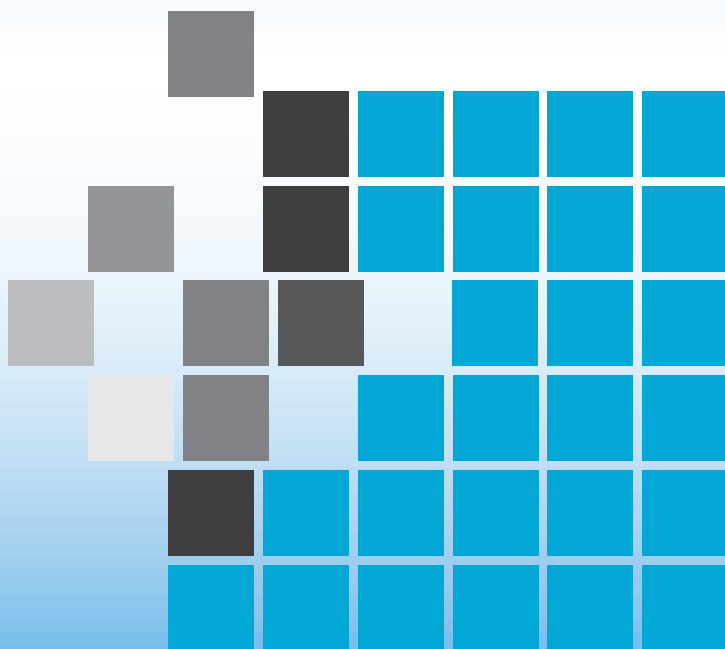


Roman Dolata, Anna Hawrot, Grzegorz Humenny,
Aleksandra Jasińska, Maciej Koniewski,
Przemysław Majkut, Tomasz Żółtak

Trafność metody edukacyjnej wartości dodanej dla gimnazjów



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

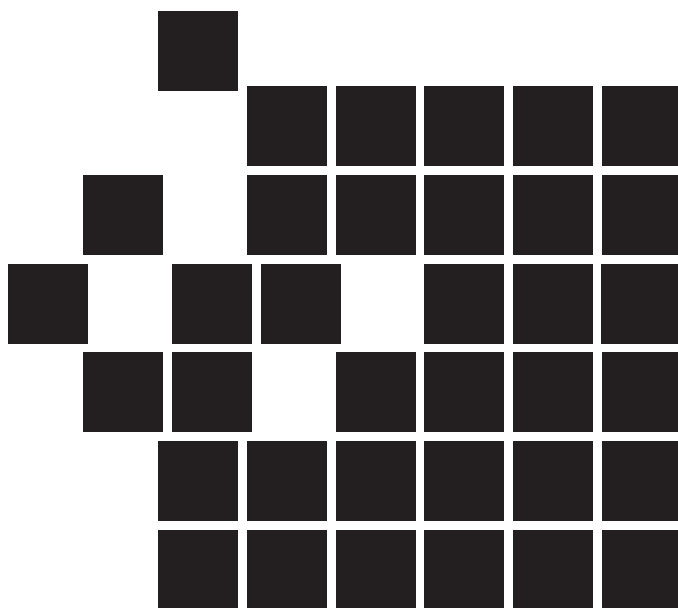


**EDUKACYJNA
WARTOŚĆ
DODANA**

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Trafność metody edukacyjnej wartości dodanej dla gimnazjów



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



EDUKACYJNA
WARTOŚĆ
DODANA

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Autorzy:

Roman Dolata, Instytut Badań Edukacyjnych
Anna Hawrot, Instytut Badań Edukacyjnych
Grzegorz Humenny, Instytut Badań Edukacyjnych
Aleksandra Jasińska, Instytut Badań Edukacyjnych
Maciej Koniewski, Instytut Badań Edukacyjnych
Przemysław Majkut, Instytut Badań Edukacyjnych
Tomasz Żółtak, Instytut Badań Edukacyjnych

Recenzenci:

dr hab. Roman Konarski, profesor UG, Instytut Psychologii Uniwersytetu Gdańskiego
dr Zbigniew Karpiński, Instytut Filozofii i Socjologii PAN

Redakcja językowa:

Karolina Kwiatosz

Wydawca:

Instytut Badań Edukacyjnych
ul. Górczewska 8, 01-180 Warszawa
tel. (22) 241 71 00; www.ibe.edu.pl

© Copyright by: Instytut Badań Edukacyjnych, Warszawa 2013

ISBN 978-83-61693-28-4

Skład:

Drukarnia TINTA, Z. Szymański, ul. Żwirki i Wigury 22, 13-200 Działdowo
www.drukarniatinta.pl

Druk:

Drukarnia GC, Grzegorz Cielecki, ul. Sycowska 20, 02-266 Warszawa

Publikacja współfinansowana przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach projektu: *Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD).*

Egzemplarz bezpłatny

Spis treści

Wprowadzenie i najważniejsze wyniki uzyskane w badaniu trafności metody EWD dla gimnazjów	9
--	----------

Rozdział 1

Organizacja i metodologia badania podłużnego w gimnazjum	26
1.1. Koncepcja badania	27
1.2. Realizacja badania	31
1.3. Etap I	31
1.3.1. Definicja populacji i opis próby badawczej	31
1.3.2. Losowanie	31
1.3.3. Próba badawcza oraz stopy realizacji w grupach respondentów	33
1.3.4. Przebieg I etapu badania	34
1.4. Etap II	36
1.4.1. Próba badawcza oraz stopy realizacji w grupach respondentów	36
1.4.2. Przebieg II etapu badania	37
1.5. Etap III	38
1.5.1. Próba badawcza oraz stopy realizacji w grupach respondentów	38
1.5.2. Przebieg III etapu badania	39
1.6. Zmienne zależne używane w analizach	40
1.7. Reprezentatywność próby	42
1.8. Wyniki egzaminacyjne	43
1.8.1. Sprawdzian po klasie szóstej	43
1.8.2. Egzamin gimnazjalny, część humanistyczna	44
1.8.3. Egzamin gimnazjalny, część matematyczno-przyrodnicza	45
1.8.4. Wyniki egzaminacyjne w szkołach, które odmówiły uczestnictwa w badaniu	46
1.8.5. Międzyszkolne i międzyoddziałowe zróżnicowanie wyników egzaminacyjnych	49
1.9. Wagi stosowane w analizach	53
1.9.1. Wagi szkół, klas i uczniów	53
1.9.2. Przypadki szczególne	54
1.9.3. Wagi nauczycieli	54
1.9.4. Modele podstawowe stosowane w analizach	54
1.10. Analiza modeli dwupoziomowych analogicznych do populacyjnych modeli EWD	55
1.11. Analiza modeli trzypoziomowych i testowanie oddziaływania zmiennych dodatkowych	57

Rozdział 2

Trafność testów egzaminacyjnych	61
2.1. Rzetelność i błędy pomiaru	62
2.2. Reprezentatywność treści nauczania	68
2.3. Zewnętrzne kryteria trafności egzaminów gimnazjalnych	72
2.4. Spójność mierzonych konstruktów	79

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

2.5. Skalowanie wyników	84
2.6. Podsumowanie	87

Rozdział 3

Gimnazjalne modele edukacyjnej wartości dodanej	90
3.1. Zmienne kontrolne w modelach EWD	90
3.2. Specyfika polskich modeli gimnazjalnych EWD	93
3.3. Jednoroczne wskaźniki EWD	95
3.4. Trzyletnie wskaźniki EWD	98

Rozdział 4

Wiek i inteligencja ucznia	103
4.1. Wiek a osiągnięcia szkolne	104
4.1.1. Dlaczego warto podjąć problem wieku w kontekście badania trafności wskaźników EWD i rozwoju metodologii ich szacowania?	106
4.2. Inteligencja	108
4.2.1. Inteligencja a osiągnięcia szkolne	110
4.3. Pytania badawcze i hipotezy	112
4.4. Metoda	113
4.4.1. Pomiar wieku	113
4.4.2. Pomiar inteligencji	114
4.4.3. Schemat analizy danych	115
4.5. Wyniki	117
4.5.1. Wiek a wyniki egzaminu matematyczno-przyrodniczego i EWD	117
4.5.2. Wiek a wyniki egzaminu z przedmiotów humanistycznych i EWD	119
4.5.3. Podsumowanie znaczenia wieku dla wyników egzaminacyjnych i EWD	121
4.5.4. Międzyszkolna i międzyoddziałowa wariancja inteligencji uczniów	122
4.5.5. Inteligencja a wyniki egzaminu matematyczno-przyrodniczego i EWD	124
4.5.6. Inteligencja a wyniki egzaminu z przedmiotów humanistycznych i EWD	127
4.5.7. Efektywność nauczania a przyrost inteligencji uczniów	129
4.6. Podsumowanie	134

Rozdział 5

Status społeczny rodziny ucznia	140
5.1. Kontekst teoretyczny i pytania badawcze	140
5.2. Metoda	145
5.2.1. Pomiar czynników statusowych	145
5.2.2. Schemat analizy danych	149
5.3. Wyniki	150
5.3.1. Międzyszkolna i międzyoddziałowa wariancja wskaźników statusowych	150
5.3.2. Status społeczny a wyniki egzaminu gimnazjalnego w części matematyczno-przyrodniczej	151
5.3.3. Znaczenie czynników statusowych dla szacowania EWD w obszarze matematyczno-przyrodniczym	154
5.3.4. Status ekonomiczny a wyniki egzaminu gimnazjalnego w części humanistycznej	156

5.3.5. Znaczenie czynników statusowych dla szacowania EWD w obszarze humanistycznym	159
5.4. Podsumowanie	160
Rozdział 6	
Korepetycje i rodzinne wsparcie edukacyjne	166
6.1. Korepetycje i wsparcie edukacyjne	168
6.1.1. Korepetycje i wsparcie edukacyjne w świetle badań	169
6.2. Metoda	174
6.2.1. Pytania badawcze	174
6.2.2. Schemat pomiarów	175
6.2.3. Schemat analizy danych	178
6.3. Wyniki	180
6.3.1. Skala zjawiska korepetycji oraz nieodpłatnej pomocy w nauce i odrabianiu lekcji	180
6.3.2. Korelaty korzystania z korepetycji	182
6.3.3. Korelaty korzystania z bezpłatnej pomocy w nauce i odrabianiu lekcji	186
6.3.4. Korepetycje a wyniki egzaminacyjne i EWD	192
6.3.5. Bezpłatne wsparcie w nauce a wyniki egzaminacyjne i EWD	198
6.4. Podsumowanie	205
Rozdział 7	
Spółeczno-gospodarcze charakterystyki gmin	210
7.1. Spółeczno-gospodarcze uwarunkowania osiągnięć szkolnych	210
7.1.1. Zamożność	211
7.1.2. Wydatki na oświatę	211
7.1.3. Bezrobocie	211
7.1.4. Struktura lokalnej gospodarki	212
7.1.5. Zasoby kulturowe	212
7.1.6. Przedszkola	213
7.1.7. Problem uwzględniania czynników lokalnych z punktu widzenia modeli EWD	213
7.2. Uwarunkowania lokalne: konstrukcja wskaźników i ich rozkłady w próbie	214
7.2.1. Dane na temat zamożności	214
7.2.2. Dane na temat lokalnej gospodarki	214
7.2.3. Dane dotyczące poziomu bezrobocia	215
7.2.4. Dane o wydatkach na oświatę	215
7.2.5. Dane na temat dostępności zasobów kulturowych	216
7.2.6. Dane dotyczące opieki przedszkolnej	216
7.2.7. Dane kontekstowe	216
7.3. Pytania i hipotezy badawcze	217
7.3.1. Schemat analizy danych	219
7.4. Wyniki	220
7.4.1. Zamożność	220
7.4.2. Bezrobocie	220
7.4.3. Gospodarka	230

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

7.4.4. Wydatki na oświatę	231
7.4.5. Zasoby kulturowe	237
7.4.6. Opieka przedszkolna	246
7.4.7. Czynniki kontrolne	247
7.4.8. Model z kontrolą kilku czynników na poziomie gmin	254
7.5. Podsumowanie	259

