

Szacowanie efektu nauczyciela na osiągnięcia edukacyjne uczniów z wykorzystaniem hierarchicznego modelowania liniowego

MACIEJ KONIEWSKI

Pracownia Edukacyjnej Wartości Dodanej, Instytut Badań Edukacyjnych*

Nauczyciele, różniąc się efektywnością pracy, wpływają na wyniki uczniów. Analizowano dane z ogólnopolskiego badania przeprowadzonego w gimnazjach w 2012 r. Analizy objęły 3883 uczniów z 246 oddziałów, ze 137 szkół i 202 nauczycieli matematyki oraz 4119 uczniów z 260 oddziałów, ze 143 szkół i 215 nauczycieli języka polskiego. Wariancja wyników egzaminów przypisana nauczycielom wyniosła 12% (matematyka) i 8% (kompetencje językowe). W modelu uwzględniającym m.in.: wcześniejsze osiągnięcia, inteligencję, zmienne statusowe rodziny ucznia, efekt nauczyciela wyjaśnił 5% (matematyka) i 4% (kompetencje językowe) wariancji wyników egzaminu. Do tej pory nie było jasne, za pomocą jakich cech nauczycieli i stylu nauczania można wyczerpująco wyjaśnić zróżnicowanie efektywności ich pracy. Skala „autorytet nauczyciela/utrzymanie dyscypliny” wyjaśniła 91% efektu nauczyciela w nauczaniu matematyki i 81% w nauczaniu języka polskiego, przy kontroli wcześniejszych osiągnięć edukacyjnych, inteligencji, zmiennych statusowych rodziny ucznia i lokalizacji gimnazjum.

SŁOWA KLUCZOWE: socjologia edukacji, efekt nauczyciela, szkolne uwarunkowania efektywności kształcenia, modele hierarchiczne, HLM.

Co najmniej od 1902 r. (Rice) badacze poszukują czynników najlepiej wyjaśniających osiągnięcia szkolne uczniów. Od 1966 r. (Coleman i in.) dominowała teza, że czynnikiem tym jest rodzina ucznia i jej miejsce w strukturze społecznej. Teza ta, wsparta licznymi badaniami (Blau i Duncan, 1967;

Bourdieu, 1986; Bourdieu i Passeron, 1990; Mosteller i Moynihan, 1972; Plowden, 1967), jest obecnie krytykowana (Byrne i in., 2010; Hart, Petrill i Kamp Dush, 2009; Kavas, Haworth, Petrill i Plomin, 2007; McGue i in., 2007; Oliver i in., 2004). Przywołani autorzy wskazują, że po uwzględnieniu czynnika genetycznego, wpływ rodziny na osiągnięcia

Artykuł powstał na podstawie rozdziału „Nauczycielskie uwarunkowania efektywności kształcenia” opublikowanego w raporcie *Trafność metody edukacyjnej wartości dodanej. Raport podsumowujący wyniki badania podłużnego w gimnazjach*. Rozdział został napisany na zlecenie Instytutu Badań Edukacyjnych (IBE) w ramach projektu „Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD)”,

współfinansowanego przez Unię Europejską ze środków Europejskiego Funduszu Społecznego.

* Adres do korespondencji: Maciej Koniewski, Pracownia Edukacyjnej Wartości Dodanej, Instytut Badań Edukacyjnych, ul. Górczewska 8, 01-180 Warszawa. E-mail: m.koniewski@ibe.edu.pl

uczniów jest znacznie mniejszy, niż pierwotnie przypuszczano.

Praktycznie zorientowani badacze, chcąc wspomagać wynikami swoich badań decyzje twórców polityk edukacyjnych, poszukują predyktorów wyników nauczania wśród zmiennych, na które można wpływać decyzjami politycznymi i finansowymi. Badania pokazują, że wśród czynników kontrolowanych przez system szkolny najbardziej obiecująca w wyjaśnianiu osiągnięć szkolnych jest efektywność nauczycieli (Akiba, LeTendre i Scribner, 2007; Hanushek, 1997; Jordan, Mendro i Weerasinghe, 1997; Rivkin, Hanushek i Kain, 2005; Sanders i Rivers, 1996; Taylor i in., 2010; Wright, Horn i Sanders, 1997).

Artykuł odpowiada na pytania: (a) Jak duży jest efekt nauczyciela, czyli w jakim stopniu nauczyciele mogą przyczyniać się do podnoszenia osiągnięć szkolnych uczniów? (b) Czy nauczyciele i szkoły istotnie różnią się w swoich możliwościach podnoszenia osiągnięć szkolnych uczniów? Jeśli tak, to (c) czy różnice te można wyjaśnić przez charakterystyki nauczycieli (np. staż pracy, wykształcenie) i ich pracy (np. style prowadzenia lekcji)?

Odpowiedzi na pytania badawcze udzielono na podstawie wyników analizy danych dotyczących nauczycieli matematyki i języka polskiego, którzy uczyli w klasach trzecich w gimnazjum oraz danych dotyczących uczniów, w tym ich wyników z części matematycznej i polonistycznej egzaminu gimnazjalnego z 2012 r.

Przegląd wyników dotychczasowych badań

Barbara Nye, Spyros Konstantopoulos i Larry Hedges (2004) podsumowali wyniki 15 oszacowań efektu nauczyciela, które były raportowane w pięciu badaniach (Armour, 1976; Goldhaber i Brewer, 1997; Hanushek, 1971; 1992; Murnane i Phillips, 1981). Autorzy raportują, że od 7% do 21% wariacji osiągnięć edukacyjnych uczniów

wyjaśnia różna efektywność nauczycieli. Wynik ten koresponduje z efektem $d = 0,32$ (Cohen, 1988)¹, co oznacza, że wraz ze wzrostem o jedno odchylenie standardowe miary efektywności nauczyciela, średnio o około jedną trzecią odchylenia standardowego wzrastają osiągnięcia szkolne uczniów (Hattie, 2009).

Brian Rowan, Richard Correnti i Robert J. Miller (2002), którzy zastosowali modelowanie hierarchiczne, raportowali od 18% do 28% (matematyka) oraz od 12% do 23% (czytanie) wyjaśnionej wariacji między oddziało-
waniem szkolnymi w modelu pustym z losową stałą (zob. Bryk i Raudenbush, 1992) oraz odpowiednio: między 8% i 18% (matematyka) i między 4% i 16% (czytanie) dla modelu uwzględniającego wcześniejsze osiągnięcia ucznia, status społeczno-ekonomiczny rodziny ucznia, skład społeczny szkoły.

Niektórzy badacze (np. Scheerens i Bosker, 1997) interpretują wariację międzyoddziało-
waniem wyników uczniów w modelu pustym jako efekt nauczyciela. Podejście to ma swoje uzasadnienie, ponieważ efektywność pracy nauczycieli jest najsilniejszym predyktorem osiągnięć uczniów na poziomie oddziału. Obok efektu nauczyciela, na poziomie oddziałów rozważa się także efekt rówieśników i efekt liczebności oddziału. Jednak stymulują one poprawę osiągnięć uczniów dopiero przy udziale nauczyciela. Efekt rówieśników stymulujący procesy nauczania może zaistnieć, gdy nauczyciele modyfikują swoje praktyki dydaktyczne, uwzględniając charakterystyki i potencjał grupy uczniów (Wilkinson, 2002). W przypadku mało licznych oddziałów należy spodziewać się braku lub niewielkiej poprawy wyników uczniów, o ile nauczyciel nie zmienia swoich metod nauczania i nie wykorzysta potencjału małej klasy do odpowiedniego

¹ Efekt d Cohena definiuje się jako różnicę między średnią wartością zmiennej zależnej w grupie eksponowanej na działanie bodźca (np. efektywny nauczyciel) i grupie kontrolnej, podzieloną przez odchylenie standardowe, tj. $d = \frac{\bar{x}_1 - \bar{x}_2}{s}$.

dostosowania programu i sposobów nauczania (Robinson, 1990).

Na wartość wariancji międzyoddziałowej składają się czynniki nie tylko na poziomie klasy, ale przede wszystkim wcześniejsze osiągnięcia uczniów i cechy ich środowiska rodzinnego. Dlatego inni badacze (np. Rowan, Correnti i Miller, 2002) jako efekt nauczyciela definiują wariancję międzyoddziałową w modelu, który kontroluje co najmniej status społeczno-ekonomiczny rodziny ucznia i wcześniejsze osiągnięcia edukacyjne.

Modelowanie hierarchiczne wykorzystali w swoich analizach Nye, Konstantopoulos i Hedges (2004). Dane pochodziły z eksperymentu *The Tennessee Class Size Experiment*, znanego bardziej jako projekt STAR (*Student-Teacher Achievement Ratio*). W swoich analizach autorzy uwzględnili na poziomie indywidualnym wcześniejsze wyniki uczniów, płeć, status społeczno-ekonomiczny rodziny, przynależność do grupy etnicznej. Na poziomie oddziałów uwzględniono liczebność klasy, pełnoetatową pomoc nauczycielską, doświadczenie nauczyciela i odbytą edukację. Wariancja międzyoddziałowa w pełnym modelu z losowymi stałymi wyniosła 13% (matematyka) i 7% (czytanie). W modelu pustym odpowiednio: 14% i 9%.

Jeszcze silniejszych dowodów na istotność efektu nauczyciela dostarczają badania uwzględniające czynnik genetyczny. Są to badania na bliźniętach. Czynnik genetyczny wyjaśnia do 65% osiągnięć szkolnych (Oliver i in., 2004). Uwzględniając czynnik genetyczny, efekt nauczyciela szacuje się na 8% (Byrne i in., 2010). Jak wykazała Jeanette Taylor i współpracownicy (2010), jakość pracy nauczyciela moderuje efekt genetycznego potencjału poznawczego ucznia. Nauczyciele, którzy są w stanie zapewnić optymalne środowisko dla realizacji naturalnych trajektorii rozwojowych swoich uczniów, najbardziej przyczyniają się do ich sukcesów edukacyjnych. Przegląd literatury dotyczącej wpływu czynnika genetycznego

na wyniki uczniów można znaleźć w pracy Artura Pokropka (2013).

Nie ma wątpliwości, że nauczyciele, różniąc się efektywnością (jakością pracy), wpływają na wyniki uczniów. Nie jest natomiast jasne, jakie konkretne cechy nauczycieli i stylu ich pracy, pozwalają wyjaśnić zróżnicowanie w efektywności ich pracy mierzonej osiągnięciami uczniów. Ten stan rzeczy ma swoje odzwierciedlenie w parafrazie popularnego twierdzenia „teacher matters” („nauczyciel ma znaczenie”), dokonanego przez Johna Hattie (2009, s. 108) – „variance due to teachers makes the difference” („wariancja w wynikach powodowana osobą nauczyciela ma znaczenie”). Z kolei, jak zauważają Nye, Konstantopoulos i Hedges (2004, s. 237), „powszechnie akceptuje się fakt, że nauczyciele różnią się efektywnością, jednak dostępne dowody empiryczne są słabe”. Steven Rivkin, Eric Hanushek i John Kain (2005, s. 419) twierdzą, że:

mała liczba empirycznych dowodów dotyczących roli nauczyciela w determinowaniu osiągnięć szkolnych oraz przyszłego sukcesu akademickiego i sukcesu na rynku pracy, może prowadzić do wniosku, że rodzice i uczniowie przeceniają rolę nauczyciela lub też, że mierzalne charakterystyki nauczyciela wyjaśniają niewiele rzeczywistego zróżnicowania (*true variation*) w efektywności nauczycieli [tłum. M. K.].

