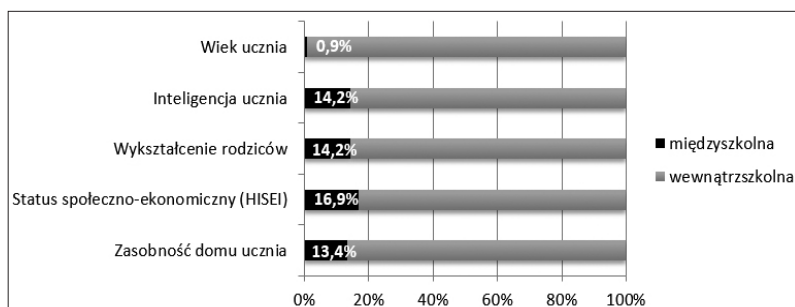
Źródło: opracowanie własne

Zaprezentowane dane z badań PIRLS i TIMSS potwierdzają wyniki analiz zróżnicowania międzyszkolnego uzyskane na podstawie danych z badania SUEK. Drobne różnice, wielkości 2-3 punktów procentowych sugerują możliwość niedoszacowania zróżnicowania osiągnięć po pierwszym etapie kształcenia w badaniu SUEK. Hipoteza ta jest o tyle prawdopodobna, że badanie reprezentatywności przeprowadzone po I etapie badania SUEK z wykorzystaniem danych ze sprawdzianu szóstoklasisty wykazało, że w wyniku odmów

dyrektorów szkół na udział w badaniu, w ostatecznej, zbadanej próbie szkół zabrakło – w stosunku do próby zamierzonej – placówek o bardzo niskich i bardzo wysokich średnich wynikach na sprawdzianie (Małuchnik, Słowicki, Modzelewski i 2012). Nie zmienia to jednak faktu, że zarówno wyniki badań IEA, jak i badania SUEK ukazują obraz w średnim stopniu zróżnicowanego systemu szkolnego. Należy przy tym pamiętać, że mówimy tu o zróżnicowaniu obserwowanym już po I etapie kształcenia – zaledwie cztery lata (wliczając jeden rok obowiązkowego wychowania przedszkolnego) po powszechnym rozpoczęciu przez uczniów edukacji.

Zróżnicowanie międzyszkolne charakterystyk uczniów mających znaczenie dla wyników nauczania

Jak zasygnalizowano we wstępie, przyczyn zróżnicowania międzyszkolnego wyników nauczania można upatrywać w dwóch zjawiskach. Po pierwsze może być ono konsekwencją tego, że jedne szkoły uczą efektywniej niż inne i tym samym dają lepsze szanse swoim uczniom na osiągnięcie wyższych wyników. Z drugiej jednak strony, zróżnicowanie to może wynikać z tego, że niektóre szkoły pracują z uczniami o korzystniejszych cechach, mających znaczenie dla osiągnięć szkolnych, a inne z uczniami pochodzącymi z mniej sprzyjających środowisk. W takiej sytuacji, nawet jeśli szkoły uczyłyby tak samo dobrze, z taką samą efektywnością, uzyskiwałyby różne wyniki w testach osiągnięć. W celu oszacowania zróżnicowania międzyszkolnego efektywności nauczania należy więc uwzględnić w modelu te cechy uczniów, które wpływają na osiągnięcia szkolne, a które są nierównomiernie rozdystrybuowane między szkołami. Zanim poznamy wyniki analiz takiego modelu, przyjrzymy się zatem zróżnicowaniu międzyszkolnemu wybranych charakterystyk uczniów, które mają znaczenie dla osiągnięć szkolnych. Pozwoli to na weryfikację, które ze zmiennych należy uwzględnić w modelowaniu dla poprawnego oszacowania zróżnicowania międzyszkolnego efektywności nauczania. Jeśli bowiem szkoły nie różniłyby się znacząco między sobą średnim poziomem jakiejś cechy uczniów, to nawet gdyby bardzo silnie determinowała ona osiągnięcia szkolne, nie miałyby większego znaczenia dla oszacowania zróżnicowania międzyszkolnego wyników nauczania.