Rozdział 8

Selekcja uczniów na progu gimnazjum **264**

8.1. Selekcja i segregacja uczniów jako problem badawczy	264
8.2. Pytania badawcze i hipotezy	266
8.3. Używane narzędzia badawcze	267
8.4. Wskaźniki selekcji	268
8.5. Schemat analizy danych	271
8.6. Wyniki dla części matematyczno-przyrodniczej	272
8.7. Wyniki dla części humanistycznej	272
8.8. Podsumowanie	275

Rozdział 9

Style zarządzania szkołą **278**

9.1. Podstawy teoretyczne	279
9.1.1. Partycypacja w zarządzaniu	279
9.1.2. Kultura organizacyjna szkoły	281
9.2. Adaptacja narzędzi badawczych kierowanych do nauczycieli	283
9.3. Schemat analizy danych	283
9.4. Przygotowanie danych do analizy	285
9.4.1. Partycypacja w zarządzaniu	285
9.4.2. Kultura organizacyjna	286
9.5. Wyniki	287
9.5.1. Partycypacja w zarządzaniu	287
9.5.2. Kultura organizacyjna	290
9.6. Podsumowanie	294

Rozdział 10

Cechy nauczyciela **299**

10.1. Wpływ nauczyciela na osiągnięcia szkolne	300
10.2. Dowody z analiz wariancji międzyoddziałowej	300
10.3. Dowody z modelowania funkcji produkcyjnej szkół	301
10.4. Przygotowanie danych do analizy	305
10.5. Style nauczania i przebiegu lekcji	307
10.6. Selekcja zmiennych niezależnych	307
10.7. Wyniki dla części matematycznej	311
10.8. Wyniki dla części polonistycznej	316
10.9. Podsumowanie	321

Aneks do rozdziału 1	
Organizacja i metodologia badania podłużnego w gimnazjum	327
Aneks do rozdziału 2	
Trafność testów egzaminacyjnych	334
Aneks do rozdziału 4	
Wiek i inteligencja uczniów	337
Aneks do rozdziału 5	
Status społeczny rodziny ucznia	346
Aneks do rozdziału 6	
Korepetycje i rodzinne wsparcie edukacyjne	349
Aneks do rozdziału 7	
Społeczno-gospodarcze charakterystyki gmin	363
Aneks do rozdziału 10	
Cechy nauczyciela	364

Wprowadzenie i najważniejsze wyniki uzyskane w badaniu trafności metody EWD dla gimnazjów

Badanie trafności metody EWD jest jednym z głównych elementów realizowanego w latach 2007–2013 projektu „Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD)”. Głównym celem projektu jest rozwój metody edukacyjnej wartości dodanej i propagowanie narzędzi analitycznych pozwalających wykorzystywać tę metodę w analizie wyników egzaminacyjnych.

Krajowe egzaminy szkolne pełnią wiele funkcji. Trzy najważniejsze to: wyznaczanie stopnia spełniania przez uczniów wymagań programowych, monitorowanie procesów edukacyjnych oraz ewaluacja pracy szkół w zakresie efektywności nauczania. Prace nad metodą edukacyjnej wartości dodanej służą wzmocnieniu ewaluacyjnej funkcji egzaminów zewnętrznych. Metoda edukacyjnej wartości dodanej jest ważnym narzędziem statystycznym, które może być używane do analizy wyników egzaminacyjnych w ewaluacji pracy szkół.

Jednolite egzaminy pozwalają spojrzeć nauczycielom, nadzorowi pedagogicznemu oraz samorządom na wyniki nauczania w szerokiej perspektywie. Porównywalność wyników w skali całej Polski umożliwia przyjrzenie się osiągnięciom szkolnym z punktu widzenia kraju, województwa czy gminy. Jednak obraz nauczania szkolnego widziany tylko przez pryzmat wyników końcowych nauczania jest niepełny. Równie ważny, a nawet ważniejszy jest postęp, jaki robią uczniowie przez kilka lat nauki w danej placówce. Metoda edukacyjnej wartości dodanej stara się w sposób poprawny statystycznie wyznaczyć wkład szkoły w ten postęp, dzięki temu staje się ważnym narzędziem analizy efektywności nauczania. Metoda edukacyjnej wartości dodanej ma służyć przede wszystkim szkołom. To dyrektorzy i nauczyciele mogą dzięki właściwej analizie wyników egzaminacyjnych zobaczyć mocne i słabe strony nauczania w swojej placówce. Analiza wyników testów egzaminacyjnych z użyciem metody edukacyjnej wartości dodanej może dostarczać ważnych przesłanek do formułowania planów rozwoju szkoły, a także umożliwia ocenę skuteczności już podjętych działań. Ewaluacja prowadzona przez samą szkołę w celu poprawy swojego funkcjonowania to najważniejszy obszar wykorzystania wyników egzaminacyjnych, a tym samym i metody edukacyjnej wartości dodanej. Beneficjentami skutecznej autoewaluacji szkoły stają się uczniowie, którzy mają szansę uczyć się w skuteczniej nauczających szkołach. Metoda edukacyjnej wartości dodanej może też służyć nadzorowi szkolnemu. Trzecim potencjalnym użytkownikiem metody edukacyjnej wartości dodanej są samorządy. Bez wchodzenia w kompetencje nadzoru pedagogicznego mogą one monitorować

efektywność nauczania w szkołach na swoim terenie. Spełnienie się tych wszystkich nadziei pokładanych w metodzie EWD zależy jednak od jej trafności. Nie można tego przyjąć za pewnik. To wymaga empirycznych dowodów. Wszystkie dotychczasowe działania aplikacyjne prowadzone w ramach projektu EWD miały charakter pilotażowy. Podejmując badania trafności, byliśmy gotowi przyjąć, że zdobyta w ich wyniku wiedza skłoni nas do negatywnej oceny metody i odstąpienia od jej aplikacji.

Historia prac nad metodą edukacyjnej wartości dodanej w Polsce

Po raz pierwszy dane niezbędne do wykorzystania modeli EWD w analizie wyników egzaminacyjnych wytworzone zostały w polskim systemie oświaty w 2005 roku. Wtedy właśnie gimnazja ukończył pierwszy rocznik gimnazjalistów, dla których był dostępny zarówno wynik sprawdzianu w klasie szóstej, jak i egzaminu gimnazjalnego. W 2006 roku ukazały się pierwsze publikacje pokazujące możliwości metody EWD w analizie wyników gimnazjalnych (Dolata, 2006a; 2006b). W tekstach tych wykorzystywano proste modele regresji, w których zmienną wyjaśnianą był wynik z danej części egzaminu gimnazjalnego, a zmienną wyjaśniającą wynik na sprawdzianie w klasie szóstej szkoły podstawowej. Wskaźniki EWD dla gimnazjów czy oddziałów gimnazjalnych szacowane były punktowo, jako średnie reszt z równania regresji. Ze względu na brak scalonych krajowych baz danych z połączonymi wynikami sprawdzianu i egzaminu gimnazjalnego analizy prowadzono na poziomie lokalnym (np. gimnazja w danym mieście). Najważniejszym osiągnięciem tego okresu było uświadomienie decydentom potencjału analitycznego tkwiącego w połączonych, podłużnie analizowanych wynikach egzaminacyjnych. Dzięki temu powołano przy Centralnej Komisji Egzaminacyjnej zespół badawczy ds. rozwoju metody EWD (pierwszy projekt EFS, 2005–2007). Pierwszy, skromny budżetowo projekt, pozwolił rozpocząć prace nad rozwojem metody EWD na potrzeby polskiego systemu egzaminacyjnego. Pomoc okręgowych komisji egzaminacyjnych umożliwiła scalenie wyników sprawdzianu i egzaminu gimnazjalnego w jedną zintegrowaną bazę danych. Dzięki temu mogła powstać pierwsza edycja Kalkulatora EWD dla gimnazjów. Był to arkusz kalkulacyjny z zaimplementowanym algorytmem wyliczania wskaźników EWD dla szkół lub innych grup uczniów (oddziały szkolne, chłopcy, dziewczęta itp.). Algorytm wykorzystywał oszacowane dla danych krajowych wartości przewidywane dla uczniów o danym wyniku na sprawdzianie. W estymacji wartości przewidywanych wykorzystywano dodatkowo zmienne kontrolne dostępne w systemie egzaminacyjnym – informacje o płci i dysleksji. Wskaźniki EWD szacowane były punktowo i przedziałowo. Szkoły mogły bezpłatnie pobierać ze strony Kalkulator EWD i po wprowadzeniu danych swoich absolwentów przeprowadzać analizy na potrzeby ewaluacji wewnętrznej. W ramach projektu odbyło

WPROWADZENIE I NAJWAŻNIEJSZE WYNIKI UZYSKANE W BADANIU TRAFNOŚCI METODY EWD DLA GIMNAZJÓW

się pierwsze, kaskadowe szkolnie dyrektorów gimnazjów w wykorzystaniu metody EWD w ewaluacji pracy szkoły. W 2007 roku powstała kolejna edycja Kalkulatora EWD, równolegle testowano bardziej złożone modele szacowania EWD: modele wielopoziomowe i modele efektów mieszanych. W roku tym uruchomiono też przy CKE kolejny zespół badawczy ds. rozwoju metody EWD (EFS, 2007–2013). Dysponował on już znacznym budżetem i możliwościami działania. W 2008 roku powstała strona www.ewd.edu.pl. Z czasem strona przerodziła się w portal wiedzy na temat wykorzystania wyników egzaminacyjnych do ewaluacji pracy szkół. W tym samym roku odbyło się kolejne masowe szkolenie potencjalnych użytkowników metody EWD pod nazwą Wiosenna Szkoła EWD. Szkoły EWD stały się w kolejnych latach główną formą upowszechniania wiedzy o metodzie i możliwościach jej wykorzystania. W 2008 roku ukazała się też przygotowana przez członków zespołu pierwsza polska monografia naukowa poświęcona metodzie EWD (Dolata, 2008). W 2008 roku miało miejsce jeszcze jedno ważne wydarzenie – wyszedł raport OECD opracowany przez międzynarodową grupę ekspertów poświęcony metodzie EWD *Measuring Improvements in Learning Outcomes*. Jednym z nich był Maciej Jakubowski, członek zespołu EWD. Następny rok zaowocował opracowaniem całkowicie nowego Kalkulatora EWD. Pod nazwą Kalkulator EWD Plus kryła się samodzielna aplikacja komputerowa do analiz wyników egzaminacyjnych dla gimnazjów. W tym samym roku po raz pierwszy zespół w trybie pilotażowym upublicznia na stronie projektu trzyletnie wskaźniki egzaminacyjne dla gimnazjów. Wskaźniki w nowatorski, graficzny sposób pokazują pozycję danego gimnazjum w dwuwymiarowej przestrzeni – jedna oś to średni wynik egzaminu, druga to miara EWD. Niepewność szacowania pokazana jest graficznie za pomocą przybierającego kształt elipsy 95-procentowego obszaru ufności. Wskaźniki szacowane są za pomocą modeli efektów mieszanych jako Bayesowskie predykcje a posteriori. Wyniki egzaminacyjne przedstawiane są na standardowej skali wyników indywidualnych uczniów o średniej krajowej 100 i odchyleniu standardowym 15. Wskaźniki EWD mają z założenia średnią krajową 0, a jednostka jest ta sama, co w wypadku wyniku egzaminacyjnego.