Badacze spierają się o możliwość wyjaśnienia zróżnicowania efektywności nauczycieli przez ich charakterystyki, takie jak wykształcenie, doświadczenie, wynagrodzenie itp. Spory te zaznaczyły się wyraźnie w debacie między Erikiem Hanushkiem (1986; 1989; 1996; 1997) a Larrym Hedgesem, Richardem Laine'em i Robem Greenwaldem (1994a; 1994b; 1996). Hanushek stoi na stanowisku, że metaanaliza dostępnych badań wskazuje na brak silnego pozytywnego związku charakterystyk nauczycieli z wynikami uczniów, przynajmniej gdy kontrolowany jest status społeczno-ekonomiczny rodziny ucznia. Greenwald, Hedges i Laine uważają, że analiza

tych samych danych prowadzi do wniosku o istnieniu pozytywnego związku między tymi zmiennymi, a jego siła ma praktyczne znaczenie dla twórców polityk oświatowych. Mimo braku zgody na temat znaczenia poszczególnych charakterystyk nauczycieli i stylów ich pracy w wyjaśnianiu zmienności osiągnięć edukacyjnych, należy przytoczyć dowody empiryczne na rzecz alternatywnych stanowisk.

Znajomość nauczanego przedmiotu. W podsumowaniu wyników 30 badań dotyczących związku poziomu wiedzy przedmiotowej nauczyciela z wynikami uczniów, 17 wykazało związek pozytywny, a 13 brak związku (Byrne, 1983). Jednakże w badaniach relacjonujących brak związku odnotowano tak małą zmienność w wiedzy przedmiotowej nauczycieli, że nieistotne statystycznie wyniki były nie do uniknięcia. Patricia Ashton i Linda Crocker (1987) spośród 14 relacjonowanych badań wskazały 5, które orzekały o pozytywnym związku wiedzy przedmiotowej z efektywnością nauczycieli, mierzoną jako średnie wyniki uczniów. Wyniki nauczycieli na testach kompetencji językowych (*verbal ability*), oceniających umiejętności czytania, wiedzę gramatyczną, zakres słownictwa, okazują się słabo korelować z wynikami na testach językowych uczniów (Bowles i Levin, 1968; 1992; Ehrenberg i Brewer, 1995; Ferguson i Ladd, 1996; Hanushek, 1971). Silniejsze związki prezentują wyniki egzaminów przedmiotowych, którym poddawani są nauczyciele, z wynikami uczniów w zakresie matematyki (Ferguson i Ladd, 1996; Rowan, Chiang i Miller, 1997). Nye, Konstantopoulos i Hedges (2004) oszacowali, że wraz ze wzrostem o jedną jednostkę miary wykształcenia nauczyciela, uczniowie otrzymują średnio o 0,037 (matematyka) oraz 0,013 (kompetencje językowe) odchylenia standardowego miary osiągnięć szkolnych więcej.

Andrew Wayne i Peter Youngs (2003) dokonali przeglądu badań dotyczących związku charakterystyk nauczyciela z przyrostami

osiągnięć szkolnych uczniów. Na podstawie publikacji objętych analizą, autorzy wykazali wpływ posiadania przez nauczyciela wyższego wykształcenia na wyniki testów kompetencji matematycznych. Nie ma pewnych dowodów tego związku w przypadku przedmiotów przyrodniczych, humanistycznych i języków obcych. Badania Lindy Darling-Hammond (2000) dowodzą silnego pozytywnego wpływu wysokich kwalifikacji nauczycieli na wyniki uczniów. Autorka wnioskuje, że zmienna „odsetek wysoko wykwalifikowanych nauczycieli w szkole” jest najsilniejszym predyktorem wyników uczniów spośród zmiennych charakteryzujących nauczycieli. Pozostałe zmienne na poziomie szkoły, tj. stosunek liczby uczniów do nauczycieli (*pupil-teacher ratio*), liczebność klas, proporcja nauczycieli w personelu szkoły, w analizach Darling-Hammond (2000) pokazują pozytywny, jednak bardzo mały lub nieistotny związek z wynikami uczniów. Daniel Goldhaber i Dominic Brewer (1997; 2000), a także Brian Rowan, Fang-Shen Chiang i Robert J. Miller (1997) wykazali brak związku wyników uczniów i poziomu wykształcenia nauczycieli. Autorzy raportowali związek wyników uczniów i dziedziny wykształcenia nauczyciela. Tylko w przypadku matematyki efekt okazał się istotny. Uczniowie, których nauczyciele posiadają tytuł magistra z matematyki, osiągnęli lepsze wyniki na testach z matematyki.

Wykształcenie pedagogiczne. Carolyn Evertson, Willis Hawley, Marilyn Zlotnik (1985) wykazali stały pozytywny efekt formalnego wykształcenia pedagogicznego nauczycieli na wyniki uczniów. Podobnie jak w przypadku poziomu wiedzy nauczanego przedmiotu, i tu występuje problem z małą wariancją, co uniemożliwia zaobserwowanie istotnego efektu. Spośród 13 badań, 11 raportowało wyższą efektywność nauczycieli posiadających pełne uprawnienia zawodowe nad nauczycielami bez dyplomu lub posiadającymi niepełne uprawnienia zawodowe. Umiejętności

dydaktyczne są powiązane z wiedzą przedmiotową i korelują z wynikami uczniów. Wielu autorów wykazało także, że lepsze wyniki osiągają wychowankowie nauczycieli, którzy uczestniczą w programach doskonalenia zawodowego z zakresu dydaktyki przedmiotowej (Wiley i Yoon, 1995; Brown, Smith i Stein, 1996).

Staż zawodowy. Zauważono dodatni, jednak nie zawsze istotny i raczej nieliniowy, związek stażu zawodowego nauczycieli z wynikami uczniów (Klitgaard i Hall, 1975; Murnane i Phillips, 1981). Początkujący nauczyciele (poniżej 3 lat doświadczenia zawodowego) są mniej efektywni niż nauczyciele z dłuższym stażem. Wyraźnie lepsze wyniki osiągają uczniowie nauczani przez nauczycieli z ponad pięcioletnim stażem (Rosenholtz, 1986). Efektywność nauczania obniża się u nauczycieli zbliżających się do wieku emerytalnego. Nye, Konstantopoulos i Hedges (2004) oszacowali efekt stażu nauczania na 0,085 (matematyka) oraz 0,094 (kompetencje językowe) odchylenia standardowego wyników uczniów.

Klimat nauczania. Dowiedziono istotnego związku klimatu nauczania z wynikami uczniów (Freiberg, 1999; Anderson, 1982). Valerie Lee i Anthony Bryk (1989) wykazali, że klimat nauczania ukierunkowany na osiągnięcia, jest istotnym czynnikiem wyjaśniającym zróżnicowanie wyników uczniów. Gdy przywództwo nauczycieli bardziej nastawione jest na korygowanie zachowań niepożądanych oraz wskazywanie rozwiązań poprawnych, efektywność nauczania staje się wyższa (Brown, 2001). Z kolei Brian Mullen i Carolyn Copper (1994) zwrócili uwagę na spójność grupy nauczycieli jako czynnika pozytywnie oddziałującego na osiągnięcia szkolne uczniów. Ronald Edmonds (1979) obok bezpiecznego i stabilnego klimatu szkoły wskazuje na pozytywny związek z wynikami uczniów takich czynników, jak: silne przywództwo dyrektora, zgodność co do celów i metod nauczania wśród kadry nauczycielskiej, wy-

sokie wymagania wobec uczniów, wykorzystywanie wyników ucznia do ewaluacji pracy nauczyciela.

Ruth Wade (1984) w przeglądzie studiów nad efektywnością nauczyciela wskazała na istotne czynniki efektywności pracy nauczyciela, takie jak: ustalanie celów kształcenia, długość kursów, charakterystyki nauczanej grupy, lokalizacja szkoły, harmonogram zajęć, motywacja do pracy, struktura i techniki nauczania. Wilbur Brookover (1982), w książce podsumowującej kilkadziesiąt badań nad efektywnością pracy szkół, dodaje jeszcze do tej listy: nauczanie grupowe, efektywne strategie nauczania, efektywne zarządzanie klasą, stosowanie wzmocnień pozytywnych uczniów, zaangażowanie się rodziców w życie szkoły.

W metaanalizie ponad 100 studiów eksperymentalnych nad pracą nauczyciela, Robert Marzano, Jana Marzano, Debra Pickering (2003) wskazali, że istotne dla osiągnięcia sukcesu edukacyjnego mogą być częste podsumowania i nakłanianie uczniów do robienia notatek, wzmocnienia pozytywne uczniów i wyrażanie oznak uznania, gdy na to zasługują, zadawanie prac domowych i praca indywidualna na lekcjach, wspólne rozwiązywanie problemów, precyzyjne definiowanie celów i zapewnianie wsparcia w ich realizacji, stawianie hipotez i testowanie ich, aktywowanie u uczniów posiadanej już wiedzy.

Hattie (2009) podsumował wyniki 31 metaanaliz dotyczących wpływu nauczyciela na wyniki uczniów. W Tabeli 1 przedstawiono wyniki jego pracy. Oprócz metaanalizy dotyczącej efektu nauczyciela w ogóle, zaprezentowano zbiorcze wyniki metaanaliz poświęconych konkretnym charakterystykom nauczycieli i stylom ich pracy. Hattie do raportowania efektów wykorzystał miarę d Cohena (Cohen, 1988). W badaniach edukacyjnych efekt $d = 0,2$ należy interpretować jako słaby; $d = 0,4$ jako średni; a $d = 0,6$ jako silny efekt (Hattie, 2009, s. 9). Do największych przyrostów wyników

uczniów przyczyniają się nauczyciele, którzy brali udział w szkoleniach z zakresu prowadzenia lekcji wykorzystujących wideoanalizę ($d = 0,88$). Wyniki uczniów podnoszą nauczyciele, którzy formułują zrozumiałe i precyzyjne komunikaty ($d = 0,75$). Zdolności językowe nauczycieli pozwalają im na komunikowanie idei w przejrzysty i przekonujący sposób (Murnane, 1985). Efektywni są także ci, którzy mają dobre relacje ze swoimi uczniami ($d = 0,72$). Spośród pozostałych zmiennych, które mierzą relacje nauczycieli i uczniów, największy wpływ na wyniki uczniów mają: promowanie samodzielności, empatia, przyjazność, zachęcanie do myślenia analitycznego, krytycznej oceny i myślenia kreatywnego, zachęcanie do nauki, indywidualizacja nauczania. W przypadku wszystkich tych zmiennych $d > 0,4$.