Rysunek 2. Zróżnicowanie szkół ze względu na cechy uczniów i ich rodzin. Procent wariacji międzyszkolnej w stosunku do wariacji całkowitej zmiennych

Na rysunku 2 pokazano, jaką część zróżnicowania danej zmiennej w populacji uczniów możemy przypisać podziałowi uczniów na szkoły. Pozostała część to wariancja wewnątrzszkolna. Im większa wartość odsetka wariancji międzyszkolnej, tym szkoły bardziej różnią się między sobą ze względu na średni poziom danej cechy uczniów. Zróżnicowanie międzyszkolne każdej z analizowanych zmiennych jest istotne statystycznie na poziomie istotności $p < 0,001$ ². Wyniki te pokazały, że spośród charakterystyk uczniów i ich rodzin uwzględnionych w analizach, szkoły najbardziej różnią się między sobą ze względu na status społeczno-ekonomiczny mierzony wskaźnikiem HISEI. Podział na szkoły wyjaśnia 17% wariancji tej zmiennej. Niewiele mniejsze jest zróżnicowanie międzyszkolne pozostałych miar statusu, jakimi są wykształcenie rodziców i zasobność domu ucznia. Zaobserwowano także znaczące zróżnicowanie międzyszkolne poziomu inteligencji uczniów, które jest na podobnym poziomie jak zróżnicowanie wykształcenia rodziców i wynosi 14%. Szkoły natomiast w bardzo niewielkim stopniu różnią się między sobą ze względu na wiek uczęszczających do nich uczniów. Odsetek wariancji wieku uczniów, który możemy wyjaśnić podziałem na szkoły, mimo iż jest istotnie statystycznie większy od zera, ma pomijalne znaczenie praktyczne, wynosi bowiem niecały 1%.

Zaprezentowane tu wyniki pokazały, że szkoły istotnie różnią się między sobą cechami uczniów, które mają znaczenie dla osiągnięć szkolnych. Różnice te są widoczne w szczególności w przypadku statusu społeczno-ekonomicznego oraz inteligencji uczniów. Oznacza to, że za część obserwowanego zróżnicowania międzyszkolnego wyników nauczania może odpowiadać to, że szkoły pracują z uczniami o różnych cechach determinujących w pewnym stopniu sukces edukacyjny. Jeśli chcemy zatem oszacować, jak bardzo szkoły różnią się między sobą efektywnością nauczania, powinniśmy wytrącić z modelu zróżnicowanie tych cech.

Zróżnicowanie międzyszkolne efektywności nauczania po pierwszym etapie kształcenia

W celu oszacowania tego, jak bardzo szkoły różnią się między sobą efektywnością nauczania, wyspecyfikowano modele, w których kontrolowano znaczenie tych zmiennych, które są istotnie statystycznie nierównomiernie rozdystrybuowane między szkołami, a mogących kształtować osiągnięcia szkolne uczniów. W tabeli 3 przedstawiono oszacowania efektów losowych z dwupoziomowych modeli z losową stałą i efektami stałymi dla wymienianych wcześniej zmiennych kontrolnych: wiek, inteligencja, wykształcenie rodziców, HISEI oraz zasobność domu ucznia. Dla czytelności tabeli nie pokazano oszacowań efektów stałych, które nie są przedmiotem naszej analizy. Wariancja efektów szkół dla każdego testu osiągnięć jest statystycznie istotnie różna od zera.

² Prezentowane dane pochodzą z pięciu niezależnych modeli. Analizy dla każdego modelu przeprowadzono dla wszystkich dostępnych danych, uwzględniając dla kolejnych zmiennych następującą liczbę uczniów: 5081, 4975, 5484, 4930, 5121. Uczniowie byli zgrupowani w 172 szkołach.