O ile Kalkulator EWD z założenia był przeznaczony do wykorzystania w ewaluacji wewnętrznej efektywności nauczania, to publikowane w internecie, powszechnie dostępne wskaźniki trzyletnie miały służyć ewaluacji zewnętrznej. W założeniu użytkownikami tych wskaźników miały być organy nadzoru pedagogicznego i organy prowadzące. Oczywiście i same szkoły mogły z nich korzystać. Wskaźniki trzyletnie były tworzone z myślą o profesjonalistach, ale mogli z nich korzystać również zainteresowani rodzice i uczniowie. Jednak nie byli oni grupą docelową w projekcie. Pod koniec 2009 roku rozpoczęła się też realizacja trzech badań podłużnych mających dostarczyć informacji potrzebnych do rozwoju metody EWD, w tym do oceny jej trafności. Właśnie jednym z nich jest badanie podłużne w gimnazjach,

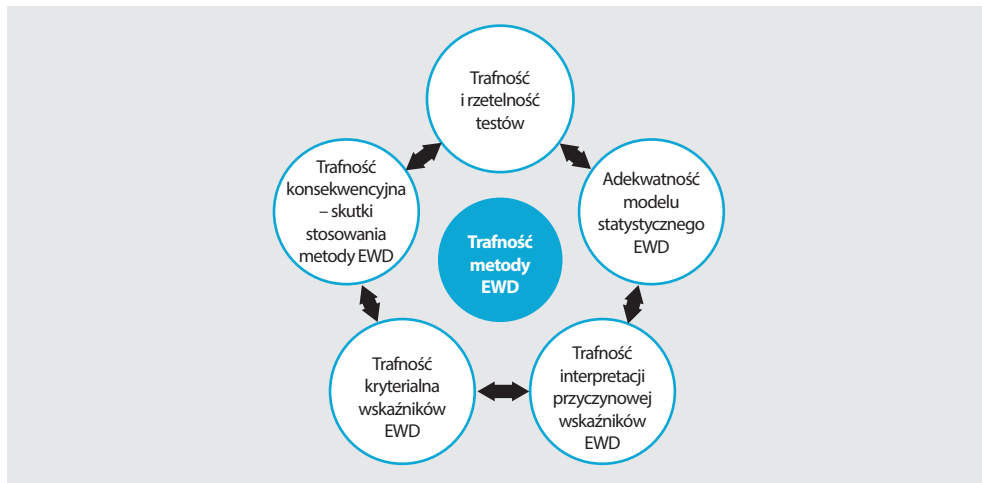
którego wyniki są opisywane w tym raporcie. Pozostałe to podłużne badanie w szkołach ponadgimnazjalnych (na zlecenie CKE badanie realizował IFiS PAN) i szkołach podstawowych. W 2010 roku upubliczniono po raz pierwszy wskaźniki egzaminacyjne dla liceów ogólnokształcących i techników w zakresie matematyki i języka polskiego. Graficzna postać wskaźników była analogiczna jak w wypadku gimnazjów, jednak proces ich wyliczania był bardziej złożony. Na podstawie wykonania zadań maturalnych na poziomie podstawowym i rozszerzonym szacowano za pomocą metody IRT poziom umiejętności matematycznych i polonistycznych maturzystów, następnie analogicznie jak w wypadku wskaźników gimnazjalnych szacowano średni wynik i EWD szkoły. Skale standaryzowano (100;15) oddzielnie dla krajowej populacji liceów ogólnokształcących i techników. W 2011 roku doskonalono modele EWD dla gimnazjów. Stworzono też spójną w czasie bazę szkół, dzięki czemu można było lepiej przypisywać wyniki egzaminacyjne poszczególnym placówkom, co stało się palącą potrzebą w obliczu publikowania wskaźników EWD dla kilku okresów trzyletnich dla każdej szkoły. W tym samym roku znacząco zmodyfikowano wskaźniki dla LO i techników. Wprowadzono trzy syntetyczne wskaźniki: umiejętności w zakresie przedmiotów humanistycznych, matematyczno-przyrodniczych i oddzielnie matematyki. Wspólne skalowanie zadań z arkuszy egzaminacyjnych z różnych przedmiotów wymagało zastosowania poza modelami IRT, również modeli selekcji pozwalających rozwiązać problem różnic w poziomie umiejętności subpopulacji przystępujących do egzaminu maturalnego z różnych przedmiotów i na różnych poziomach. Zmiana struktury egzaminu gimnazjalnego wymusiła w 2012 roku opracowanie nowej aplikacji komputerowej – Kalkulatora EWD 100. Narzędzie to choć funkcjonalnie podobne do Kalkulatora EWD Plus ma inną architekturę, dzięki której łatwo można implementować nowe funkcjonalności i nowe typy wskaźników egzaminacyjnych. Ważną zmianą jest też przejście w Kalkulatorze EWD 100, analogicznie do wskaźników wieloletnich, na skale standaryzowane 100;15. W 2012 roku opublikowano również trzyletnie wskaźniki egzaminacyjne dla LO i techników w zakresie przedmiotów humanistycznych, matematyczno-przyrodniczych i oddzielnie polskiego i matematyki. Od początku 2013 roku Kalkulator EWD 100 może już analizować wyniki maturalne – na początek z matematyki na poziomie podstawowym.

Koncepcja badania trafności metody EWD

Jak zbadać trafność metody EWD? Przed konstruktorami modeli stoi wiele, trudnych do pokonania przeszkód. Na podstawie współczesnych ujęć trafności i wymagań stawianych testom używanym w modelowaniu EWD (por. Skorupiński, 2013; Pokropek, 2013) można wskazać na pięć powiązanych z sobą, ale dających się analitycznie rozdzielić aspektów badania trafności metody EWD. Schematycznie przedstawia je rysunek 1

WPROWADZENIE I NAJWAŻNIEJSZE WYNIKI UZYSKANE W BADANIU TRAFNOŚCI METODY EWD DLA GIMNAZJÓW

Rysunek 1. Aspekty badania trafności metody EWD



Pierwszy aspekt trafności wiąże się z faktem, że podstawowych danych do modelowania dostarczają testy egzaminacyjne. Zatem wartość tych danych jest warunkiem koniecznym, choć niewystarczającym dla trafności metody. W zakresie jakości danych testowych przede wszystkim stawia się pytanie o to, czy wykorzystane do modelowania EWD testy są dobrą miarą realizacji celów kształcenia. Wiąże się to z problemem zakresu dozwolonej generalizacji treściowej wyników testów. Test jest zawsze tylko próbką wymagań programowych, ale jeżeli próbka została dobrze dobrana, to wyniki testu można z dającym się zaakceptować błędem uogólniać na całość programu (przynajmniej tę część, która była podstawą budowy testu). Krytycy testów sądzą, że narzędzia te co najwyżej mówią nam, jak uczeń poradził sobie z danym zestawem zadań. Dodatkowo oskarżają testy o negatywny, zawężający ich wpływ na realizowany program. Ważne jest zatem, by sprawdzić, czy wyniki testów, które wykorzystujemy w modelowaniu EWD, pozwalają na uogólnienie. Jednak sam fakt, że dwa testy dobrze reprezentują wymagania programowe na poziomie kolejnych lat nauczania jest niewystarczający. Jeżeli wykorzystywane testy nie są zrównane pionowo, to należy sprawdzić, czy narzędzia wykorzystane do szacowania danego wskaźnika EWD mierzą ten sam lub przynajmniej bardzo zbliżony konstrukt. Jest to konieczne, by móc interpretować uzyskane wskaźniki jako miary względnego przyrostu osiągnięć szkolnych. Ważna jest również rzetelność pomiaru osiągnięć szkolnych i to, by jego wyniki nie były obciążone działaniem czynników niezwiązanych z poziomem osiągnięć. Ponadto skala, na której wyrażane są wyniki pomiaru, powinna być skalą co najmniej interwałową. Niektórzy badacze zwracają jeszcze uwagę na znaczenie porównywalności warunków, w których przeprowadzony jest pomiar osiągnięć oraz motywacji uczniów do rozwiązywania testów. Wymienione problemy zostaną podjęte w rozdziale 2.

Drugi aspekt trafności metody EWD wiąże się z właściwym statystycznym wymodelowaniem postępu w zakresie osiągnięć szkolnych uczniów w danej szkole. Jak właściwie uwzględnić w modelowaniu hierarchiczny charakter danych edukacyjnych: szkoła–oddział–uczeń? Jak uwzględnić zjawisko regresji ku średniej? Za pomocą jakiej statystyki opisywać przeciętny postęp w zakresie osiągnięć szkolnych uczniów w danej szkole? Jakie zmienne kontrolne uwzględniać w modelowaniu? Te problemy zostaną rozpatrzone w rozdziale 3. Ostatni z nich – zmienne kontrolne – prowadzi nas bezpośrednio do trzeciego aspektu trafności metody EWD.

Wiąże się on z przyczynową interpretacją wskaźników EWD. Przyjmujemy, że mówią one, jak efektywnie naucza dana szkoła. Metoda EWD operacjonalizuje efektywność nauczania w kategoriach względnego przyrostu osiągnięć uczniów. Jeżeli używamy testów dobrej jakości, to taka definicja kryterium efektywności jest do przyjęcia. Jednak definicja kryterium nie kończy problemu. Jeżeli wskaźniki EWD mają być użyteczne, to musimy mieć argumenty pozwalające oszacowany za pomocą modelu EWD postęp w zakresie osiągnięć szkolnych zasadnie przypisać działaniom szkoły, czyli wskaźniki efektywności nauczania powinny dawać się sensownie interpretować w kategoriach przyczynowo-skutkowych. Można popatrzeć na problem przez pryzmat metodologii eksperymentu. W kategoriach eksperymentalnych mamy do czynienia ze schematem wielogrupowym z pretestem i posttestem. Szkoła to jeden z warunków eksperymentalnych. Przyrost między pomiarem na wejściu i wyjściu jest miarą efektu oddziaływań danej szkoły. Sytuacja byłaby klarowna, gdyby pomiary były dokonywane na jednej skali, mielibyśmy grupę kontrolną, i – przede wszystkim – spełniony byłby wymóg randomizacji. Z dwoma pierwszymi problemami modele EWD radzą sobie relatywizując efekty (efekt szkoły X na tle średniego efektu we wszystkich szkołach), jednak problem braku losowego przydziału uczniów i nauczycieli do szkół jest niemożliwy do doskonałego rozwiązania. Zgodnie z metodologicznymi wymaganiami dla badań korelacyjnych możemy jedynie zmniejszyć ryzyko brania korelacji pozornych za relacje przyczynowo-skutkowe poprzez kontrolę zmiennych, z jednej strony związanych ze zmienną zależną, czyli postępem w zakresie osiągnięć szkolnych, z drugiej strony uwikłanych w procesy selekcji międzyszkolnych (lub międzyoddziałowych w wypadku wskaźników EWD dla oddziałów). Innymi słowy chodzi o problem skutecznej kontroli w modelowaniu EWD czynników, które mogłyby sprawiać, że przypisanie wskaźników EWD szkołom, miałoby charakter korelacji pozornej, czyli czy w modelach EWD skutecznie kontrolujemy czynniki niezależne od szkoły rzutujące na postępy uczniów. Przeanalizujemy pod tym kątem takie czynniki jak inteligencja ucznia i jego wiek (rozdział 4), status ekonomiczno-społeczny rodziny ucznia (rozdział 5), korepetycje i rodzinne wsparcie (rozdział 6), społeczno-gospodarcze charakterystyki gmin (rozdział 7)

oraz procesy rekrutacji do gimnazjów (rozdział 8). Niektóre z tych czynników mają dobrze udokumentowany wpływ na osiągnięcia (inteligencja, SES), znaczenie pozostałych dla osiągnięć szkolnych jest w świetle wyników badań problematyczne, ale są one na tyle często przywoływane w dyskusjach wokół polskich miar EWD jako obciążające szacunki, że z punktu widzenia trafności fasadowej (przekonania użytkowników o trafności) należy starannie zbadać ich wpływ w warunkach polskich.