Podsumowując omówione metaanalizy i przeglądy badań, wskazać można pozytywny wpływ na wyniki uczniów: udziału nauczycieli w szkoleniach z zakresu

prowadzenia lekcji; formułowania przez nauczyciela zrozumiałych i precyzyjnych komunikatów do uczniów; wiedzy z nauczanego przedmiotu, zwłaszcza w przypadku matematyki; formalnego wykształcenia pedagogicznego; stażu; spójności grupy nauczycieli; zgodności kadry nauczycielskiej co do celów i metod nauczania; wysokich wymagań wobec uczniów; wykorzystywania wyników uczniów do ewaluacji pracy.

Szacowanie efektu nauczyciela

Analizowane dane dotyczą uczniów i nauczycieli gimnazjów i zostały zgromadzone w ramach „Badania uwarunkowań osiągnięć szkolnych uczniów szkół podstawowych i gimnazjów”, realizowanego przez zespół Edukacyjnej Wartości Dodanej (EWD) od stycznia 2010 do czerwca 2012 (01–04.2010 – 1. etap; 09–10.2011 – 2. etap; 05–06.2012 – 3. etap). Badanie objęło 5249 uczniów, w 292 klasach, w 150 szkołach oraz 3579

Tabela 1

Podsumowanie informacji z metaanaliz dotyczących efektu nauczyciela na osiągnięcia edukacyjne uczniów (Hattie, 2009, s. 109)

Czynniki nauczycielskie o największym wpływie na wyniki uczniów	Liczba metaanaliz	Liczba badań	Liczba oszacowań efektów	d	SE
Szkolenia nauczycieli wykorzystujące wideoanalizę	4	402	439	0,88	–
Klarowność komunikacji	1	–	–	0,75	–
Relacje nauczyciela z uczniami	1	229	1 450	0,72	(0,011)
Doskonalenie zawodowe	5	537	1 884	0,62	(0,034)
Nieetykietowanie uczniów	1	79	79	0,61	–
Jakość nauczania oceniana przez uczniów	5	141	195	0,44	(0,060)
Formułowanie wymagań wobec uczniów	8	674	784	0,43	(0,081)
Przygotowanie pedagogiczne	3	53	286	0,11	(0,044)
Wiedza przedmiotowa	2	92	424	0,09	(0,016)
Efekt nauczyciela (ogólnie)	1	18	18	0,32	(0,020)
Razem	31	2 225	5 559	0,49	(0,049)

nauczycieli². Z operatu pozyskanego z Systemu Informacji Oświatowej (SIO) wykluczono szkoły charakteryzujące się odmiennym programem nauczania lub strukturą nauczania, szkoły prowadzone przez jednostki inne niż samorządowe, a także szkoły publiczne z tylko jednym oddziałem, mniejszym niż 10 osób. Zastosowano warstwowy, proporcjonalny, dwustopniowy schemat losowania. W ramach szkół losowane były oddziały. W badaniu brali udział wszyscy uczniowie z wylosowanego oddziału, którzy w momencie badania byli obecni w szkole. Próba została podzielona na 6 warstw, według kategorii wielkości miejscowości oraz informacji dotyczącej tego, czy szkoła jest jednooddziałowa, czy wielooddziałowa.

Wykorzystane w badaniu narzędzia zostały opracowane w przez zespół EWD samodzielnie lub na podstawie narzędzi opracowanych w ramach polskich i międzynarodowych projektów badawczych. Prezentowane tu wyniki opierają się na wskaźnikach i pytaniach zaczerpniętych z badań PISA OECD w zakresie wskaźnika „zamożność gospodarstwa domowego” (HOMEPOS) oraz skali „autorytet nauczyciela/utrzymanie dyscypliny”. Do pomiaru inteligencji uczniów wykorzystano test matrycy Ravena.

Do uczniów biorących udział w 3. etapie badania przypisano nauczycieli, którzy uczyli poszczególne oddziały przez co najmniej rok do momentu badania. Okres ten przypadał na ostatni rok nauki w gimnazjum. Przypisanie nauczycieli do klas zrealizowano osobno dla nauczycieli matematyki i języka polskiego. Oddziały, które przez rok poprzedzający termin badania miały więcej niż jednego nauczyciela z matematyki lub polskiego, zostały usunięte z analiz.

Analizy prowadzono na dwóch zbiorach danych. Wyniki uczniów z części matematycznej egzaminu były zmienną zależną w przypadku pierwszego zbioru. Baza zawierała dane dotyczące 3883 uczniów w 246 klasach w 137 szkołach, nauczanych przez 202 nauczycieli matematyki. Wyniki z części polonistycznej egzaminu były zmienną zależną w przypadku drugiego zbioru danych. Baza zawierała dane dotyczące 4119 uczniów w 260 klasach w 143 szkołach, nauczanych przez 215 nauczycieli języka polskiego. Nie przeprowadzono imputacji braków danych. Pozostawiono je w bazie. Do wyliczania wartości skal stosowano procedurę wyłączania braków parami. Wszystkie analizy prowadzono z użyciem wag osobno dla poziomu uczniów, oddziałów, szkół. Na potrzeby raportowania właściwości skal, przeprowadzono analizy bez ważenia.

Wyniki analiz należy raportować z uwzględnieniem zagrożeń dla trafności interpretacji, wynikających zarówno z zastosowanego schematu badania, jak i przyjętej metody analizy danych. Wyniki mogą być obciążone, ponieważ: (a) rodzice mogą wybierać miejsce zamieszkania, a przez to sieć dostępnych dla swoich dzieci szkół; (b) rodzice mogą wybierać konkretną szkołę dla swoich dzieci, inną niż ta, do której są one przypisane ze względu na miejsce zamieszkania; (c) dyrektor lub administracja szkoły może przydzielać uczniów do konkretnych oddziałów i nauczycieli na podstawie jawnych lub ukrytych reguł lub sugestii rodziców. Wyniki badania mogą być więc obciążone przez autoselekcję (rodzice wybierają szkoły dla swoich uczniów) i selekcję (jawne i ukryte mechanizmy rekrutacji) uczniów do szkół i oddziałów (Dolata, 2008; Jordan, Mendro i Weerasinghe, 1997; Sanders i Rivers, 1996).

Aby zredukować ewentualne obciążenie wyników wynikające z uznaniowego (nielosowego) przypisania uczniów do nauczycieli o różnym poziomie efektywności, uwzględniono w analizach charakterystyki uczniów

² Pełna dokumentacja badania jest dostępna w materiałach do pobrania na stronach: ewd.edu.pl, ibe.edu.pl, cke.edu.pl.

i ich środowiska rodzinnego. Uwzględniono także średnie wcześniejszych osiągnięć uczniów w oddziałach, kontrolując tym samym ewentualną segregację uczniów w klasach według wcześniejszych osiągnięć. Uwzględniono także średnie statusu społeczno-ekonomicznego uczniów w oddziałach, kontrolując tym samym skład społeczny klasy (tj. ewentualną segregację uczniów w klasach ze względu na ich statusu społeczno-ekonomicznego).

Analizy wykonano z wykorzystaniem modelowania hierarchicznego. Modele hierarchiczne to modele regresji, w których zmienne egzogenne występują na kilku poziomach najczęściej na dwóch (np. uczniowie, szkoły) lub trzech (np. uczniowie, oddziały, szkoły). Spośród obszernej literatury na temat modeli wielopoziomowych na uwagę zasługuje przede wszystkim książka Stephena Raudenbusha i Anthony'ego Bryka (2002). Obszerne informacje zawiera także dokumentacja programu HLM7, z pomocą którego wykonano analizy, dostępna na stronie ssicentral.com. Na uwagę polskiego czytelnika zasługują w zakresie modeli hierarchicznych publikacje Henryka Domańskiego i Artura Pokropka (2011) oraz Domańskiego i in. (2012).

Wyniki dla części matematycznej

W Tabeli 2 zaprezentowano współczynniki i błędy standardowe dla modeli uwzględniających zmienne niezależne na poziomie uczniów, oddziałów i szkół. Nie są raportowane wartości statystyki p . Zainteresowany czytelnik może je obliczyć samodzielnie, dzieląc współczynnik przy danej zmiennej przez błąd standardowy, a następnie odczytując wartość krytyczną z rozkładu normalnego Z lub wykorzystać dostępne kalkulatory.

Na podstawie stałej w modelu pustym możemy przewidzieć wynik ucznia z części matematycznej egzaminu. Uczeń średnio otrzymał 0,06 odchylenia standardowego miary wyników. Wyniki surowe zostały wyskalowane za pomocą trójparametrycznego modelu regresji logistycznej (3PLM)

(Jakubowski i Pokropek, 2009). Wyniki raportowane są na skali o średniej 0 i odchyleniu standardowym 1. Średnia w próbie różni się od stałej w modelu ze względu na braki danych.

Współczynnik korelacji wewnątrzgrupowej można interpretować jako procent wyjaśnionej wariancji wyników egzaminu. W modelu pustym, uwzględniającym jedynie fakt grupowania uczniów w oddziałach i szkołach, współczynnik korelacji wewnątrzgrupowej wyniósł 0,121. Oznacza to, że grupowanie uczniów w klasach wyjaśniło 12,1% wariancji wyników z części matematycznej egzaminu. Z kolei grupowanie uczniów w szkołach wyjaśniło 9,1%, wariancji wyników części matematycznej egzaminu.

Do modelu pustego dodano charakterystyki uczniów i ich rodzin, czyli wcześniejsze osiągnięcia szkolne (wynik sprawdzianu po szkole podstawowej), płeć, trwałą dysleksję (orzeczenie dysleksji ucznia na sprawdzianie i egzaminie), inteligencję (mierzoną testem matryc Ravena), wykształcenie rodziców (mierzone wyższą liczbą lat nauki rodziców), liczbę osób w gospodarstwie domowym, zamożność gospodarstwa domowego. Otrzymany model EWD³ szkół (1MAT)⁴ wyjaśnił

³ Zaproponowany model EWD szkół różni się od modeli EWD stosowanych w Polsce. Wskaźniki EWD szkół w wersji polskiej tworzone są z modeli dwupoziomowych, które uwzględniają wynik ucznia na egzaminie zewnętrznym na koniec wcześniejszego etapu kształcenia, płeć i dysleksję. W przypadku kalkulatorów EWD są to modele jednopoziomowe. Wyczerpujące materiały na temat stosowanych w Polsce wskaźników EWD dostępne są na stronie ewd.edu.pl oraz w materiałach pokonferencyjnych Polskiego Towarzystwa Diagnostyki Edukacyjnej (PTDE) (m.in. Pokropek, 2009; Pokropek i Żółtak, 2012). Polskie wskaźniki EWD należy interpretować jako wskaźniki efektywności pracy szkół. Dodając poziom oddziału, tj. uwzględniając fakt grupowania uczniów w oddziałach, a także zmienne na poziomie szkół, można mówić o wskaźnikach efektywności pracy nauczycieli.