Tabela 3. Zróżnicowanie międzyszkolne efektywności nauczania po I etapie kształcenia – oszacowania na podstawie dwupoziomowego modelu z losową stałą i efektami stałymi dla zmiennych kontrolnych (analiza kowariancji z efektami losowymi)

	Test umiejętności czytania	Test świadomości językowej	Test umiejętności matematycznych
<i>Oszacowanie efektów losowych</i>			
Wariancja efektów szkół	13,07	13,70	13,71
Wariancja na poziomie uczniów	136,06	117,45	111,63
Wskaźnik zróżnicowania międzyszkolnego	8,76%	10,45%	10,93%
Liczba uczniów	4725	4725	4784
Liczba szkół	172	172	172

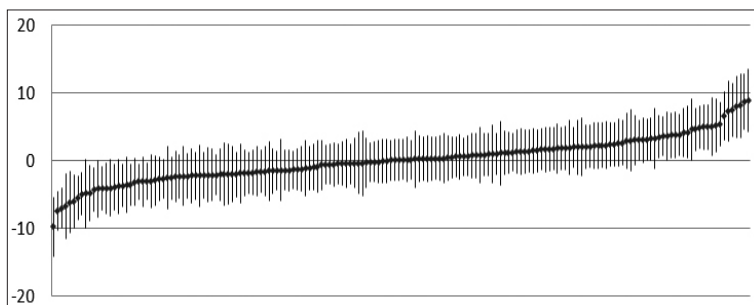
Źródło: opracowanie własne

Wyniki te pokazują, że jeśli kontrolujemy cechy biologiczno-społeczne uczniów, to zróżnicowanie międzyszkolne osiągnięć szkolnych jest trochę niższe. Tak więc, zgodnie z przewidywaniami, za część zróżnicowania międzyszkolnego wyników nauczania odpowiada zróżnicowanie międzyszkolne w zakresie potencjału uczniów na wejściu. Uzyskane tu wyniki informują o tym, na ile szkoły różniłyby się między sobą ze względu na wyniki nauczania, jeśli wszystkie pracowałyby z uczniami o takich samych cechach biologiczno-społecznych. Można je więc interpretować w kategoriach zróżnicowania w zakresie wkładu szkoły w postępy uczniów niezależnie od ich potencjału czy – inaczej mówiąc – w zakresie efektywności nauczania. Pokazują one, że nawet jeśli kontrolujemy potencjał uczniów, podział na szkoły wyjaśnia znaczną część zróżnicowania wyników.

Dodanie do modelu zmiennych kontrolnych zredukowało jednak nie tylko wariancję efektów na poziomie szkół, ale w dużej mierze także wariancję na poziomie uczniów. Oznacza to, że sporą część zróżnicowania indywidualnego osiągnięć szkolnych możemy przypisać cechom biologicznym uczniów i środowisku rodzinnemu, z którego pochodzą. Redukcja wariancji na poziomie uczniów, poza tą oczywistą interpretacją, prowadzi jednak do jeszcze jednej konsekwencji – zmniejszenia błędu standardowego oszacowania efektów grupowych. Co to oznacza? Spójrzmy na rysunek 3, na którym przedstawiono oszacowania wskaźników efektywności nauczania³ dla testu umiejętności czytania wraz z 95% przedziałem ufności. Wyniki te w pewnym uproszczeniu mówią o tym, o ile średnio punktów został przesunięty rozkład wyników indywidualnych z zakresu umiejętności czytania w danej szkole w stosunku do rozkładu przewidywanego na podstawie cech uczniów uwzględnionych w modelu i zależności między tymi zmiennymi a zmienną zależną w populacji. Skala,