W wypadku stwierdzenia znaczącego efektu danego czynnika dla oszacowania EWD będziemy postulować włączenie go w aplikacyjne modele EWD. Jednak należy pamiętać, że w sytuacji nielosowego przydziału uczniów do szkół – co ma miejsce właściwie w każdym systemie szkolnym, choć z różnym nasileniem – część wpływu wynikającego z działań szkół może być w procesie estymacji modelu przypisywana oddziaływaniu zmiennych kontrolnych. Poszerzanie zakresu czynników kontrolowanych w modelu, z jednej strony chroni więc przed „niesprawiedliwym” ocenianiem szkół, jednak z drugiej może utrudniać rozróżnianie od siebie szkół pracujących efektywnie i nieefektywnie. Jeżeli np. bardziej efektywnie nauczające szkoły „przyciągają” uczniów o wysokim SES, to włączenie miar SES do modelu szacowania EWD, zaniży szacunki dla tych szkół. Niestety, nie ma dobrego sposobu szacowania korelacji efektywności nauczania ze zmiennymi uwikłanymi w procesy selekcyjne na progu gimnazjum.

Kolejnym aspektem trafności metody EWD jest skonfrontowanie miar postępu z zewnętrznymi w stosunku do wyników nauczania kryteriami. O tym, czy interpretacja przyczynowo-skutkowa jest zasadna, możemy wnioskować nie tylko na podstawie skuteczności kontroli zmiennych powiązanych z selekcją, ale również patrząc, czy szkoły o wysokim EWD to takie, w których zgodnie z naszą normatywną wiedzą o funkcjonowaniu szkoły dobrze się dzieje. W naszym badaniu skupiliśmy uwagę na różnych aspektach zarządzania szkołą (rozdział 9) i pracy nauczyciela (rozdział 10). Jednak najważniejsze zewnętrzne kryterium trafności metody EWD to przyrost inteligencji uczniów (fragment rozdziału 4). Dzięki dwukrotnemu pomiarowi testem matryc Ravena możliwa była weryfikacja bardzo ryzykownej teoretycznie, ale w razie pozytywnej weryfikacji niezwykle ważnej dla trafności metody EWD, hipotezy o wpływie efektywnego nauczania na progres w inteligencji ogólnej.

Ostatni wyróżniony na schemacie aspekt trafności metody EWD, to konsekwencje stosowania tej metody z punktu widzenia ewaluacyjnej funkcji egzaminów zewnętrznych. Czy metoda służy skutecznej ewaluacji i autoewaluacji szkół, czy jest motorem doskonalenia nauczania? Czy stosowanie kryteriów egzaminacyjnych w ewaluacji wywołuje zjawisko nauczania pod test? W tym raporcie badawczym nie zajmujemy się powyższymi pytaniami. Odpowiedzi na nie poszukiwać będziemy w kolejnym przygotowywanym przez zespół raportcie.

Najważniejsze wyniki

Problem reprezentatywności badania

Dla oceny uzyskanych wyników kluczowy jest problem reprezentatywności badania. Dzięki danym z egzaminów możemy ocenić jakość próby nie tylko na podstawie procedury losowania i stopy realizacji, ale i ex post na podstawie porównania próby z populacją ze względu na rozkład kluczowej zmiennej zależnej. Zarówno w wylosowanej, jak i w zrealizowanej próbie szkół przeciętne wyniki sprawdzianu oraz egzaminu gimnazjalnego okazały się nieznacznie wyższe niż w populacji szkół uwzględnionych w operacie losowania. Rozkłady okazały się przesunięte w stronę wyników wysokich o około 1/10–1/15 odchylenia standardowego. Analizy wykazały też, że nie ma podstaw by sądzić, iż odmowy szkół związane były z poziomem osiągnięć szkolnych uczniów. Stanowi to dodatkowy argument przemawiający za odzwierciedleniem w próbie populacji, z której była losowana. Analiza zróżnicowania wyników sprawdzianu i egzaminu gimnazjalnego pokazała, że te właściwości zrealizowanej próby są tylko nieznacznie różne od parametrów populacji. Analizy pozwoliły stwierdzić, że wartości parametrów dwupoziomowych modeli estymowanych na danych z próby nie różnią się od wartości z analogicznych modeli populacyjnych, używanych do wyliczania wskaźników EWD, a co za tym idzie, że przyjęty sposób doboru próby zapewnił reprezentatywność zebranych wyników. Podsumowując, można powiedzieć, że pobrana próba była wysokiej jakości i w zadowalającym stopniu reprezentowała populację.

Jakość testów egzaminacyjnych

Fundamentem trafności metody EWD są dobre własności treściowe i psychometryczne używanych w modelowaniu testów osiągnięć szkolnych. W polskiej metodzie używane są – inaczej niż w wielu innych systemach EWD – testy wysokiej stawki, co jest dobrym prognozykiem, jeżeli chodzi o kompletność danych i problem motywacji testowej. Jednak używanie w modelowaniu danych z krajowych egzaminów nie gwarantuje rzetelności i trafności wyników. Dlatego analizy rozpoczęto od oceny jakości egzaminacyjnych narzędzi. Uzyskane wyniki wskazują, że egzaminy zewnętrzne wykorzystane do konstrukcji gimnazjalnych wskaźników EWD mają wystarczająco dobre właściwości, by mogły stanowić podstawę wyliczania trafnej miary EWD. Związki z zewnętrznymi kryteriami przemawiają za tym, że niosą one za sobą informację również o skuteczności realizacji szerzej rozumianych celów nauczania. Potwierdzono istnienie związków wyników egzaminacyjnych z innymi konstruktami, z którymi możemy oczekiwać, że powinny być one związane: z inteligencją i nauczycielskimi ocenami osiągnięć szkolnych. Dodatkowo wykazano ich związek z szerzej ujętymi efektami kształcenia, takimi jak motywacja, funkcjonowanie szkolne czy brak bezradności intelektualnej. Oznacza to, że wyniki

egzaminów gimnazjalnych są miarą nie tylko tego, jak uczeń poradził sobie z rozwiązaniem danego zestawu zadań, ale zdają sprawę z ogólnego poziomu wiadomości i umiejętności z danej dziedziny przedmiotowej, a także pośrednio niosą za sobą informację o skuteczności realizacji szerzej rozumianych celów nauczania. Wykazano, że sprawdzian mierzy na tyle bliski egzaminowi gimnazjalnemu konstrukt, że może być z powodzeniem wykorzystywany w modelach EWD jako miara uczniowskich zasobów na wejściu. Rzetelność testów egzaminacyjnych także okazała się wystarczająca dla tworzenia miar zagregowanych.

Warto jednak zaznaczyć, że poprawa pewnych własności polskich egzaminów zewnętrznych mogłaby przyczynić się do zwiększenia jakości wyliczanych wskaźników EWD. Rekomendacje w tym zakresie sprowadzają się do zwiększenia troski o jakość egzaminów na etapie ich planowania i konstruowania, wykorzystania w pełni możliwości badań pilotażowych w celu diagnozy zadań niespełniających założeń przyjętego modelu oraz zadań stronicznych, a także dostosowania trudności egzaminów tak, by ich wyniki miały stabilne rozkłady w kolejnych latach, które zgodnie z założeniem o badanej umiejętności będą zbliżone swoim kształtem do rozkładu normalnego. Ponadto dokumentacja podsumowująca proces tworzenia oraz właściwości każdego testu egzaminacyjnego powinna być publicznie dostępna niedługo po zakończonej sesji. Zwyczaj taki ułatwiłby badaczom korzystanie z wyników egzaminacyjnych w celach naukowych, a dla szerszego grona odbiorców mógłby stać się czynnikiem uwiarygodniającym wykorzystywane narzędzia. Poprawa testów egzaminacyjnych we wspomnianych zakresach podniosłaby nie tylko jakość wskaźników EWD, ale każdej miary opartej na tych testach: średniego wyniku szkoły, a w szczególności indywidualnych wyników uczniów.

Trafność interpretacji przyczynowej wskaźników EWD

Przegląd najważniejszych wyników pozwalających ocenić skuteczność modeli EWD w kontroli wpływu czynników pozaszkolnych na ocenę efektywności nauczania zaczniemy od analizy znaczenia wieku i inteligencji uczniów.

W zakresie wieku pytaliśmy o konsekwencje nieuwzględniania tej zmiennej w modelach EWD. Analizy wykazały brak zależności między wiekiem uczniów z głównej kohorty wiekowej a ich wynikami na egzaminie gimnazjalnym, choć zaobserwowano istotny efekt dla uczniów młodszych i starszych niż główna kohorta. W obu częściach egzaminu gimnazjalnego pierwsi z nich uzyskali wyniki przeciętnie o pół odchylenia standardowego wyższe niż pozostali uczniowie, natomiast drudzy – o około jedno odchylenie standardowe niższe. Jednakże w sytuacji kontroli uprzednich osiągnięć oraz innych zmiennych uwzględnianych w modelach EWD, efekty te traciły istotność statystyczną za wyjątkiem uczniów starszych niż główna kohorta w modelu dla przedmiotów humanistycznych. Pozwala to na sformułowanie wniosku, że

modele EWD dobrze radzą sobie z efektami bycia młodszym lub starszym niż główna kohorta wiekowa, choć dodanie tej informacji do modelu pozwoliłoby nieco precyzyjniej szacować EWD. Analizy wykazały także, że w modelach EWD efekt wieku biologicznego jest ujemny, choć niewielki. Jeśli szkoły znacząco różniłyby się ze względu na wiek uczniów z głównej kohorty wiekowej, to placówki, w których uczy się więcej młodszych uczniów, zyskiwałyby, a te z większą liczbą starszych uczniów traciłyby na wartościach wskaźników EWD. Obecnie mamy jednak do czynienia z niewielkim zróżnicowaniem międzyszkolnym na poziomie gimnazjum w zakresie wieku uczniów w głównej kohorcie i liczby uczniów młodszych lub starszych niż główna kohorta. Dlatego wprowadzenie do modeli EWD informacji o wieku uczniów nie zmieniłoby wskaźników w sposób znaczący. Problem uwzględnienia wieku w szacowaniu EWD nabierze znaczenia w przyszłości, gdy do gimnazjów trafią uczniowie, którzy rozpoczęli naukę szkolną w okresie przejściowym reformy obniżenia wieku szkolnego. Wtedy zróżnicowanie wiekowe uczniów stanie się znacznie większe, co sprawi, że istotne stanie się uwzględnienie danych na ten temat w modelowaniu.

Kolejnym czynnikiem analizowanym w raporcie jest inteligencja ucznia. Jeżeli relatywny postęp edukacyjny, jaki uczeń robi podczas nauki w gimnazjum, jest silnie zdeterminowany poziomem inteligencji ucznia i jednocześnie szkoły oraz oddziały wewnątrz szkół znacząco różniłyby się poziomem inteligencji uczniów, to brak w gimnazjalnych modelach EWD tej zmiennej prowadziłby do znaczącego obciążenia oszacowań EWD. Przeprowadzone analizy wykazały, że szkoły różnią się istotnie statystycznie poziomem inteligencji uczniów (mierzonym testem matryc Ravena). Sprawdzono zatem, czy wykorzystywane modele EWD pozwalają w wystarczającym stopniu kontrolować znaczenie inteligencji dla wyników egzaminacyjnych. Kontrolę tę w zadowalającym stopniu zapewnia uwzględnienie w modelach wyników sprawdzianu po klasie szóstej. Wprowadzenie w modelach EWD z uwzględnioną inteligencją jej efekt pozostaje istotny statystycznie, jednak wyjaśnia przede wszystkim wariancję na poziomie ucznia. Wobec tego uwzględnianie inteligencji w modelach zwiększyłoby precyzję szacowania EWD, jednak w średnich oszacowaniach EWD szkół nie zasłżyłby znaczące zmiany. Nadmienić należy, że bardziej rzetelny pomiar na sprawdzianie zapewniłby lepszą kontrolę inteligencji w modelach EWD. To, że efekt poziomu inteligencji pozostaje wciąż istotny może wynikać właśnie z niskiej rzetelności testu „na wejściu”.