⁴ Symbole w nawiasach w tekście odpowiadają symbolom w nagłówkach Tabeli 2 i Tabeli 4, co ma ułatwiać identyfikację opisywanych modeli.

58% wariancji wyników części matematycznej egzaminu na poziomie ucznia.

Reszty z modelu EWD (1MAT) są miarą efektywności pracy szkoły w nauczaniu matematyki. Na efektywność pracy szkoły składa się m.in. praca nauczycieli, kultura organizacyjna szkoły, klimat pracy, infrastruktura. Aby uzyskać miarę efektywności pracy nauczyciela, należy „wytrącić” efekty inne niż nauczycielskie. W tym celu do modelu EWD (1MAT) dodawano kolejno zmienne na poziomie szkoły, a także gminy, w której szkoła się znajduje. Zweryfikowano ponad 40 potencjalnych zmiennych wyjaśniających wyniki szkolne uczniów, m.in.: liczbę nauczycieli, strukturę ich wieku, stażu i doświadczenia, średnią liczebność klas, liczbę mieszkańców gminy, strukturę gospodarki gminy, wysokość wydatków gminy w ogóle i na oświatę na mieszkańca, liczbę bibliotek na mieszkańca, wielkość księgozbioru na mieszkańca, stopę bezrobocia, odsetek mieszkańców korzystających z pomocy społecznej. Niestety, autor nie dysponował wyczerpującymi danymi na poziomie szkół, takimi jak wysokość finansowania działalności szkół, infrastruktura szkół itp. Problem z dostępnością dobrej jakości danych opisujących funkcjonowanie szkół jest związany z brakiem efektywnego systemu gromadzenia i monitoringu kompleksowego zestawu charakterystyk szkół. Dostępne i dobrej jakości są natomiast dane charakteryzujące gminy.

Spośród ponad 40 testowanych zmiennych, istotną zmienną wyjaśniającą wyniki części matematycznej egzaminu na poziomie szkoły była lokalizacja gimnazjum, mierzona na czterostopniowej skali porządkowej (wieś, miasto do 20 tys. mieszkańców, miasto od 20 tys. do 100 tys., miasto powyżej 100 tys.). Korelacja lokalizacji gimnazjum ze średnim wynikiem części matematycznej egzaminu w gminie wyniosła 0,393. Przy kontroli zmiennych na poziomie ucznia, tj. wcześniejszych osiągnięć edukacyjnych, inteligencji, wykształcenia rodziców

itd. efekt lokalizacji na wyniki uczniów był negatywny. Im większe miasto, tym niższe wyniki części matematycznej egzaminu przy kontroli wcześniejszych osiągnięć, inteligencji i zmiennych statusowych.

Innym istotnym predyktorem wyników części matematycznej egzaminu na poziomie szkoły była dynamika bezrobocia, mierzona jako średnia zmiana rok do roku udziału bezrobotnych w liczbie osób w wieku produkcyjnym w gminie, w której znajduje się gimnazjum. Im większe wzrosty stopy bezrobocia, tym gorsze wyniki osiągalni gimnazjaliści.

Po dodatkowym uwzględnieniu w modelu EWD szkół (1MAT) lokalizacji gminy, dynamiki bezrobocia, a także średniego wyniku sprawdzianu uczniów w klasie, reszty uzyskane z modelu 2MAT, można było uznać w przybliżeniu za trafny wskaźnik efektywności nauczyciela. Średni wynik sprawdzianu uczniów w klasie został uwzględniony, aby kontrolować segregację uczniów w klasach ze względu na wcześniejsze osiągnięcia edukacyjne. Średni poziom zmiennych statusowych w klasie okazał się być nieistotny. Nie było więc potrzeby wprowadzać ich do modelu.

Należy zauważyć, że w modelu EWD nauczyciela (2MAT) odseparowane zostały tylko zmienne uwzględnione w badaniu. Poza kontrolą były zmienne nieobserwowalne lub obserwowalne, ale w badaniu niezmiernie. Oczywiście, pełne odseparowanie czynników innych niż nauczycielskie, można by osiągnąć jedynie w kontrolowanym eksperymencie. Można więc argumentować, że uzyskany wskaźnik dotyczy nie tyle efektywności nauczyciela, ile cech oddziały. Można to sprawdzić.

Jeśli model EWD (2MAT) dla części matematycznej rozszerzyć o wskaźnik EWD nauczycieli języka polskiego (użyty z analogicznego modelu 2POL), a model EWD (2POL) dla części polonistycznej rozszerzyć o wskaźnik EWD nauczycieli matematyki, to w modelach

Tabela 2

Uwarunkowania osiągnięć matematycznych. Wyniki trzydziestomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

Zmienna zależna: wynik egzaminu gimnazjalnego w części matematycznej [standaryzacja Z] (ZIRTMAT) ^(a)									
	Pusty	(SE)	(1MAT)	(SE)	(2MAT)	(SE)	(3MAT)	(SE)	(4MAT)
Oszacowanie efektów stałych									
POZIOM UCZNIĄ									
Sprawdzian (std. Z) (ZSP3PL)			0,55*	0,018	0,55*	0,019	0,55*	0,019	0,55*
Plec ^(b) (EWD_PLEC)			-0,12*	0,019	-0,12*	0,019	-0,12*	0,019	-0,13*
Dysleksja na sprawdzianie i na egz. gim. ^(c) (DYS)			-0,18*	0,043	-0,18*	0,043	-0,18*	0,043	-0,18*
Inteligencja (Test matryc Ravena) (std. Z) (RAVEN)			0,30*	0,017	0,30*	0,017	0,30*	0,017	0,29*
Wyższa z liczby lat nauki rodziców (HEDULL)			0,02*	0,005	0,02*	0,005	0,02*	0,005	0,02*
Zamożność gospodarstwa domowego (std. Z) (HOMEPOS)			0,04*	0,016	0,03*	0,016	0,04*	0,016	0,03*
Liczba osób w gospodarstwie domowym (WIELRODZ)			0,02*	0,007	0,02*	0,007	0,02*	0,007	0,02*
Stała	0,06	(0,039)	-0,11	0,077	-0,12	0,080	-0,12	0,080	-0,14
POZIOM ODDZIAŁU									
Skala „aktywny styl nauczania” (std. Z) (ACTIVM5)							0,10*	0,032	
Skala „autoritet nauczyciela/utrzymanie dyscypliny” (std. Z) (DYSM)									0,13*
Średni wynik sprawdzianu uczniów w klasie (std. Z) (ZSPRAWD2)					0,14*	0,041	0,17*	0,040	0,13*
POZIOM SZKOŁY									
Lokalizacja gimnazjum (LOK)					-0,05*	0,018	-0,05*	0,017	-0,04*
Średnia dynamika bezrobocia YoY w % (2006–2010) (YOYBEZ)					-0,86*	0,179	-0,81*	0,166	-0,78*

<i>Oszacowanie efektów losowych</i>						
Wariancja efektów szkół	0,09	0,03	0,02	0,02	0,02	0,02
Wariancja efektów klas (w naw. % wyjaśnionej wariancji)	0,12	0,02	0,02	(86%)	0,02	(91%)
Wariancja na poziomie ucznia (w naw. % wyjaśnionej wariancji)	0,77	0,32	(58%)	0,32	0,32	0,32
<i>Współczynniki korelacji wewnątrzgrupowej^(d)</i>						
Poziom szkół	0,091	0,074	0,061	0,054	0,057	
Poziom oddziałów	0,121	0,053	0,045	0,045	0,031	
Poziom szkół i oddziałów łącznie	0,211	0,127	0,105	0,099	0,088	
<i>Podsumowanie</i>						
Deviance	10 401,07	6 920,89	6 888,52	6 880,08	6 858,71	
Liczba szacowanych parametrów	4	11	14	15	15	
Porównanie dopasowania modelu do modelu pustego	–	33,46%	33,77%	33,85%	34,06%	
Poprawa dopasowania przed dodanie predyktorów	–	33,46%	0,47%	0,12%	0,43%	

Liczba uczniów: $n = 3883$, liczba oddziałów: $n = 246$, liczba szkół: $n = 137$ (analizy na danych ważonych).

* Korelacja jest istotna na poziomie 0,05.

^(a) W nawiasach podano nazwy zmiennych, dla ułatwienia odczytania równań w aneksie.

^(b) Grupa odniesienia: chłopcy.

^(c) Grupa odniesienia: uczniowie bez dysleksji.

^(d) W przypadku modeli z predyktorami dla współczynników korelacji wewnątrzgrupowej zastosowano taką samą interpretację jak dla modelu pustego. Wartości te to % wyjaśnionej wariancji wyników uczniów z wariancji pozostałej do wyjaśnienia na poziomie oddziałów, po uwzględnieniu predyktorów na poziomie ucznia.

takich kontrolowany będzie ewentualny wpływ na wyniki uczniów innych niż nauczyciel czynników na poziomie oddziału. Współczynnik dla EWD nauczycieli języka polskiego w modelu EWD dla części matematycznej (2MAT) wyniósł 0,34 ($SE = 0,14$) i jest istotny statystycznie ($p < 0,05$). Z kolei współczynnik dla EWD nauczycieli matematyki w modelu EWD dla części polonistycznej (2POL) wyniósł 0,26 ($SE = 0,17$) i jest nieistotny statystycznie ($p > 0,05$). Zarówno w pierwszym, jak i drugim przypadku zmienne te nie powinny mieć znaczenia. Można więc sformułować wniosek, że zaproponowany model EWD dla nauczycieli matematyki jest gorszy niż model EWD dla nauczycieli języka polskiego, ponieważ wskaźnik EWD dla nauczycieli matematyki jest tak na prawdę miarą zarówno efektywności nauczycieli, jak i innych charakterystyk oddziału, np. atmosfery w klasie, efektu rówieśników itp. W Tabeli 3 przedstawiono korelacje między zmiennymi resztowymi z modelu 2MAT oraz z modelu rozszerzonego o wskaźnik EWD nauczycieli matematyki a wybranymi potencjalnymi predyktorami efektywności nauczycieli.

Zgodnie z oczekiwaniami, po odseparowaniu od efektu nauczyciela czynników oddziału, wyższe korelacje obserwujemy dla charakterystyk nauczycieli i stylu ich pracy (staż, stopień awansu, skala „aktywny styl nauczania”), a niższe korelacje dla zmiennych związanych w całości (liczba uczniów) lub częściowo (wyniki egzaminu, skala „autorytet.../dyscyplina...”) z charakterystykami klasy.