³ Między tak skonstruowaną miarą a klasycznymi wskaźnikami edukacyjnej wartości dodanej (EWD) jest spore podobieństwo wynikające z faktu uwzględniania w modelu zmiennych świadczących o potencjale uczniów. W szacowaniu wskaźników EWD bierze się pod uwagę uprzednie osiągnięcia szkolne jako podstawową charakterystykę uczniów wyznaczającą szansę na dobry wynik na teście podsumowującym kolejny etap kształcenia. Tak skonturowana miara, jak ta wyliczona na potrzeby analiz zaprezentowanych w tym tekście, nazywana jest w literaturze *kontekstualną miarą osiągnięć* (por. np. Lenkeit, 2013; Organisation for Economic Co-operation and Development, 2008).

na której przedstawiono wyniki, jest skalą indywidualnych wyników w teście czytania, o odchyleniu standardowym w populacji równym 15. Średnia w populacji tych efektów jest równa 0. Inaczej mówiąc, wyniki zobrazowane na wykresie pokazują, ile średnio punktów w teście czytania dana szkoła dodała (lub odjęła) uczniom w stosunku do wyników przewidywanych na podstawie ich potencjału. Efekty te zostały oszacowane w sposób analogiczny do omawianego dla modelu pustego z tym zastrzeżeniem, że zostały one policzone na podstawie modelu uwzględniającego zmienne wyjaśniające.



Źródło: opracowanie własne

Rysunek 3. Oszacowania wskaźników efektywności nauczania szkół wraz z przedziałami ufności dla testu umiejętności czytania

Jak widzimy, na podstawie porównania wykresów znajdujących się na rysunkach 1 i 3 spadek wariancji międzyszkolnej uwidacznia się w bardziej płaskiej linii, która rysuje się nam z kropek oznaczających średnie dla wyznaczonych efektów. Widzimy jednak także, że przedziały ufności są węższe, dzięki czemu nadal ok. 24% szkół ma – wyznaczony w ten sposób – wskaźnik efektywności nauczania istotnie różny od przeciętnej. Nadal też różnica pomiędzy najlepszą i najsłabszą szkołą jest niemała, bo trochę ponad jedno odchylenie standardowe. Dla testu matematycznego jest niewiele większa, a dla testu świadomości językowej w granicach półtora odchylenia standardowego.

Podsumowanie

Miara zróżnicowania międzyszkolnego wyników nauczania na danym etapie kształcenia jest przydatnym wskaźnikiem charakteryzującym system szkolny. Pozwala na określenie, jak bardzo szkoły różnią się między sobą wynikami swoich uczniów. Tak więc może ona służyć za wskaźnik jednolitości systemu kształcenia. Z drugiej strony, jest ona punktem wyjścia dla badań mających na celu poszukiwanie szkolnych uwarunkowań wyników nauczania. Obserwowane zróżnicowanie międzyszkolne badacze starają się wyjaśnić charakterystykami szkoły, nauczycieli czy cechami procesów dydaktycznych.

Powyższe analizy pokazują, że zróżnicowanie międzyszkolne wyników nauczania w zakresie umiejętności czytania, świadomości językowej i umiejętności matematycznych po pierwszym etapie kształcenia jest znaczące. Zmniejsza się ono jednak, jeśli policzymy je, kontrolując cechy uczniów i rodzin, z których

pochodzą. Oznacza to, że w pewnym stopniu zróżnicowanie międzyszkolne wyników na poziomie nauczania zintegrowanego jest rezultatem zróżnicowania międzyszkolnego potencjału uczniów na wejściu. Jednak nawet po wytrąceniu tego zróżnicowania szkoły różnią się wciąż w istotnym stopniu efektywnością nauczania, co prowadzi do znacznych różnic, jeśli porównamy szkoły o skrajnych wynikach. Biorąc pod uwagę, że mówimy o podsumowaniu efektów kształcenia po III klasie szkoły podstawowej, obserwowane wyniki sugerują istnienie niepokojącego zjawiska – zróżnicowanej efektywności nauczania szkół już na samym początku formalnego procesu kształcenia.