Kolejny czynnik mogący rzutować na trafność metody EWD to status ekonomiczno-społeczny rodziny ucznia. Brak uwzględnienia tej informacji przy szacowaniu EWD jest często podnoszonym argumentem krytycznym. Czy zatem czynniki statusowe determinują postęp edukacyjny w trakcie nauki w gimnazjum? Czy gimnazja znacząco różnią się składem społecznym uczniów? W analizach wykorzystano kilka wskaźników statusu: poziom wykształcenia rodziców, indeks statusu społeczno-ekonomicznego (ISEI), indeks prestiżu zawodu

WPROWADZENIE I NAJWAŻNIEJSZE WYNIKI UZYSKANE W BADANIU TRAFNOŚCI METODY EWD DLA GIMNAZJÓW

wykonywanego przez rodziców (SIOPS), wskaźnik zasobności (głównie dobra kulturowe) rodziny ucznia oraz oddzielny wskaźnik wyposażenia gospodarstwa domowego. Największe różnice między gimnazjami stwierdzono dla miary ISEI i poziomu wykształcenia rodziców – około 20% wariancji całkowitej tych zmiennych to wariancja międzyszkolna. Równocześnie miary te bardzo silnie skorelowane są z wynikami egzaminacyjnymi. Determinacja statusowa wyników egzaminacyjnych jest szczególnie silna na poziomie szkół: zagregowane miary statusu wyjaśniają prawie 50% zmienności średnich wyników egzaminacyjnych gimnazjów. Czy uwzględnienie w modelach EWD uprzednich osiągnięć skutecznie kontroluje wpływ czynników statusowych? Dla wszystkich miar statusu obserwujemy istotne statystycznie zależności na poziomie indywidualnym. Jednak siła efektu jest dwu–czterokrotnie słabsza w porównaniu z modelem nieuwzględniającym wyników sprawdzianu. Co więcej, w modelach EWD zmienne statusowe wyjaśniają głównie wariancję na poziomie indywidualnym, w niewielkim stopniu wariancję międzyszkolną wyników egzaminacyjnych. Oszacowane z wykorzystaniem zmiennych statusowych kontekstowe wskaźniki EWD dla szkół korelują ze zwykłymi miarami EWD na poziomie 0,98–0,99. Oznacza to, że brak w modelu EWD zmiennych statusowych w bardzo niewielkim stopniu obciąża oszacowania wskaźników, a tym samym ma marginalne znaczenie dla trafności metody.

Czynnikiem silnie powiązanim ze statusem społecznym rodziny są korepetycje, czy szerzej mówiąc, rodzinne wsparcie w szkolnej nauce. Wpływ tego czynnika to najczęściej i z największym przekonaniem podnoszony argument krytyczny wobec trafności metody EWD nie uwzględniającej tej informacji. Przeprowadzone analizy wykazały, że znacznie częściej gimnazjalista korzysta z dodatkowego wsparcia w nauce z przedmiotów matematyczno-przyrodniczych niż z przedmiotów humanistycznych. Odsetek uczniów korzystających z korepetycji z przedmiotów humanistycznych jest niewielki, z przedmiotów ścisłych co szósty gimnazjalista deklaruje wzięcie choć raz w trakcie nauki w gimnazjum korepetycji. Obserwujemy większe międzyszkolne zróżnicowanie korzystania z bezpłatnego wsparcia z przedmiotów matematyczno-przyrodniczych. Analizy ujawniły negatywną zależność pomiędzy korzystaniem z dodatkowego wsparcia w nauce poza lekcjami w szkole a osiąganymi wynikami egzaminacyjnymi. Uczniowie, którzy zadeklarowali korzystanie z korepetycji z przedmiotów matematyczno-przyrodniczych, uzyskują w tej części egzaminu gimnazjalnego wyniki niższe niż uczniowie, którzy z korepetycji nie korzystali. Związek ten utrzymuje się w modelu EWD, a także przy kontroli charakterystyk ucznia i jego rodziny, istotnych zarówno dla uczęszczania na korepetycje, jak i dla wyników egzaminacyjnych. Podobne rezultaty otrzymano w przypadku korzystania z bezpłatnego wsparcia w nauce – wiąże się ono z niższymi osiągnięciami na egzaminie. Schemat badania nie pozwala na wnioskowanie, że jest to zależność przyczynowo-skutkowa (lecz jednocześnie jej nie wyklucza). Widoczne jest zatem, że

wbrew potocznej opinii nie możemy powiedzieć, że szkoły, w których uczniowie otrzymują dodatkowe wsparcie w nauce (płatne lub bezpłatne), zyskują we wskaźnikach EWD. Widoczne jest także, że w modelach EWD negatywna zależność między korzystaniem z dodatkowego wsparcia w nauce pozostaje istotna, jednak znacznie słabnie. To bardzo ważny rezultat, gdyż dane dotyczące korzystania przez uczniów ze wsparcia w nauce nie są gromadzone w systemie egzaminów zewnętrznych i z dużym prawdopodobieństwem nigdy się to nie zmieni. Nie ma więc możliwości uwzględniania ich w modelach populacyjnych.

Kolejny problem, który można ulokować w ramach badania wewnętrznej trafności metody EWD to wpływ lokalnych czynników społeczno-gospodarczych. Czy takie czynniki jak zamożność gminy, infrastruktura kulturalna czy wielkość wydatków na oświatę, wyjaśniają egzaminacyjne sukcesy szkoły? Gdyby wyniki końcowe, a przede wszystkim miary postępu czynionego przez uczniów w danym gimnazjum, były determinowane przez czynniki społeczno-gospodarcze, trafny byłoby zarzut, że to ekonomia, a nie starania szkół decydują o sukcesach egzaminacyjnych. Przyjrzymy się zatem, czy układ korelacji potwierdza ekonomiczny determinizm. Czy brak w używanych modelach EWD lokalnych charakterystyk społeczno-gospodarczych znacząco obciąża miary EWD? Społeczno-gospodarcze zmienne, które okazały się być istotnymi predyktorami wyników egzaminu gimnazjalnego, to wielkość gminy, udział dzieci 3–5-letnich w opiece przedszkolnej, udział zarejestrowanych bezrobotnych w populacji osób w wieku produkcyjnym oraz liczba podmiotów zaawansowanej technologii (HT od ang. *high technology*) zrelatywizowana liczbą mieszkańców. Dla osiągnięć w części humanistycznej egzaminu gimnazjalnego istotne okazały się dodatkowo wydatki na oświatę i wychowanie na mieszkańca gminy. Po wprowadzeniu zmiennych, które uwzględniane są w modelach EWD, istotnymi predyktorami w obu modelach pozostał jedynie udział zarejestrowanych bezrobotnych w populacji osób w wieku produkcyjnym. Dodatkowo w modelu dla części matematyczno-przyrodniczej istotna pozostała wielkość gminy oraz liczba podmiotów HT zrelatywizowana liczbą mieszkańców. Zmienne te oddziałują na oszacowania osiągnięć szkolnych również przy kontroli zmiennych kontekstowych z poziomu indywidualnego (IQ, SES) i z tego powodu należy rozważyć uwzględnienie tych zmiennych w modelach EWD w celu odciążenia oszacowań efektywności szkół z lokalnych czynników o charakterze pozaszkolnym. Dane te są dostępne i mogą być wykorzystane w modelowaniu. Jednak problemem, który wymagałby rozstrzygnięcia w kontekście wpływu czynników lokalnych, jest przestrzenna precyzja danych dostarczanych przez GUS. Może mieć to znaczenie dla dużych, zróżnicowanych wewnętrznie gmin. Jeśli np. w dużym mieście występuje znaczne zróżnicowanie w poziomie bezrobocia pomiędzy dzielnicami, to brak jest powszechnie dostępnych danych, które pozwoliłyby na przypisanie szkołom poziomu bezrobocia właściwego dla dzielnicy w której się znajdują. Brak takiej

możliwości może utrudniać ocenę efektywności szkół w dużych, zróżnicowanych jednostkach administracyjnych.

Z punktu widzenia interpretacji przyczynowo-skutkowej wskaźników EWD ważnym problemem jest związek selekcji do gimnazjów z wynikami egzaminacyjnymi. Czy oszacowana na podstawie danych dotyczących rekrutacji uczniów spoza rejonu selekcyjność gimnazjum koreluje z EWD szkoły? Innymi słowy, czy „spijanie śmietanki” jest gwarancją sukcesu w miarach EWD? W książce przedstawiono wyniki analiz mających na celu zbadanie relacji między selekcją przeprowadzaną przez gimnazja w momencie naboru do klas pierwszych, a osiągnięciami uczniów mierzonymi wynikami egzaminacyjnymi oraz wskaźnikami EWD. Typologia selekcji w szkołach została opracowana na podstawie wskaźników określających odsetek uczniów spoza rejonu w szkole oraz wskaźnik biorący pod uwagę liczbę osób spoza rejonu, które chciały do szkoły się dostać. By wyeliminować losowy wpływ rekrutacji z danego roku, wskaźniki zostały opracowane na podstawie danych o rekrutacji do klas pierwszych badanych szkół z lat 2007–2012. Otrzymane wyniki wskazują, że szkoły zaklasyfikowane do grupy silnie selekcyjnych mają wyraźnie wyższe średnie wyniki egzaminów końcowych. Przewaga ta okazała się bardzo podobna w obu częściach egzaminu gimnazjalnego. W przypadku części matematyczno-przyrodniczej jest to około 13 punktów, natomiast w części humanistycznej około 11 punktów. Uwzględnienie w analizach zmiennych wykorzystywanych do modelowania EWD, czyli wyników na sprawdzianie, informacji o dysleksji na egzaminach oraz płci, zmienia znacząco rezultaty analiz. W modelach EWD efekt czynnika selekcji okazał się nieistotny statystycznie. Zatem możemy uznać, że w przypadku modeli biorących pod uwagę wcześniejsze osiągnięcia szkolne, szkoły różniące się od siebie siłą selekcji nie różnią się wartościami wskaźników EWD. Możemy zatem dojść do wniosku, że modele EWD liczone dla gimnazjów w dostateczny sposób kontrolują wpływ praktyk selekcyjnych odbywających się na progu gimnazjum.

Trafność kryterialna metody

Jak sądzimy najmocniejszy argument na rzecz zewnętrznej trafności metody EWD zdobyliśmy dzięki uwzględnieniu w badaniu danych o inteligencji uczniów. Pomiaru inteligencji dokonano dwukrotnie: w pierwszej i trzeciej klasie gimnazjalnej. Podłużny charakter pomiarów pozwolił zweryfikować niezwykle ryzykowną teoretycznie, ale pociągającą poznawczo hipotezę: w szkołach o wysokim EWD nie tylko uczniowie robią większy niż przeciętnie postęp w zakresie przyswajania wiadomości i umiejętności szkolnych, ale szybciej im przyrasta iloraz inteligencji. Jest to hipoteza tym bardziej ryzykowna, że do pomiaru inteligencji użyto testu matrycy Ravena, a więc testu inteligencji płynnej.

Testując hipotezę w pierwszym kroku, stwierdzono, że szkoły różnią się średnią wielkością przyrostów inteligencji swoich uczniów (podział na szkoły wyjaśnia 14,5% tego zróżnicowania, model analogiczny do EWD). W drugim, kluczowym etapie analiz potwierdzono istnienie pozytywnej, znaczącej statystycznie korelacji pomiędzy wskaźnikami EWD szkół a średnimi przyrostami inteligencji ich uczniów w toku nauki w gimnazjum. Oznacza to, że wraz ze wzrostem EWD szkoły rosną średnie przyrosty inteligencji uczniów. Związek ten jest silniejszy w wypadku EWD w obszarze matematyczno-przyrodniczym. Uzyskany wynik stanowi silny argument przemawiający za zewnętrzną trafnością metody EWD. Powiązanie wskaźników EWD z przyrostami inteligencji sugeruje, że nie wskazują one jedynie na to, na ile dobrze szkoła radzi sobie z przygotowaniem uczniów do egzaminów zewnętrznych, lecz są wyrazem zachodzących w szkole procesów, które mają znaczenie dla rozwoju poznawczego uczniów.