Podjęto próbę odpowiedzi na pytanie, za pomocą jakich cech nauczycieli i charakterystyk ich pracy można wytłumaczyć zróżnicowanie w efektywności nauczania w poszczególnych oddziałach. Przetestowano 26 zmiennych niezależnych, w tym 14 charakterystyk nauczycieli, związanych ze stażem pracy, stopniem awansu zawodowego, doskonaleniem zawodowym nauczycieli. Wśród testowanych zmiennych znalazło się 7 skal skonstruowanych na podstawie pytań z kwestionariuszy nauczycielskich oraz 5 skal skonstruowanych na podstawie pytań z kwestionariuszy uczniowskich. Skale dotyczyły stylów pracy nauczycieli. Wyniki zaprezentowane zostały w Tabeli 2 (modele 3MAT i 4MAT).

Tabela 3

Korelacje zmiennej resztowej z modelu (2MAT) z wynikami egzaminu i wybranymi potencjalnymi predyktorami efektywności nauczyciela w nauczaniu matematyki

Wybrane predyktory efektywności nauczyciela	Efektywność nauczyciela	Efektywność nauczyciela po odseparowaniu czynników oddziału
Średni wynik egzaminu z matematyki (std. Z)	0,375*	0,359*
Wiek nauczyciela	0,100	0,094
Staż pracy w szkolnictwie	0,071	0,080
Stopień awansu zawodowego	0,067	0,157*
Aktywny styl nauczania (skala) (std. Z)	0,128*	0,144*
Autorytet nauczyciela/utrzymanie dyscypliny (skala) (std. Z)	0,323*	0,317*
Liczba uczniów w klasie	0,031	0,014

Liczba nauczycieli: $n = 246$ (analiza na danych nieważonych).

*Korelacja jest istotna na poziomie 0,05.

Ze względu na bardzo małą wariancję, płeć (90% kobiet w próbie) i wykształcenie (98% nauczycieli z wykształceniem co najmniej wyższym magisterskim w próbie) okazały się nieistotnymi predyktorami wyników części matematycznej egzaminu, podobnie jak wiek i staż nauczyciela, stopień awansu zawodowego, uczestnictwo w programach doskonalenia zawodowego, a także większość skal dotyczących stylów nauczania.

W istotny sposób efektywność nauczyciela opisuje skala „aktywny styl nauczania”. Skala ta sprawdza sposób prowadzenia zajęć w czterech wymiarach: „indywidualizacja nauczania”, „korzystanie z multimedii”, „podsumowania przerobionego materiału”, „stymulowanie pracy w grupach”. Dokumentacja techniczna skali znajduje się w aneksie. Nauczyciele, którzy podchodzą indywidualnie do uczniów, wykorzystują w procesie nauczania multimedia, wprowadzają podsumowania przerobionego materiału, organizują pracę w grupach, przyczyniają się do podniesienia wyników średnio o 0,10 odchylenia standardowego wyniku z części matematycznej egzaminu. Silniejszym predyktorem, który znosi istotność skali „aktywny styl nauczania”, jest skala „autorytet nauczyciela/utrzymanie dyscypliny”. Dokumentacja techniczna skali znajduje się w aneksie. Wraz ze wzrostem o 1 odchylenie standardowe na skali, gimnazjaliści otrzymują średnio o 0,13 odchylenia standardowego wyższy wynik z części matematycznej egzaminu. Skala „autorytet nauczyciela/utrzymanie dyscypliny” wyjaśnia 91% wariancji efektu nauczyciela w nauczaniu matematyki, przy kontroli wcześniejszych osiągnięć edukacyjnych, płci, dysleksji, inteligencji, zmiennych statusowych rodziny ucznia, lokalizacji gimnazjum i dynamiki bezrobocia w gminie, w której znajduje się gimnazjum. Choć, oczywiście, pozostałej do wyjaśniania wariancji na poziomie oddziałów, po uwzględnieniu predyktorów na poziomie

ucznia, nie pozostaje wiele. Podana wartość procentowa to tzw. miara pseudo R^2 (Raudenbush i Bryk, 2002)⁵.

Na podstawie statystyki *Deviance*, można ocenić jakość dopasowania modelu do danych. Im niższa statystyka *Deviance*, tym lepsze dopasowanie. Dodając do zmiennych na poziomie indywidualnym skalę „aktywny styl nauczania” i średni wynik sprawdzianu w klasie, dopasowanie poprawia się o 0,59% (model 3MAT). Uwzględniając skalę „autorytet nauczyciela/utrzymanie dyscypliny” zamiast skali „aktywny styl nauczania”, dopasowanie poprawia się o 0,90% (model 4MAT).

Wyniki dla części polonistycznej

W Tabeli 4 zaprezentowano współczynniki i błędy standardowe dla modeli z wynikami części polonistycznej egzaminu jako zmienną zależną. Uczeń średnio otrzymał 0,07 odchylenia standardowego miary wyników. Fakt grupowania uczniów w klasach wyjaśnił 8,2%, a fakt grupowania uczniów w szkołach 7,8% wariancji wyników części polonistycznej egzaminu.

Do modelu pustego dodano charakterystyki uczniów i ich rodzin, tj. wcześniejsze osiągnięcia edukacyjne, płeć, miarę inteligencji, wykształcenie rodziców, liczbę osób w gospodarstwie domowym, zamożność gospodarstwa domowego. Nieistotnym predyktorem wyników z części polonistycznej na poziomie uczniów jest trwała dysleksja. Otrzymany model EWD szkół (1POL) wyjaśnił 55% wariancji wyników części polonistycznej egzaminu na poziomie ucznia.

Istotnym predyktorem wyników części polonistycznej egzaminu na poziomie szkoły jest lokalizacja gimnazjum. Podobnie jak w przypadku matematyki, przy

⁵ Pseudo R^2 mierzy stopień redukcji niewyjaśnionej wariancji na danym poziomie analizy (uczniowie, oddziały, szkoły) w modelu pustym w stosunku do modelu ze zmiennymi wyjaśniającymi. Zaleca się jednak ostrożność w interpretacji tej miary (Kreft i de Leeuw, 1998).

Tabela 4
Uwarunkowania osiągnięć językowych. Wyniki trzypięciomowych analiz regresji, modele z losową statą, odporne (robust) oszacowania błędów standardowych

Zmienna zależna: wynik egzaminu gimnazjalnego w części polonistycznej (standaryzacja Z) (ZIRTPOL)									
	Pusty	(SE)	(1POL)	(SE)	(2POL)	(SE)	(3POL)	(SE)	(4POL)
Oszacowanie efektów stałych									
POZIOM UCZNIĄ									
Sprawdzian (std. Z) (ZSP3PL)			0,54*	0,02	0,54*	0,02	0,54*	0,02	0,54*
Płeć ^(a) (EWD_PLEC)			0,29*	0,02	0,29*	0,02	0,28*	0,02	0,28*
Inteligencja (Test matryc Ravena) (std. Z) (RAVEN)			0,17*	0,02	0,17*	0,02	0,17*	0,02	0,16*
Wyższa z liczby lat nauki rodziców (HEDULL)			0,03*	0,01	0,03*	0,01	0,03*	0,01	0,03*
Zamożność gospodarstwa domowego (std. Z) (HOMEPOS)			0,08*	0,02	0,08*	0,02	0,08*	0,02	0,08*
Liczba osób w gospodarstwie domowym (WIELRODZ)			-0,02*	0,01	-0,03*	0,01	-0,03*	0,01	-0,03*
Stała	0,07*	0,03	-0,31*	0,08	-0,23*	0,08	-0,23*	0,08	-0,23*
POZIOM ODDZIAŁU									
Skala „pomocny nauczyciel” (std. Z) (POMOCNP)							0,07*	0,03	
Skala „autorytet nauczyciela/utrzymanie dyscypliny” (std. Z) (DYSP)									0,11*
POZIOM SZKOŁY									
Lokalizacja gimnazjum (LOK)					-0,04*	0,02	-0,04*	0,02	-0,04*
Oszacowanie efektów losowych									
Wariancja efektów szkół	0,07		0,02		0,02		0,02		0,02
Wariancja efektów oddziałów (w naw. % wyj. war.)	0,08		0,02		0,02	(79%)	0,02	(80%)	0,01
Wariancja na poziomie ucznia (w naw. % wyj. war.)	0,78		0,35	(55%)	0,35		0,35		0,35

<i>Współczynniki korelacji wewnątrzgrupowej</i>						
Poziom szkół	0,078	0,057	0,051	0,051	0,043	
Poziom oddziałów	0,082	0,042	0,042	0,039	0,039	
Poziom szkół i oddziałów łącznie	0,160	0,100	0,093	0,090	0,082	
<i>Podsumowanie</i>						
Deviance	11 033,91	7 615,64	7 605,77	7 599,31	7 586,19	
Liczba szacowanych parametrów	4	10	11	12	12	
Porównanie dopasowania modelu do modelu pustego	–	30,98%	31,07%	31,13%	31,25%	
Poprawa dopasowania przed dodanie predyktorów	–	30,98%	0,13%	0,08%	0,26%	

Liczba uczniów: 4119, liczba oddziałów: 260, liczba szkół: 143.

*Korelacja jest istotna na poziomie 0,05.

^(a) Grupa odniesienia: chłopcy.

kontroli zmiennych na poziomie ucznia, tj. wcześniejszych osiągnięć edukacyjnych, inteligencji, wykształcenia rodziców itd., efekt lokalizacji na wyniki uczniów był negatywny. Im większe miasto, tym niższe wyniki części polonistycznej egzaminu, przy kontroli wcześniejszych osiągnięć, inteligencji i zmiennych statusowych.

Dynamika bezrobocia, będąca istotnym predyktorem wyników części matematycznej egzaminu, nie jest istotnym predyktorem wyników z części polonistycznej egzaminu. Nieistotne znaczenie ma średnia wyników sprawdzianu. Po uwzględnieniu w modelu EWD (1POL) dodatkowo lokalizacji gminy, reszty uzyskane z modelu (2POL), można uznać w przybliżeniu za trafny wskaźnik efektywności nauczyciela.

W Tabeli 5 przedstawiono korelacje między zmiennymi resztowymi z modelu 2POL oraz z modelu rozszerzonego o wskaźnik EWD nauczycieli języka polskiego a wybranymi potencjalnymi predyktorami efektywności nauczycieli.

Istotna i wysoka (0,485–0,524) korelacja efektywności nauczyciela z wynikami uczniów z części polonistycznej egzaminu

potwierdza hipotezę, że efektywny nauczyciel znacznie przyczynia się do sukcesów edukacyjnych swoich uczniów. Najlepszym predyktorem efektywności nauczyciela w nauczaniu języka polskiego jest – tak jak w nauczaniu matematyki – skala „autorytet nauczyciela/utrzymanie dyscypliny” (0,231–0,246).

Podobnie jak w przypadku nauczania matematyki, tak w przypadku nauczania języka polskiego, podjęto próbę ustalenia, za pomocą jakich cech nauczycieli i charakterystyk ich pracy można wytłumaczyć zróżnicowanie w efektywności nauczania w poszczególnych oddziałach. Wyniki prezentowane są w Tabeli 4 (modele 3POL i 4POL).