Bibliografia

1. Adams, R. J. (2005), *Reliability as a Measurement Design Effect*, *Studies in Educational Evaluation*, 31(2), 162–172.
2. De Ayala, R. J. (2009), *The theory and practice of item response theory*, New York: Guilford Press.
3. Dolata, R. (2011), *Analiza różnicowania się systemu oświaty w Polsce na poziomie szkół podstawowych i gimnazjów, edycja 2011*. Pobrano 19 marzec 2012, z http://ewd.edu.pl/downloads/publikacje/analiza_roznicowania_2011.pdf
4. Dolata, R., Jasińska, A., Modzelewski, M. (2012), *Wykorzystanie krajowych egzaminów jako instrumentu polityki oświatowej na przykładzie procesu różnicowania się gimnazjów w dużych miastach*, „Polityka Społeczna”, nr tematyczny 1, 41–46.
5. Glas, C. A. (2010). *Preliminary manual of the software program Multidimensional Response Theory (MIRT)*.
6. Jasińska, A., Modzelewski, M. (2012a), *Można inaczej. Wykorzystanie IRT do konstrukcji testów osiągnięć szkolnych*, [w:] B. Niemierko, M. K. Szmigel (red.), *Regionalne i lokalne diagnozy edukacyjne, Wrocław, 21–23 września 2012 r.* (ss. 157–168).
7. Jasińska, A., Modzelewski, M. (2012b). *Pomiar wyników nauczania w zakresie czytania, świadomości językowej i matematyki*, [w:] *Szkolne, psychologiczne i społeczne uwarunkowania wyników nauczania na I etapie edukacyjnym. Raport na podstawie fazy wstępnej ogólnopolskiego badania podłużnego Szkolnych Uwarunkowań Efektywności Kształcenia*. Warszawa: Instytut Badań Edukacyjnych.
8. Jaworowska, A., Szustrowa, T. (1991), *Podręcznik do Testu Matryc Ravena. Wersja Standard (1956), Polska standaryzacja 1989 (5;11-15;11)*, Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
9. Konarzewski, K. (2012), *TIMSS i PIRLS 2011. Osiągnięcia szkolne polskich trzecioklasistów w perspektywie międzynarodowej*, Warszawa: CKE.
10. Lenkeit, J. (2013), *Effectiveness measures for cross-sectional studies: a comparison of value-added models and contextualised attainment models*, *School Effectiveness and School Improvement*, 24(1), 1–25. doi:10.1080/09243453.2012.680892
11. Maluchnik, M., Słowicki, L., Modzelewski, M. (2012), *Próba badawcza i sposób zbierania danych*, [w:] *Szkolne, psychologiczne i społeczne uwarunkowania wyników nauczania na I etapie edukacyjnym. Raport na podstawie fazy wstępnej ogólnopolskiego badania podłużnego Szkolnych Uwarunkowań Efektywności Kształcenia*. Warszawa: Instytut Badań Edukacyjnych.
12. OECD (2010), *PISA 2009 Results: Overcoming Social Background Equity in Learning Opportunities and Outcomes (Volume II): Equity in Learning Opportunities and Outcomes*. OECD Publishing.

13. Organisation for Economic Co-operation and Development (2008), *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paris: OECD.
14. Raudenbush, S. W., Bryk, A. S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Thousand Oaks, CA: SAGE.
15. Raudenbush, S. W., Bryk, A. S., Congdon, R. (2004). *HLM 6 for Windows [Computer software]*. Skokie, IL: Scientific Software International.
16. Snijders, T. A. B., Bosker, R. J. (2012), *Multilevel analysis: an introduction to basic and advanced multilevel modeling* (2nd ed.), Los Angeles: Sage.
17. Wu, M. (2005), *The role of plausible values in large-scale surveys*, *Studies in Educational Evaluation*, 31(2), 114–128.