Kolejny test trafności metody EWD w jej zewnętrznym aspekcie dotyczy związków EWD ze stylami zarządzania szkołą. Czy szkoły uzyskujące wysokie wskaźniki EWD są zarządzane w sposób, który budzi nasz niepokój czy też odwrotnie, odnajdujemy w nich przejawy stylów zarządzania, adekwatnych do edukacyjnej misji szkoły? Uzyskane rezultaty są niejednoznaczne i świadczą o zróżnicowanym wpływie różnych wymiarów stylów zarządzania szkołą na wyniki egzaminacyjne. Partycypacja i zmienne jej towarzyszące (zaufanie do współpracowników, zaufanie do przełożonego, możliwość uzyskania informacji zwrotnej o swojej pracy, możliwość samorozwoju, identyfikacja z instytucją) nie mają wpływu na wyniki egzaminacyjne w przypadku części matematyczno-przyrodniczej. W przypadku części humanistycznej egzaminu w modelu, w którym analizowano tylko wynik egzaminu na zakończenie gimnazjum, uzyskano negatywny efekt dla skali „informacja zwrotna” oraz pozytywny dla skali „zaufanie do współpracowników”. Drugim z analizowanych w rozdziale wymiarów funkcjonowania szkoły była kultura organizacyjna. W obu częściach egzaminu istotne, pozytywne efekty uzyskano dla rynkowego typu kultury – nastawienie na rezultaty, na rywalizację z innymi szkołami i motywowanie pracowników do realizacji celów, zorientowanie w działaniu na otoczenie zewnętrzne. Pozytywne efekty daje także klanowy typ kultury. W modelu EWD staje się on istotny w przypadku obu części egzaminu gimnazjalnego. Organizacja posiadająca ten typ kultury jest nastawiona na współpracę między pracownikami. Więzy między nimi są bardzo silne, mogą nabrać charakteru osobistego, prawie rodzinnego, a dyrektor często przyjmuje rolę mentora. Przedstawione w rozdziale analizy wskazują na związki stylów zarządzania ze wskaźnikami EWD. Niektóre z nich należy uznać za pożądane z punktu widzenia traktowania EWD jako miary efektywności pracy szkoły. Wykazano, że istnieje pozytywny efekt tych elementów stylów zarządzania, które nastawione są na pracowników, ich rozwój oraz utrzymywanie w szkole

WPROWADZENIE I NAJWAŻNIEJSZE WYNIKI UZYSKANE W BADANIU TRAFNOŚCI METODY EWD DLA GIMNAZJÓW

poczucia wspólnoty. Z drugiej strony, interesujący z punktu widzenia trafności samego wskaźnika EWD jest wpływ rynkowego typu kultury organizacyjnej na wyniki i EWD. Ten typ kultury, charakteryzujący się dużą konkurencyjnością, jest relatywnie mocno związany z surowymi wynikami egzaminu, natomiast w przypadku analiz z modelem analogicznym do EWD, efekt staje się słabszy o połowę. Tego rodzaju rezultat można traktować jako pośredni dowód na trafność wskaźnika EWD.

Ostatni obszar peregrynacji badawczych prowadzonych w poszukiwaniu potwierdzenia lub falsyfikacji trafności zewnętrznej metody EWD, to efekty związane z nauczycielem. Uzyskane oszacowania wyjaśnionej wariancji wyników egzaminacyjnych uczniów przez cechy nauczyciela są spójne z dostępnymi dowodami w literaturze przedmiotu. Wariancja wyników egzaminów, która może być przypisana nauczycielowi, wyniosła w raportowanych analizach 14% (matematyka) i 12% (kompetencje językowe). W przypadku uwzględnienia wyniku sprawdzianu, płci, dysleksji, statusu ekonomiczno-społecznego rodziny ucznia, uzyskano wynik 6% (matematyka) i 9% (kompetencje językowe). Przy uwzględnieniu dodatkowo miary inteligencji i kontroli składu oddziału pod kątem kompozycji społecznej i wcześniejszych osiągnięć uzyskano odpowiednio 7% i 9%. Odnotowano pozytywny związek wykształcenia nauczycieli z wynikami uczniów w nauczaniu matematyki. Związek ten staje się jednak nieistotny przy kontroli inteligencji i statusu społeczno-ekonomicznego rodziny ucznia. Doskonalenie zawodowe okazało się być nieistotnym czynnikiem w wyjaśnianiu wyników uczniów. Ciekawym efektem okazał się pozytywny wpływ płci nauczyciela na wyniki egzaminu z części polonistycznej. W nauczaniu języka polskiego nauczycielki przyczyniają się bardziej do podniesienia wyników uczniów niż nauczyciele. Efekt ten wymaga jednak dalszej eksploracji na większych próbach, gdzie odsetek nauczycieli-mężczyzn jest wyższy, niż w próbie tu analizowanej.

W zakresie stylu pracy nauczyciela należy odnotować, że uwzględnienie w modelu analizy informacji o wcześniejszych osiągnięciach edukacyjnych uczniów, płci, statusie ekonomiczno-społecznym i inteligencji w większości wypadków znoszą istotność lub osłabiają zaobserwowane efekty. Jedynym stabilnym i o jednoznacznej interpretacji efektem związanym ze sposobem prowadzenia lekcji jest utrzymanie dyscypliny. Nauczyciele potrafiący utrzymać dyscyplinę na lekcjach, zaciekać tematem uczniów, zająć ich uwagę, przyczyniają się do podniesienia wyników uczniów. Podsumowując, literatura dotycząca efektu pracy nauczyciela nie dostarcza prostych odpowiedzi na pytanie za pomocą jakich charakterystyk samych nauczycieli, jak i charakterystyk ich stylu pracy można wyjaśnić zróżnicowanie efektywności nauczania. Również w naszych badaniach nie wykryto zbyt wielu znaczących korelacji. W tym wypadku można uznać, że badanie nie dostarczyło ważnych argumentów na rzecz trafności metody EWD, ale też nie wykazało, że wysokie

wartości wskaźników EWD wiąże się z jakimiś niepokojącymi cechami lub sposobami pracy nauczyciela.

Reasumując wszystkie uzyskane wyniki, można stwierdzić, że w wielu aspektach metoda EWD wykazała swą trafność, równocześnie badanie wskazało kierunki doskonalenia modeli EWD. Jako zespół możemy pozytywnie rekomendować użycie metody EWD w analizach wyników egzaminacyjnych na potrzeby ewaluacji pracy gimnazjum w zakresie efektywności nauczania. Jednak wątpliwości co do przyczynowo-skutkowej interpretacji wskaźników EWD, choć osłabione, nie mogą być całkowicie oddalone. Należy zatem pamiętać, że ważnym remedium na problemy z przyczynową interpretacją wskaźników EWD jest analiza trendów w czasie. W wypadku gimnazjów szkoły dysponują już wynikami z ubiegłych lat i analiza ich dynamiki z pewnością jest w większym stopniu wolna od tych interpretacyjnych wątpliwości. O ile w otoczeniu szkoły nie zaszły jakieś rewolucyjne zmiany, to zmiany w czasie wskaźników EWD mogą być ze znaczną pewnością przypisywane działaniom szkoły.

Podziękowania i wkład poszczególnych autorów

Choć tekst raportu badawczego napisało siedmiu autorów, to książka nie mogłaby powstać bez solidarnego, wieloletniego wysiłku całego zespołu badawczego zajmującego się rozwojem metody EWD w Polsce. Autorzy szczególnie pragną podziękować Ewie Stożek, Maciejowi Jakubowskiemu, Arturowi Pokropkowi, Zofii Lisieckiej, Filipowi Kulonowi, Annie Rappe, Krystynie Szmigel, Cezaremu Kacprzakowi, Klarze Sołtan-Kościeleckiej i Danucie Ostrochulskiej. Bez ich zaangażowania nie udałooby się przygotować i przeprowadzić badań.

Praca ma charakter spójnej monografii naukowej. Z tego powodu nie zamieszczono przy poszczególnych rozdziałach nazwisk autorów. Jednak dla udokumentowania ich wkładu w tym miejscu wymienione zostaną ich nazwiska. Rozdział 1. *Organizacja i metodologia badania podłużnego w gimnazjum* napisali Anna Hawrot, Grzegorz Humenny, Przemysław Majkut i Tomasz Żółtak, rozdział 2. *Trafność testów egzaminacyjnych* napisali Aleksandra Jasińska i Tomasz Żółtak, rozdział 3. *Gimnazjalne modele edukacyjnej wartości dodanej* napisał Tomasz Żółtak, rozdział 4. *Wiek i inteligencja ucznia* napisały Aleksandra Jasińska i Anna Hawrot, rozdział 5. *Status społeczny rodziny ucznia* napisali Roman Dolata i Grzegorz Humenny, rozdział 6. *Korepetycje i rodzinne wsparcie edukacyjne* napisały Anna Hawrot i Aleksandra Jasińska, rozdział 7. *Spółeczno-gospodarcze charakterystyki gmin* napisał Grzegorz Humenny, rozdział 8. *Selekcja uczniów na progu gimnazjum* napisał Przemysław Majkut, rozdział 9. *Style zarządzania szkołą* napisał Przemysław Majkut, rozdział 10. *Cechy nauczyciela* napisał Maciej Koniewski, w końcu *Wprowadzenie* napisał Roman Dolata.

Literatura

Dolata, R. (2006a). Edukacyjna wartość dodana w komunikowaniu wyników egzaminów zewnętrznych. Egzamin. *Biuletyn Badawczy CKE*, 8, s. 9–20.

Dolata, R. (2006b). Efektywność nauczania w gimnazjach miasta X. Analiza edukacyjnej wartości dodanej. Egzamin. *Biuletyn Badawczy CKE*, 8, s. 28–37.

Dolata, R. (red. naukowa) (2008). *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania*. Warszawa: Centralna Komisja Egzaminacyjna.

Organisation for Economic Co-operation and Development. (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paris: OECD.

Skorupiński, P. M. (2013). Modele trafności pomiaru. [w:] Karwowski, M. (red.) *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne. Trafność wskaźników edukacyjnej wartości dodanej dla szkół maturalnych*. Warszawa: Wydawnictwo IFiS PAN.

Pokropek, A. (2013). Trafność testów egzaminacyjnych. [w:] Karwowski, M. (red.) *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne. Trafność wskaźników edukacyjnej wartości dodanej dla szkół maturalnych*. Warszawa: Wydawnictwo IFiS PAN.

Rozdział 1

Organizacja i metodologia badania podłużnego w gimnazjum

Zasadniczym celem badania podłużnego w gimnazjum było zebranie danych pozwalających określić trafność metody EWD. Dodatkowo, informacje zebrane w ramach badania pozwolą pogłębić wiedzę o związkach czynników indywidualnych, środowiskowych i szkolnych z wynikami nauczania. W ramach badania monitorowano także opinie nauczycieli i dyrektorów gimnazjów na temat systemu egzaminów zewnętrznych. Dane z tego ostatniego wątku zostaną przeanalizowane w oddzielnym raporcie.

Wszystkie czynności badawcze zostały zaplanowane przez zespół EWD. Członkowie zespołu odpowiadają za przygotowanie koncepcji badania, dobór i przygotowanie odpowiednich narzędzi badawczych, określenie dokładnej procedury każdego z etapów badania oraz wylosowanie próby badawczej. Należy zaznaczyć, że w trakcie trwania badania podłużnego zespół EWD był afiliowany przy Centralnej Komisji Egzaminacyjnej. Z tego względu w opisie badania i procedur mogą pojawiać się odniesienia do tej właśnie instytucji jako zlecającej badania. Od września 2012 roku projekt kontynuowany jest w Instytucie Badań Edukacyjnych w Warszawie.

Realizacja badania w terenie, ze względu na swą skalę, powierzona została firmom zewnętrznym. Wykonawców wyłoniono w ramach przetargów publicznych, zgodnie z obowiązującymi przepisami prawa. Pierwszy etap badania został zrealizowany przez konsorcjum Pracownia Badań Społecznych DGA i CBOS, drugi – przez InterActive Agencja Komunikacji Marketingowej, natomiast trzeci przez firmę EU-CONSULT Sp. z o.o.