Ze względu na bardzo małą wariancję, płeć (97,5% kobiet w próbie) i wykształcenie (98% nauczycieli z wykształceniem co najmniej wyższym magisterskim w próbie) okazały się nieistotnymi predyktorami wyników części polonistycznej egzaminu. Nieistotnymi predyktorami wyników części polonistycznej egzaminu okazały się także wiek i staż nauczyciela, stopień awansu zawodowego, uczestnictwo w programach doskonalenia zawodowego.

W istotny sposób efektywność nauczyciela opisuje skala „pomocny nauczyciel”.

Tabela 5

Korelacje zmiennej resztowej z modelu (2POL) z wynikami egzaminu i wybranymi potencjalnymi predyktorami efektywności nauczyciela w nauczaniu języka polskiego

Wybrane predyktory efektywności nauczyciela	Efektywność nauczyciela	Efektywność nauczyciela po odseparowaniu czynników oddziału
Średni wynik egzaminu z polskiego (std. Z)	0,524*	0,485*
Wiek nauczyciela	0,050	-0,040
Staż pracy w szkolnictwie	-0,078	-0,084
Stopień awansu zawodowego	0,086	0,029
Pomocny nauczyciel (skala) (std. Z)	0,176*	0,192*
Autorytet nauczyciela /utrzymanie dyscypliny (skala) (std. Z)	0,231*	0,246*
Liczba uczniów w klasie	0,024	0,016

Liczba nauczycieli: $n = 260$ (analiza na danych nieważonych).

*Korelacja jest istotna na poziomie 0,05.

Wraz ze wzrostem o 1 odchylenie standardowe na skali „pomocny nauczyciel”, gimnazjaliści otrzymują średnio o 0,07 odchylenia standardowego wyższy wynik z części polonistycznej egzaminu. Silniejszym predyktorem, który znosi istotność skali „pomocny nauczyciel”, jest skala „autorytet nauczyciela/utrzymanie dyscypliny”. Wraz ze wzrostem o 1 odchylenie standardowe na skali „autorytet nauczyciela/utrzymanie dyscypliny”, gimnazjaliści otrzymują średnio o 0,11 odchylenia standardowego wyższy wynik z części polonistycznej egzaminu. Skala „autorytet nauczyciela/utrzymanie dyscypliny” wyjaśnia 81% wariancji efektu nauczyciela, przy kontroli wcześniejszych osiągnięć edukacyjnych, płci, dysleksji, inteligencji, zmiennych statusowych rodziny ucznia, lokalizacji gimnazjum i dynamiki bezrobocia w gminie, w której znajduje się gimnazjum.

Wnioski

Uzyskane oszacowania wyjaśnionej przez efektywność nauczyciela wariancji wyników uczniów są spójne z wynikami dostępnymi w literaturze przedmiotu. Wariancja wyników egzaminów, która może być przypisana nauczycielowi, wyniosła w raportowanych tu analizach 12% (matematyka) i 8% (kompetencje językowe). Rowan, Correnti i Miller (2002) raportują 18–28% (matematyka) i 12–23% (kompetencje językowe) wyjaśnionej przez efekt nauczyciela wariancji wyników uczniów. Nye, Konstantopoulos i Hedges (2004) odpowiednio 13–14% i 7–9%.

W modelu uwzględniającym wcześniejsze osiągnięcia, płeć, dysleksję, inteligencję i zmienne statusowe rodziny ucznia, efekt nauczyciela wyjaśnia 5% (matematyka) i 4% (kompetencje językowe) wariancji wyników egzaminu. Rowan, Correnti i Miller (2002) raportują odpowiednio 8–18% i 4–16%. Raportowany

efekt ma doniosłe praktyczne znaczenie wobec wielkości efektów innych czynników na poziomie klasy i szkoły, np. liczebności klasy (Koniewski, 2012).

Efektywny nauczyciel ma większy wpływ na procesy uczenia niż jakiegokolwiek inne czynniki kontrolowane przez system szkolny (Wright, Horn i Sanders, 1997; Rivkin, Hanushek i Kain, 2005). Stąd też nauczyciel powinien być centralnym punktem zainteresowania decydentów polityki oświatowej. Jeśli bowiem chcemy naprawiać coś w systemie edukacji, to powinniśmy zacząć od umiejętności nauczycielskich (Hanushek, 1997; Akiba, LeTendre i Scribner, 2007).

Skala „autorytet nauczyciela/utrzymanie dyscypliny” okazała się stabilnym i silnym predyktorem efektywności nauczyciela, zarówno w nauczaniu matematyki, jak i języka polskiego. Nauczyciele potrafiący utrzymać dyscyplinę na lekcjach, zaciekać tematem uczniów i zająć ich uwagę, przyczyniają się do podniesienia wyników uczniów średnio o 0,13 odchylenia standardowego wyników z matematyki i 0,11 odchylenia standardowego wyników z języka polskiego. Skala „autorytet nauczyciela/utrzymanie dyscypliny” wyjaśnia 91% wariancji efektu nauczyciela w nauczaniu matematyki i 81% w nauczaniu języka polskiego, przy kontroli wcześniejszych osiągnięć edukacyjnych, płci, inteligencji, zmiennych statusowych rodziny ucznia, lokalizacji gimnazjum.

Otrzymane wyniki utwierdzają w przekonaniu, że szacowanie efektu nauczyciela na podstawie polskich danych daje zbieżne rezultaty z wynikami otrzymanymi na podstawie danych gromadzonych głównie w Stanach Zjednoczonych. Godna uwagi jest przede wszystkim zgodność otrzymanych wyników z wynikami Nye, Konstantopoulos i Hedgesa, (2004), którzy analizowali wysokiej jakości dane eksperymentalne.

Literatura

- Akiba, M., LeTendre, G. K. i Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, 36(7), 369–387.
- Anderson, C. S. (1982). The search for school climate – a review of the research. *Review of Educational Research*, 52(3), 368–420.
- Armour, D. T. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica, CA: Rand Corporation.
- Ashton, P. i Crocker, L. (1987). Systematic study of planned variations – the essential focus of teacher – education reform. *Journal of Teacher Education*, 38(3), 2–8.
- Blau, P. M. i Duncan, O. D. (1967). *The American occupational structure*. New York: John Wiley and Sons.
- Bourdieu, P. (1986). The forms of social capital. W: J. Richardson (red.) *Handbook of theory and research for the sociology of education* (s. 241–258). New York, NY: Greenwood Press.
- Bourdieu, P. i Passeron, J. C. (1990). *Reprodukcja: elementy teorii systemu nauczania*. Warszawa: PWN.
- Bowles, S. S. i Levin, H. M. (1968). More on multicollinearity and the effectiveness of schools. *The Journal of Human Resources*, 3(3), 393–400.
- Brookover, W. B. (1982). *Creating effective schools: an in-service program for enhancing school learning climate and achievement*. Holmes Beach, FL: Learning Publications.
- Brown, C., Smith, M. i Stein, M. (1996). Linking teacher support to enhanced classroom instruction. Referat wygłoszony na konferencji American Educational Research Association, New York, NY.
- Brown, D. F. (2001). The value of advisory sessions for urban young adolescents. *Middle School Journal*, 32(4), 14–22.
- Bryk, A. S., i Raudenbush S. (1992). *Hierarchical linear models in social and behavioral research: applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- Byrne, B., Coventry, W. L., Olson, R. K., Wadsworth, S. J., Samuelsson, S., Petrill, S. A., Willcutt, E. G. i Corley, R. (2010). „Teacher effects” in early literacy development: evidence from a study of twins. *Journal of Educational Psychology*, 102(1), 32–42.
- Byrne, C. (1983). Teacher knowledge and teacher effectiveness: a literature review, theoretical analysis and discussion of research strategy. Referat wygłoszony na konferencji Northwestern Educational Research Association, Ellenville, NY.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (wyd. 2). Hillsdale, NJ: L. Erlbaum.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. B., Weinfeld, F. D., i York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: Government Printing Office.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1–44.
- Dolata, R. (2008). *Szkoła – segregacje – nierówności*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Domański, H. i Pokropek, A. (2011). *Podziały terytorialne, globalizacja a nierówności społeczne, Wprowadzenie do modeli wielopoziomowych*. Warszawa: Wydawnictwo IFiS PAN.
- Domański, H., Karpiński, Z., Pokropek, A., Przybysz, D., Sawiński, Z., Słomczyński, K. M. i Trzciński, R. (2012). *Metodologia badań nad stratyfikacją społeczną*. Warszawa: Wydawnictwo Naukowe Scholar.
- Edmonds, R. (1979). Effective schools for urban poor. *Educational Leadership*, 37(1), 15–24.
- Ehrenberg, R. G. i Brewer, D. J. (1995). Did teachers’ verbal ability and race matter in the 1960’s? Coleman revisited. *Economics of Education Review*, 14(1), 1–21.
- Evertson, C. M., Hawley, W. D. i Zlotnik, M. (1985). Making A difference in educational quality through teacher – education. *Journal of Teacher Education*, 36(3), 2–12.
- Ferguson, R. F. i Ladd, H. L. (1996). How and why money matters: an analysis of Alabama schools. W: H. F. Ladd (red.). *Holdine schools accountable*. Washington, D.C.: The Brookings Institution.
- Forero, C. G., Maydeu-Olivares, A., i Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling A Multidisciplinary Journal*, 16(4), 625–641.
- Freiberg, H. J. (red.). (1999). *School climate: measuring, improving and sustaining healthy learning environments*. Philadelphia, PA: Falmer Press.
- Goldhaber, D. D. i Brewer, D. J. (1997). Why don’t schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, 32(3), 505–523.

- Goldhaber, D. D. i Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129–145.
- Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement – estimation using micro data. *American Economic Review*, 61(2), 280–288.
- Hanushek, E. A. (1986). The economics of schooling production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141–1177.
- Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18(4), 45–62.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 100(1), 84–117.
- Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research*, 66(3), 397–409.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: an update. *Educational Evaluation and Policy Analysis*, 19(2), 141–164.
- Hart, S. A., Petrill, S. A. i Thompson, L. A. (2009). A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learning and Individual Differences*, 20(2), 63–69.
- Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Hedges, L. V., Laine, R. D., i Greenwald, R. (1994a). Does Money Matter? A Meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(3), 5–14.
- Hedges, L. V., Laine, R. D., i Greenwald, R. (1994b). Money does matter somewhere: a response to Hanushek. *Educational Researcher*, 23(4), 9–10.
- Hedges, L. V., Laine, R. D., i Greenwald, R. (1996). The Effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361–396.
- Jakubowski, M., Pokropek, A. (2009). *Badając egzamin. Podejście ilościowe w badaniach edukacyjnych*. Warszawa: CKE.
- Jordan, H. R., Mendro, R., Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement: a preliminary report on research on teacher effectiveness*. Referat wygłoszony w National Evaluation Institute, Indianapolis, IN.
- Klitgaard, R. E. i Hall, G. R. (1975). Are there unusually effective schools? *Journal of Human Resources*, 10(1), 90–106.
- Koniewski, M. (2012). Szacowanie wpływu liczebności klasy na osiągnięcia edukacyjne uczniów z wykorzystaniem eksperymentu ex post facto. *Edukacja*, 117(1), 23–43.
- Kovas, Y., Haworth, C. M. A., Petrill, S. A. i Plomin, R. (2007). Mathematical ability of 10-year-old boys and girls: genetic and environmental etiology of typical and low performance. *Journal of Learning Disabilities*, 40(6), 554–568.
- Kreft, I. i Leeuw, J. de (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage Publications.
- Lee, V. E. i Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 62(3), 172–192.
- Marzano, R. J., Marzano, J. S. i Pickering, D. J. (2003). *Classroom management that works*. Alexandria, VA: ASCD.
- McDonald, R. P., Ho, M.-H. (2002). Principles and Practice in reporting structural equation analyses. *Psychological methods*, 7(1), 64–82.
- McGue, M., Keyes, M., Sharma, A., Elkins, I., Legrand, L., Johnson, W. i Iacono, W. G. (2007). The environments of adopted and non-adopted youth: evidence on range restriction from the Sibling Interaction and Behavior Study (SIBS). *Behavior Genetics*, 37(3), 449–462.
- Mosteller, F. i Moynihan, D. P. (1972). *On equality of educational opportunity*. New York: Random House.
- Mullen, B. i Copper, C. (1994). The Relation between group cohesiveness and performance – an integration. *Psychological Bulletin*, 115(2), 210–227.
- Murnane, R. J. (1985). Do effective teachers have common characteristics: interpreting the quantitative research evidence. Referat wygłoszony na National Research Council Conference on Teacher Quality in Science and Mathematics, Washington, D.C.
- Murnane, R. J. i Phillips, B. R. (1981). What Do Effective Teachers of Inner-city Children Have In Common. *Social Science Research*, 10(1), 83–100.
- Nestler, S. (2012). A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, 66(1), 127–143.
- Nye, B., Konstantopoulos, S. i Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257.

- Oliver, B., Harlaar, N., Hayiou, T., Hayiou, M. E., Kovas, Y., Walker, S. O., Petrill, S. A., Spinath, F. M., Dale, P. S. i Plomin, R. (2004). A twin study of teacher – reported mathematics performance and low performance in 7-year-olds. *Journal of Educational Psychology*, 96(3), 504–517.
- Plowden, B. (1967). *Children and their primary schools*. London: HMSO.
- Pokropek, A. (2009). Metody statystyczne wykorzystywane w szacowaniu trzyletnich wskaźników egzaminacyjnych. W: B. Niemierko i M. K. Szmigiel (red.), *Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej* (s. 137–153). Kraków: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- Pokropek, A. (2013). *Efekt rówieśników w nauczaniu szkolnym*. Warszawa: Wydawnictwo IBE.
- Pokropek, A. i Żółtak, T. (2012). *Nowe modele jednorocznej EWD*. W: B. Niemierko i M. K. Szmigiel (red.), *Regionalne i lokalne diagnozy edukacyjne* (s.178–187). Kraków: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- Raudenbush, S. W. i Bryk, A. S. (2002). *Hierarchical Linear Models* (wyd. 2). Thousand Oaks: Sage Publications.
- Rice, J. M. (1902). *Educational research: a test in arithmetic*. The Forum, October, 281–297.
- Rivkin, S. G., Hanushek, E. A. i Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Robinson, G. E. (1990). Synthesis of research on the effects of class size. *Educational Leadership*, 47(7), 80–90.
- Rosenholtz, S. J. (1986). Organizational conditions of teacher learning. *Teaching and Teacher Education*, 2(2), 91–104.
- Rowan, B., Chiang, F. S. i Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, 70(4), 256–284.
- Rowan, B., Correnti, R. i Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: insights from the prospects study of elementary schools. *Teachers College Record*, 104(8), 1525–1567.
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.
- Sanders, W. L., i Rivers J. (1996). *Cumulative and residual effects of teachers on future student academic achievement. Research progress report*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Scheerens, J., i Bosker, R. (1997). *The foundations of educational effectiveness*. New York: Pergamon.
- Taylor, J., Roehrig, A. D., Hensler, B. S., Connor, C. M. i Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science*, 328(5977), 512–514.
- Wade, R. K. (1984). What makes a difference in inservice teacher education? A meta-analysis of research. *Educational Leadership*, 42(4), 48–54.
- Wayne, A. J. i Youngs, P. (2003). Teacher characteristics and student achievement gains: a review. *Review of Educational Research*, 73(1), 89–122.
- Wiley, D. E. i Yoon, B. (1995). Teacher reports on opportunity to learn-analyses of the 1993 California Learning Assessment System (CLAS). *Educational Evaluation and Policy Analysis*, 17(3), 355–370.
- Wilkinson, I. A. G. (2002). Introduction: peer influences on learning: where are they? *International Journal of Educational Research*, 37(5), 395–401.
- Woods, C. M. (2002). Factor analysis of scales composed of binary items: illustration with the Maudsley Obsessional Compulsive Inventory. *Journal of Psychopathology and Behavioral Assessment*, 24(4), 215–223.
- Wright, P. S., Horn. S. P. i Sanders. W. L. (1997). Teacher and classroom context effects on student achievement: implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57–67.

Podziękowania

Autor pragnie podziękować za wartościowe i inspirujące uwagi recenzentom EDUKACJI, a także uczestnikom seminarium doktoranckiego dra hab. Jarosława Górniaka, prowadzonego w Instytucie Socjologii Uniwersytetu Jagiellońskiego w Krakowie.

Aneks 1 – skale

Do utworzenia skal „autorytet nauczyciela/utrzymanie dyscypliny”, „aktywny styl nauczania” i „pomocny nauczyciel” wykorzystano konfirmacyjną analizę czynnikową (*Confirmatory Factor Analysis – CFA*). Analizy przeprowadzono w programie Mplus.

Zastosowano domyślną w Mplus parametryzację Delta. Wykorzystano estymację *Weighted Least Squares* (WLS). Zdecydowano się na wykorzystanie metody WLS, ponieważ jest metodą rekomendowaną w literaturze dla dużych prób ($N > 200$) i wysokich ładunków czynnikowych ($\lambda > 0,7$) (Nestler, 2013). Metoda WLS umożliwia także ocenę dopasowania modeli z wykorzystaniem statystyki χ^2 w tradycyjny sposób. W przypadku małej liczby wskaźników, tak jak w przedstawionych w artykule skalach, WLS jest rekomendowaną metodą, zamiast innej, często wykorzystywanej metody dla zmiennych kategoryalnych, tj. DWLS (*Diagonally Weighted Least Squares*) (Woods, 2002). DWLS ma tendencję do przeszacowywania parametrów (Forero, Maydeu-Olivares i Gallardo-Pujol, 2009; Nestler, 2013), jak również do akceptowania źle wyspecyfikowanych modeli (Nestler, 2013).

Jako miary dopasowania raportowane są: statystyka χ^2 , miara relatywnego dopasowania *Comparative Fit Index* (CFI) oraz miara przybliżonego dopasowania *Root Mean Square Error of Approximation* (RMSEA). Uznaje się je za standardowe, podawane są najczęściej w publikacjach raportujących wyniki CFA (McDonald i Ho, 2002). Statystyka χ^2 powinna być możliwie niska przy możliwie wielu stopniach swobody dla dobrego dopasowania. Przy istotności powyżej 0,05 uznaje się model za dobrze dopasowany (w przypadku dużych prób $n > 200$ poziom istotności zazwyczaj jest istotny). Dlatego w ocenie dopasowania należy kierować się miarami RMSEA i CFI. Zgodnie uznaje się RMSEA niższe niż 0,05, jako wskaźnik dobrego dopasowania, awartość w przedziale 0,05–0,08 jako akceptowalne dopasowanie. 90-procentowy przedział ufności dla RMSEA nie powinien zawierać wartości $> 0,8$. Miara CFI powyżej 0,9 wskazuje na dobre dopasowanie (McDonald i Ho, 2002).

Mimo pewnych różnic w dopasowaniu skal dla zmiennych dotyczących lekcji języka

polskiego i matematyki, dla jasności interpretacji zdecydowano zastosować takie same skale dla języka polskiego i matematyki.

Wskaźnikami skali „autorytet nauczyciela/utrzymanie dyscypliny” były odpowiedzi na pytania⁶:

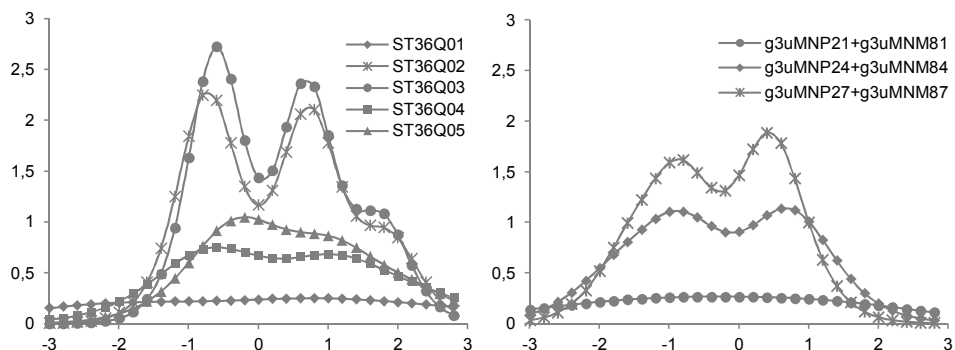
- A. „Nauczyciel musi długo czekać, zanim uczniowie się uspokoją i w klasie będzie cisza.” [mat: 0,820 0,007; pol: 0,743 0,010], (nazwa zmiennej dla mat: g3uMNN81, dla pol: g3uMNP21);
- B. „Uczniowie nie słuchają tego, co mówi nauczyciel.” [mat: 0,818 0,007; pol: 0,796 0,010], (dla mat: g3uMNN84, dla pol: g3uMNP24);
- C. „Uczniowie podczas lekcji zajmują się swoimi sprawami.” [mat: 0,865 0,006; pol: 0,792 0,010], (dla mat: g3uMNN87, dla pol: g3uMNP27).

Przy trzech wskaźnikach i jednym czynniku skale z definicji są idealnie dopasowane do danych. Im wyższe wartości skali, tym większy „autorytet nauczyciela/utrzymanie dyscypliny”. Skala ta została zaadaptowana z kwestionariusza uczniowskiego PISA OECD 2009 (kwestionariusz jest do pobrania ze strony pisa2009.acer.edu.au/downloads.php). Oryginalna skala składa się z następujących pozycji:

- a) Students don't listen to what the teacher says (ST36Q01);
- b) There is noise and disorder (ST36Q02);
- c) The teacher has to wait a long time for the students to quieten down (ST36Q03);
- d) Students cannot work well (ST36Q04);
- e) Students don't start working for a long time after the lesson begins (ST36Q05).