Jednym z głównych problemów w badaniach podłużnych jest utrata części respondentów w toku badania. Z góry należy założyć, że z powodu różnego rodzaju zdarzeń losowych odsetek respondentów, do których uda się dotrzeć z etapu na etap będzie coraz niższy. Istotne jest, by ograniczyć skalę tego spadku. W tym celu podjęto wiele działań. Po pierwsze, określono wysokie minimalne progi realizacji badania (minimalne odsetki respondentów z próby badawczej, których należy przebadać, by badanie zostało uznane za należyście zrealizowane), a wynagrodzenie wykonawcy uzależniono od osiągniętych progów realizacji. Miało to zachęcić wykonawców do podejmowania wysiłku docierania do każdego respondenta. Po drugie, wyznaczono koordynatora badania ze strony zespołu badawczego, który sprawował stały nadzór nad jego realizacją, wspierał wykonawcę, jak również osobiście kontaktował się ze szkołami, w których pojawiły się problemy realizacyjne.

W tym rozdziale przybliżona zostanie ogólna koncepcja badania, scharakteryzowane zostaną badane populacje, sposób losowania próby, stopa jej realizacji oraz reprezentatywność. Ponadto opisany zostanie przyjęty schemat badania, przeprowadzone pomiary, jak również przebieg realizacji kolejnych jego etapów.

1.1. Koncepcja badania

Badanie podłużne w gimnazjum miało charakter panelu. Dzięki obserwacji tej samej próby uczniów na trzech etapach badania, od momentu rozpoczęcia przez nich nauki w gimnazjum, aż do egzaminu gimnazjalnego, możliwe jest precyzyjne badanie osiągnięć uczniowskich i czynników z nimi związanych.

Oprócz danych uzyskanych od respondentów podczas badania, z baz egzaminacyjnych pozyskano informacje o wynikach sprawdzianu po klasie szóstej i egzaminu gimnazjalnego uczniów objętych badaniem (dane populacyjne pozyskiwane są co roku przez zespół na potrzeby modelowania wskaźników EWD). Przyłączenie wyników egzaminacyjnych do zbiorów danych z badań było kluczowe dla weryfikacji trafności wskaźników EWD. Tylko dzięki połączeniu tych danych można sprawdzić, jak zmieniają się modele EWD przy uwzględnieniu danych powiązanych z osiągnięciami uczniów.

Badanie objęło pięć grup respondentów. Pierwszą z nich byli uczniowie. Uczestniczyli oni w każdym z etapów badania. Oprócz nich badaniem objęto rodziców uczniów, nauczycieli wybranych przedmiotów oraz dyrektorów szkół, do których uczęszczali uczniowie będący zasadniczą grupą respondentów. Dokładną strukturę badania, wraz z opisem populacji badanych na każdym z etapów, ilustruje tabela 1.

Etap I przeprowadzono zimą i wiosną 2010 roku. Rozpoczął on cały cykl badawczy i objął uczniów z wybranych klas pierwszych z wylosowanych gimnazjów, ich rodziców, nauczycieli oraz dyrektorów. Miał na celu zebranie podstawowych informacji dotyczących inteligencji ucznia oraz pozycji społecznej jego rodziców. Są to dwie zmienne, które w literaturze tematu wskazuje się jako istotne wyznaczniki osiągnięć szkolnych ucznia. Z punktu widzenia trafności wskaźników EWD kluczowe jest sprawdzenie, w jakim stopniu zmieniają się wskaźniki EWD, jeśli przy ich wyliczaniu będziemy brali pod uwagę inteligencję ucznia oraz pozycję społeczną jego rodziców.

Jako podstawy przygotowania wskaźników, pozwalających mierzyć pozycję społeczną rodziców, użyto klasyfikacji ISCO w wersji 2008 (The International Standard Classification Of Occupations – Międzynarodowy Standard Klasyfikacji Zawodów), która pozwala, przy zadaniu dodatkowych pytań, na stworzenie na jej podstawie skali ISEI (The Socio-economic Index of Occupational Status – socjoekonomiczny wskaźnik statusu zawodów; Ganzeboom, De Graaf i Treiman, 1992). Skala ta jest wykorzystywana między innymi w badaniach PISA (OECD, 2009). Zebrano także informacje dotyczące wykształcenia rodziców. Dodatkowo badano tutaj strukturę rodziny ucznia, kapitał kulturowy (zoperacjonalizowany jako liczba książek w domu) i ekonomiczny rodziny (indeks posiadanych dóbr). W celu zminimalizowania ilości braków danych, pytania dotyczące pozycji społecznej zostały zadane zarówno rodzicom, jak i uczniom (odpowiedzi uczniowskie są włączane w analizę jedynie w sytuacji, gdy dla obojga rodziców występują braki danych).

Tabela 1. Schemat badania podłużnego w gimnazjum

etap	grupa respondentów	pomiary
etap 1 (I–III 2010) I klasa gimnazjum	uczniowie	inteligencja, wcześniejsze doświadczenia edukacyjne, motywwy wyboru szkoły, aspiracje edukacyjne, pozycja społeczna rodziców
	rodzice	środowisko społeczne, pozycja społeczna, aspiracje edukacyjne w stosunku do dziecka, wcześniejsze doświadczenia szkolne ucznia, motywwy wyboru gimnazjum, opinie o egzaminach zewnętrznych
	nauczyciele	opinie o egzaminach zewnętrznych i EWD, wykorzystanie wyników egzaminów zewnętrznych w pracy dydaktycznej
	dyrektorzy	opinie o egzaminach zewnętrznych i EWD, wykorzystanie wyników egzaminów zewnętrznych w pracy dydaktycznej, selekcja uczniów do szkoły
etap II (IX–XI 2011) III klasa gimnazjum	uczniowie	osiągnięcia i umiejętności szkolne uczniów w zakresie języka polskiego (świadomości językowej oraz czytania) i matematyki.
etap III (IV–VI 2012) III klasa gimnazjum	uczniowie	inteligencja, motywacja do nauki, integracja uczniów ze szkołą, zaufanie do instytucji publicznych, aspiracje edukacyjne, sposób przygotowania się do egzaminu gimnazjalnego, korepetycje, socjometria, styl nauczania, bezradność intelektualna, stopień samokontroli
	rodzice	aspiracje edukacyjne w stosunku do dziecka, uczestnictwo w życiu szkoły
	nauczyciele	styl zarządzania, kultura organizacyjna, styl przywództwa, opinie o egzaminach zewnętrznych i EWD, styl nauczania, uczestnictwo w różnych formach doskonalenia
	dyrektorzy	styl zarządzania, kultura organizacyjna, styl przywództwa, opinie o egzaminach zewnętrznych i EWD, stosowanie selekcji uczniów do szkoły, metody selekcji uczniów do klas oraz przydziału nauczycieli do klas

Do pomiaru inteligencji zdecydowano się użyć testu matryc Ravena w wersji standard – forma klasyczna (Jaworowska i Szustrowa, 1991a). Pozwala on mierzyć inteligencję płynną u dzieci, młodzieży i dorosłych.

Ponadto zebrano informacje dotyczące wcześniejszych doświadczeń szkolnych ucznia (m.in. uczęszczania do przedszkola), jego aspiracji edukacyjnych oraz aspiracji edukacyjnych rodziców względem dziecka. Przedmiotem pomiaru były także opinie rodziców, nauczycieli i dyrektorów o systemie egzaminów zewnętrznych. Dwie ostatnie grupy pytano także o opinie na temat EWD.

W II etapie badania, jesienią 2011 roku, przeprowadzono pomiar osiągnięć szkolnych. Uczestniczyli w nim uczniowie, którzy zostali objęci pierwszym etapem badania. W czasie trwania II etapu rozpoczynali oni naukę w klasie trzeciej. Wypełniali dwa testy: czytania oraz matematyki (każdy z testów miał dwie wersje, przy czym część zadań – tzw. zadania kotwiczące – występowała w obu wersjach). Zebrane dane pozwoliły na sprawdzenie, w jakim stopniu zmieniłby się wskaźnik EWD, jeśli mielibyśmy dodatkową informację o osiągnięciach uczniów przed etapem intensywnych przygotowań do egzaminu gimnazjalnego.

Testy osiągnięć zostały opracowane przez zespół EWD na podstawie zadań z wcześniejszych egzaminów zewnętrznych: sprawdzianu, egzaminu gimnazjalnego oraz odtajnionych zadań wykorzystywanych w badaniach PISA. W zakresie czytania skupiono się na następujących umiejętnościach uczniów:

- wyszukiwanie informacji – znalezienie w tekście informacji,
- interpretacja – zrozumienie tekstu, od zakreślenia zakresu tematycznego do określenia znaczenia tekstu, wraz z odniesieniem się do zawartych w nim przeciwstawnych idei lub niejednoznaczności,
- refleksja i ocena – odniesienie znaczeń zawartych w tekście do szerszej wiedzy, także umiejętności rozumowania hipotetycznego i uzasadnienia swojego stanowiska.

W przypadku matematyki skupiono się na następujących umiejętnościach:

- przestrzeń i kształt – znajomość figur geometrycznych i rozumienie relacji i związków między obiektami w przestrzeni,
- zmiana i związki – relacje matematyczne między zbiorami (zawieranie, podzielność itp.), reprezentowane symbolicznie, algebraicznie, graficznie lub tabelarycznie,
- ilość – odniesienie do liczb jako atrybutów rzeczy w otaczającym nas świecie (kursy walut, objętości, czas), także szacowania i przybliżanie.

Ostateczne przygotowanie zeszytów testowych było poprzedzone pilotażem, w ramach którego przeprowadzono badanie na 179 uczniach klas trzecich z 10 gimnazjów. Na podstawie wyników pilotażu usunięto te zadania, które miały niezadowalające własności pomiarowe.

W ramach II etapu badania każdy uczeń rozwiązywał jedną z dwóch wersji testu czytania i jedną z dwóch wersji testu matematycznego. Dzięki puli zadań kotwiczących (jednakowych w obu wersjach) możliwe było wyrażenie wyników obu wersji danego testu na wspólnej skali (dzięki zastosowaniu odpowiednich metod statystycznej

analizy danych). Pozwoliło to na wspólną analizę wszystkich zadań, zwiększając jednocześnie różnorodność zadań wykorzystywanych w teście.

W przypadku czytania, zeszyt w wersji pierwszej zawierał 13 niepowtarzalnych zadań, natomiast w drugiej – 14. Testy zawierały także 10 zadań kotwiczących. Obie wersje testu matematycznego zawierały po 10 niepowtarzalnych zadań oraz 6 zadań kotwiczących.

Etap III, trwający od kwietnia do czerwca 2012 roku, ponownie objął wszystkie grupy respondentów. Jego realizacja rozpoczęła się zaraz po zakończeniu egzaminu gimnazjalnego. W jego toku zebrano dane na temat szeregu wyznaczników osiągnięć szkolnych. W przypadku uczniów, dokonano powtórnego pomiaru inteligencji. Ze względu na niewystarczającą trudność testu matryc Ravena w wersji standard – forma klasyczna (Jaworowska, Szustrowa, 1991a), już w pierwszym etapie zdecydowano się na wykorzystanie także testu matryc Ravena w wersji „dla zaawansowanych” (Jaworowska, Szustrowa, 1991b). Umożliwiło to szacowanie przyrostów inteligencji uczniów podczas nauki w gimnazjum. Respondenci wypełniali najtrudniejsze serie testu matryc Ravena w wersji standard – forma klasyczna (serie C, D, E) oraz pozycje od 1 do 24 II serii testu w wersji dla zaawansowanych (Jaworowska, Szustrowa, 1991b). Dodatkowo uczniowie wypełniali kwestionariusz, w którym znajdowały się pytania oraz całe narzędzia do badania między innymi stopnia integracji szkolnej, motywacji do nauki, zaufania do instytucji publicznych, sposobów przygotowania do egzaminu gimnazjalnego, aspiracji edukacyjnych, korzystania z korepetycji oraz bezpłatnej pomocy w nauce, poziomu bezradności intelektualnej w zakresie języka polskiego i matematyki oraz poziomu samokontroli. W kwestionariuszach skierowanych do nauczycieli pytano także o style nauczania. Szczególną uwagę zwrócono także na funkcjonowanie szkół, głównie na sposób zarządzania nimi. W tym kontekście badano partycypację nauczycieli i rodziców w funkcjonowaniu szkoły, styl przywództwa w szkole oraz jej kulturę organizacyjną. Ponadto w kwestionariuszu kierowanym do dyrektorów powtórzono pytania dotyczące selekcji uczniów do szkoły. Uzupełniono je otwartymi pytaniami dotyczącymi sposobów podziału nowo przyjętych uczniów na klasy oraz przydziału nauczycieli do klas.