Oryginalna skala ma następujące właściwości: $n = 4892$; $\chi^2 = 450,549$; $df = 5$; $p = 0,000$; RMSEA = 0,135 (0,90 CI = 0,125 do 0,146;

⁶ Uczniowie odpowiadali na czteropunktowej skali: 1 – na każdej lekcji; 2 – na większości lekcji; 3 – czasami; 4 – nigdy lub prawie nigdy. W nawiasach kwadratowych podano ładunki czynnikowe, kursywą błędy standardowe.



Rysunek 1A

Krzywe informacyjne dla pozycji skali stosowanej w badaniu PISA 2009 i jej adaptacji w postaci skal „autorytet nauczyciela/utrzymanie dyscypliny” łącznie dla języka polskiego i matematyki.

$p = 0,000$; CFI = 0,985. Alfa Cronbacha wynosi 0,878. Alfa Cronbacha dla zaadaptowanej skali składającej się z 3. pozycji dla języka polskiego wynosi 0,769, a dla matematyki 0,821.

Na Rysunku 1A przedstawiono charakterystyki skal. Dane dopasowano do modelu *Graded Response Model* (GRM) (Samejima, 1969). Wykresy prezentują krzywe informacyjne pozycji skal. Odpowiednikami pozycji ST36Q03 z oryginalnej skali są pozycje g3uMNP21 i g3uMNM81; dla ST36Q01 to g3uMNP24 i g3uMNM84; dla ST36Q02 to g3uMNP27 i g3uMNM87.

Skala „aktywny styl nauczania” składa się z czterech skal podrzędnych, których wskaźniki utworzono na podstawie odpowiedzi na pytania:

A. Skala podrzędna: „indywidualizacja nauczania” [mat: 0,863 0,010; pol: 0,924 0,011]:

(A1) „Nauczyciel przydziela uczniom różne zadania, biorąc pod uwagę ich wyniki w nauce.” [mat: 0,804 0,009; pol: 0,756 0,010];

(A2) „Nauczyciel przydziela uczniom różne zadania, biorąc pod uwagę ich zainteresowania.” [mat: 0,832 0,009; pol: 0,797 0,010].

B. Skala podrzędna: „korzystanie z multimedialnych” [mat: 0,807 0,012; pol: 0,840 0,013]:

(B1) „Nauczyciel zachęca do korzystania z dodatkowych materiałów, np. artykułów, informacji z internetu.” [mat: 0,864 0,010; pol: 0,44 0,012];

(B2) „Nauczyciel korzysta z materiałów multimedialnych, programów edukacyjnych.” [mat: 0,701 0,011; pol: 0,650 0,013].

C. Skala podrzędna: „podsumowanie przeobrażonego materiału” [mat: 0,862 0,010; pol: 0,880 0,014]:

(C1) „Pod koniec lekcji nauczyciel podsumowuje najważniejsze zagadnienia.” [mat: 0,824 0,009; pol: 0,724 0,012];

(C2) „Na początku lekcji nauczyciel podsumowuje to, co zrobiliśmy na poprzednich zajęciach.” [mat: 0,755 0,009; pol: 0,657 0,012].

D. Skala podrzędna: „stymulowanie pracy w grupach” [mat: 0,782 0,013; pol: 0,699 0,015]:

(D1) „Nauczyciel dzieli uczniów na grupy, w których uczniowie wspólnie pracują.” [mat: 0,770 0,011; pol: 0,675 0,013];

(D2) „Uczniowie wspólnie tworzą plakaty odnoszące się do zagadnień poruszanych na lekcji.” [mat: 0,877 0,013; pol: 0,800 0,014].

Dla matematyki skala ma następujące właściwości: mat: $n = 5071$; $\chi^2 = 232,110$; $df = 16$;

$p = 0,000$; RMSEA = 0,052 (0,90 CI = 0,046 do 0,058); $p = 0,314$; CFI = 0,983. Dla języka polskiego: pol: $n = 5069$; $\chi^2 = 241,969$; $df = 16$; $p = 0,000$; RMSEA = 0,053 (0,90 CI = 0,047 do 0,059); $p = 0,208$; CFI = 0,966. Dopasowanie do danych jest akceptowalne. Dla ułatwienia interpretacji wartości skali zostały pomnożone przez -1. Im wyższe wartości na skali, tym bardziej „aktywizujący styl nauczania”.

Wskaźnikami skali „pomocny nauczyciel” były odpowiedzi na pytania:

- A. „Nauczyciel wykazuje zainteresowanie postępami w nauce każdego z uczniów.” [mat: 0,709 0,009; pol: 0,660 0,010];
- B. „Nauczyciel tłumaczy tak długo, aż wszyscy uczniowie zrozumieją.” [mat: 0,871 0,006; pol: 0,796 0,007];
- C. „Nauczyciel chętnie pomaga uczniom w nauce.” [mat: 0,896 0,006; pol: 0,874 0,007];
- D. „Zajęcia są starannie zaplanowane.” [mat: 0,652 0,010; pol: 0,650 0,010];
- E. „Nauczyciel podczas lekcji zachęca uczniów do zadawania pytań, dotyczących

przerabianego materiału.” [mat: 0,669 0,010; pol: 0,603 0,011].

Dla matematyki skala ma następujące właściwości: $n = 5072$; $\chi^2 = 97,626$; $df = 5$; $p = 0,000$; RMSEA = 0,060 (0,90 CI = 0,050 do 0,071); $p = 0,045$; CFI = 0,994. Dla języka polskiego: $n = 5071$; $\chi^2 = 86,771$; $df = 5$; $p = 0,000$; RMSEA = 0,057 (0,90 CI = 0,047 do 0,068); $p = 0,131$; CFI = 0,991. Dopasowanie do danych jest akceptowalne. Dla ułatwienia interpretacji wartości skali zostały pomnożone przez -1. Im wyższe wartości skali, tym bardziej „pomocny” nauczyciel.

Aneks 2 – modele hierarchiczne

Modele hierarchiczne obliczono w programie HLM7. Stosowano tylko modele z losowymi stałymi. Wszystkie zmienne podczas wprowadzania do modeli były „niecentrowane”. Oczywiście, te zmienne, które w oryginalnej postaci wyrażone są na skali Z, mają średnią równą 0 i odchylenie standardowe równe 1.

$$\begin{aligned}
 1\text{MAT: ZIRTMAT} &= \gamma_{000} + \gamma_{100} * \text{ZSP3PL} + \gamma_{200} * \text{EWD_PLEC} + \gamma_{300} * \text{DYS} + \gamma_{400} * \text{RAVEN} + \\
 &\quad \gamma_{500} * \text{HEDULL} + \gamma_{600} * \text{HOMEPOS} + \gamma_{700} * \text{WIELRODZ} + r_0 + u_{00} + e \\
 2\text{MAT: ZIRTMAT} &= \gamma_{000} + \gamma_{001} * \text{LOK} + \gamma_{002} * \text{YOYBEZ} + \gamma_{010} * \text{ZSPRAWD2} + \gamma_{100} * \text{ZSP3PL} \\
 &\quad + \gamma_{200} * \text{EWD_PLEC} + \gamma_{300} * \text{DYS} + \gamma_{400} * \text{RAVEN} + \gamma_{500} * \text{HEDULL} + \gamma_{600} * \text{HOMEPOS} + \\
 &\quad \gamma_{700} * \text{WIELRODZ} + r_0 + u_{00} + e \\
 3\text{MAT: ZIRTMAT} &= \gamma_{000} + \gamma_{001} * \text{LOK} + \gamma_{002} * \text{YOYBEZ} + \gamma_{010} * \text{ZSPRAWD2} + \gamma_{020} * \text{ACTIVM5} + \\
 &\quad \gamma_{100} * \text{ZSP3PL} + \gamma_{200} * \text{EWD_PLEC} + \gamma_{300} * \text{DYS} + \gamma_{400} * \text{RAVEN} + \gamma_{500} * \text{HEDULL} + \gamma_{600} * \text{HOMEPOS} + \\
 &\quad \gamma_{700} * \text{WIELRODZ} + r_0 + u_{00} + e \\
 4\text{MAT: ZIRTMAT} &= \gamma_{000} + \gamma_{001} * \text{LOK} + \gamma_{002} * \text{YOYBEZ} + \gamma_{010} * \text{ZSPRAWD2} + \gamma_{020} * \text{DYSM} + \\
 &\quad \gamma_{100} * \text{ZSP3PL} + \gamma_{200} * \text{EWD_PLEC} + \gamma_{300} * \text{DYS} + \gamma_{400} * \text{RAVEN} + \gamma_{500} * \text{HEDULL} + \gamma_{600} * \text{HOMEPOS} \\
 &\quad + \gamma_{700} * \text{WIELRODZ} + r_0 + u_{00} + e \\
 1\text{POL: ZIRTPOL} &= \gamma_{000} + \gamma_{100} * \text{ZSP3PL} + \gamma_{200} * \text{EWD_PLEC} + \gamma_{300} * \text{RAVEN} + \gamma_{400} * \text{HEDULL} + \\
 &\quad \gamma_{500} * \text{HOMEPOS} + \gamma_{600} * \text{WIELRODZ} + r_0 + u_{00} + e \\
 2\text{POL: ZIRTPOL} &= \gamma_{000} + \gamma_{001} * \text{LOK} + \gamma_{100} * \text{ZSP3PL} + \gamma_{200} * \text{EWD_PLEC} + \gamma_{300} * \text{RAVEN} + \gamma_{400} * \text{HEDULL} + \\
 &\quad \gamma_{500} * \text{HOMEPOS} + \gamma_{600} * \text{WIELRODZ} + r_0 + u_{00} + e \\
 3\text{POL: ZIRTPOL} &= \gamma_{000} + \gamma_{001} * \text{LOK} + \gamma_{010} * \text{POMOCNP} + \gamma_{100} * \text{ZSP3PL} + \gamma_{200} * \text{EWD_PLEC} + \\
 &\quad \gamma_{300} * \text{RAVEN} + \gamma_{400} * \text{HEDULL} + \gamma_{500} * \text{HOMEPOS} + \gamma_{600} * \text{WIELRODZ} + r_0 + u_{00} + e \\
 4\text{POL: ZIRTPOL} &= \gamma_{000} + \gamma_{001} * \text{LOK} + \gamma_{010} * \text{DYSP} + \gamma_{100} * \text{ZSP3PL} + \gamma_{200} * \text{EWD_PLEC} + \gamma_{300} * \text{RAVEN} + \\
 &\quad \gamma_{400} * \text{HEDULL} + \gamma_{500} * \text{HOMEPOS} + \gamma_{600} * \text{WIELRODZ} + r_0 + u_{00} + e
 \end{aligned}$$

Rysunek 2A. Wzory dla raportowanych w artykule modeli hierarchicznych.