Wszystkie etapy badania miały charakter ilościowy. . Badania z uczestnictwem uczniów miały charakter audytoryjny. Kwestionariusze kierowane do nauczycieli, dyrektorów i rodziców były rozdawane do samodzielnego wypełnienia i zbierane przez ankieterów w szkołach. Jedyńm wyjątkiem były pytania dotyczące podziału uczniów na klasy i przydziałów zespołów nauczycielskich do klas – zostały one zadane dyrektorom w formie otwartej przez ankietera, w formie quasi-wywiadu (III etap badania).

1.2. Realizacja badania

Zrealizowane badanie podłużne w gimnazjum ma skomplikowaną strukturę. Z tego względu, opisując jego przebieg, za każdym razem będziemy odwoływać się osobno do jego poszczególnych etapów. Przedstawimy tutaj jedynie najważniejsze elementy dotyczące realizacji badania.

1.3. Etap I

1.3.1. Definicja populacji i opis próby badawczej

Przy rozważaniu kwestii dotyczących populacji i próby badawczej należy pamiętać, że zrealizowane badanie miało charakter panelowy. Z tego względu definicja próby badawczej, przyjęta na I etapie badania, jest punktem wyjścia do określenia prób w następnych rundach. Drugi i trzeci etap badania przeprowadzono wśród respondentów, którzy zostali zakwalifikowani do badania na I etapie.

1.3.2. Losowanie

Próbę badawczą wylosowano przed rozpoczęciem I etapu badania. Populacja docelowa została określona jako uczniowie klas pierwszych publicznych szkół gimnazjalnych dla młodzieży, z wyłączeniem szkół specjalnych. Operatem pierwotnym losowania była lista szkół sporządzona na podstawie informacji z Systemu Informacji Oświatowej, według stanu na grudzień 2009 roku. Populacja docelowa została ograniczona do szkół, w których w klasie pierwszej uczyło się więcej niż 20 uczniów. Przyjęte ograniczenie populacji wykluczyło z operatu około 3% populacji uczniów.

Założono, że w badaniu znajdzie się 150 szkół. Schemat losowania został z jednej strony dostosowany do tej liczebności, jednak uwzględniono w nim także potrzebę zminimalizowania błędu statystycznego. Z tego względu losowanie miało charakter warstwowy, wielopoziomowy. Próba została podzielona na sześć warstw według kategorii wielkości miejscowości oraz informacji, czy szkoła jest jednooddziałowa, czy wieloodziałowa (wszystkie wyniki odwołują się do liczebności klas pierwszych). Liczbę szkół losowaną w ramach każdej z tych sześciu warstw dobrano w taki sposób, by przewidywane (na podstawie informacji o średniej liczebności oddziałów wśród szkół należących do danej warstwy) liczby uczniów w poszczególnych warstwach w zrealizowanej próbie były proporcjonalne do liczby uczniów w poszczególnych warstwach w populacji uczniów klas pierwszych gimnazjum. Prezentowany układ warstw, wraz z zakładaną liczebnością próby, znajduje się w tabeli 2.

Tabela 2. Charakterystyka warstw w próbie badawczej

warstwa (strata)	lokalizacja szkoły	wielkość szkoły	liczba (%) szkół wylosowanych do badania
1	wieś	dwa i więcej oddziałów na poziomie klasy I	48 (32%)
2	miasta do 20 tys. mieszkańców		31 (20,6%)
3	miasta od 20 do 100 tys. mieszkańców		30 (20%)
4	miasta powyżej 100 tys. mieszkańców		34 (22,6%)
5	wieś	jeden oddział na poziomie klasy I	5 (3,3%)
6	miasta od 20 do 100 tys. mieszkańców		2 (1,3%)
łącznie:			150 (100%)

W ramach warstw szkoły losowane były z prawdopodobieństwem proporcjonalnym do liczby uczniów klas pierwszych. W szkołach z jednym lub dwoma oddziałami na poziomie klasy pierwszej do badania kwalifikowani byli wszyscy uczęszczający do tej klasy uczniowie. Jeżeli oddziałów było trzy lub więcej, do badania wybierane były w sposób losowy dwa z nich (wykorzystywano do tego losowania systematyczne na podstawie liczb losowych przypisanych każdej szkole – wskazywały one na miejsce startowe i krok losowania).

Wraz z próbą główną została wylosowana próba rezerwowa. Najpierw, zgodnie z wybranym schematem losowania, wybranych zostało 300 szkół. Następnie spośród nich wylosowano warstwowo w sposób prosty 150 placówek, które trafiły do próby głównej. Pozostałe losowo uporządkowane w warstwach trafiły do próby rezerwowej. W przypadku, gdy szkoła z próby głównej odmówiła udziału w badaniu lub z powodu niespełnienia warunków realizacji nie mogła pozostać w próbie zasadniczej, wskazywano na jej miejsce pierwszą w kolejności szkołę z próby rezerwowej pochodzącą z tej samej warstwy. Kryteriami usunięcia z próby badawczej na I etapie badania były:

- brak zgody dyrektora na udział w badaniu,
- odsetek uczniów, którzy nie zostali objęci badaniem w badanych klasach, przewyższający 30% (powodami braku zakwalifikowania do badania był brak zgody rodziców na udział ucznia w badaniu lub niepełnosprawność ucznia uniemożliwiająca jego uczestnictwo w badaniu).

Spółród 150 szkół, w których przeprowadzono badanie, 34 pochodziły z próby rezerwowej (stanowi to około 22,6% badanych szkół). Należy jednak zauważyć, że w niektórych przypadkach badanie udawało się zrealizować dopiero w trzeciej lub czwartej z kolei szkole

rezerwowej. W tabeli 3 przedstawiono informacje dotyczące wykorzystania szkół rezerwowych z poszczególnych warstw.

Tabela 3. Wykorzystanie próby rezerwowej

warstwa	liczba szkół rezerwowych w próbie badawczej	liczba szkół w warstwie
1	7	48
2	10	31
3	15	30
4	23	34
5	2	5
6	2	2
ogółem	59	150

Podczas rekrutacji szkół do badania, w wyniku wykluczania szkół, w których dyrektor nie wyraził zgody na badanie lub nie osiągnięto w nich minimalnej założonej liczby zgód rodziców, skontaktowano się z większą liczbą szkół z próby rezerwowej, niż ostatecznie znalazło się ich w próbie (jak wspomniano, niekiedy badanie przeprowadzano w trzeciej lub czwartej szkole rezerwowej). Największe problemy z kwalifikacją szkół do badania pojawiły się w 3. i 4. warstwie – przypomnijmy, że są to duże szkoły w dużych i średnich miastach

Do oznaczenia szkół wylosowanych w każdej z prób (zasadniczej i rezerwowej) zostały wykorzystane kody przypisane szkołom w Systemie Informacji Oświatowej (SIO). Ankieta dyrektora danej szkoły była oznaczana tym właśnie kodem. Pozostałe kody, używane do oznaczenia klas, nauczycieli, uczniów i rodziców, utworzono według wcześniej przygotowanego schematu. Listy z nazwiskami uczniów i nauczycieli zostały zgłoszone do biura Generalnego Inspektora Ochrony Danych Osobowych (GIODO).

1.3.3. Próba badawcza oraz stopy realizacji w grupach respondentów

Najważniejszą z badanych populacji byli uczniowie rozpoczynający naukę w klasie pierwszej gimnazjum w 2009 roku. Uczniowie w wylosowanych szkołach, by mogli zostać objęci badaniem, musieli spełnić dwa warunki: ich rodzice musieli zgodzić się na udział dzieci w badaniu oraz uczeń nie mógł mieć stwierdzonej niepełnosprawności (pisał egzamin standardowy na sprawdzianie). Oba te fakty były weryfikowane w szkole podczas kontaktu wykonawcy badania z placówką. Oprócz uczniów, badaniem objęci byli także ich rodzice. Przyjęto, że próba rodziców będzie równa próbie uczniów. Badaniem w wylosowanych szkołach objęto także

nauczycieli wybranych przedmiotów (język polski, języki obce, matematyka, fizyka, chemia, biologia, geografia, historia, wiedza o społeczeństwie). Skierowano także kwestionariusz do dyrektorów gimnazjów.

Ostateczną próbę badawczą określono dopiero w momencie zebrania informacji w szkołach o liczbie uczniów, którzy mogli zostać objęci badaniem. Przypomnijmy, że jeśli w szkole odsetek uczniów, którzy nie mogli zostać objęci badaniem, przekroczył 30%, była ona usuwana z badania, a na jej miejsce wybierano szkołę rezerwową.

W tabeli poniżej znajduje się informacja o stopniu realizacji próby badawczej dla każdej z grup respondentów na I etapie badania. W każdej z nich uzyskano stopy przewyższające zakładane minimalne progi realizacyjne. Ten rezultat należy uznać za bardzo satysfakcjonujący. Wysokie progi realizacyjne były dobrym prognostykiem dla kontynuacji badania na następnych etapach, głównie w odniesieniu do zagrożenia badań panelowych, jakim jest umieralność próby.

Tabela 4. Próba I etapu badań panelowych wraz z poziomami realizacji

grupa respondentów	liczebność próby	zakładana stopa realizacji	uzyskana w badaniu stopa realizacji
uczniowie	6046	90%	92% (5567)
rodzice uczniów objętych badaniem	6046	100%	100% (6046)
nauczyciele	3028	80%	94% (2839)
dyrektorzy	150	100%	100% (150)

1.3.4. Przebieg I etapu badania

Realizacja badania w szkołach została poprzedzona wielomiesięczną pracą nad narzędziami badawczymi, zakończoną przeprowadzeniem badania pilotażowego. Objęto nimi ogólnopolską, losową próbę 12 gimnazjów. Głównym celem pilotażu było sprawdzenie trafności i rzetelności pytań zastosowanych w kwestionariuszach przygotowanych przez zespół badawczy oraz weryfikacja procedur przeprowadzania badania w szkole. W szkołach badanych podczas pilotażu wylosowano oddziały (jeden lub dwa), w których przeprowadzono badanie.

Po poprawkach kwestionariuszy, dokonanych na podstawie wyników badania pilotażowego, przystąpiono do badania zasadniczego. Na początku zwrócono się listownie do dyrektorów z prośbą o wyrażenie zgody na udział szkoły w badaniu. Do szkół wysłano także formularz zgody na badanie, informację o celach i zakresie badania oraz list od ministra edukacji.

Badanie realizowane było przez ankietatorów posiadających doświadczenie w prowadzeniu badań edukacyjnych z udziałem uczniów. Badanie testem matryc Ravena przeprowadzili

psycholodzy. Zostali oni przeszkoleni oraz otrzymali szczegółowe procedury realizacji badania i informacje na temat samego projektu. Umożliwiło to zarówno realizację badania we wszystkich klasach i szkołach według tej samej procedury, jak i przygotowanie badaczy do kontaktu z dyrektorami i pracownikami szkół.

Na ankietę spoczywał obowiązek uzyskania zgody rodziców na udział ich dzieci we wszystkich etapach badania. Był to kluczowy element rekrutacji szkół, który często rozstrzygał o pozostaniu szkoły w próbie badawczej. Rodzice uczniów otrzymywali odpowiedni list zapowiedni, kwestionariusz ankiety oraz formularz zgody na realizację badań i przetwarzanie danych osobowych. Szczegółowe rozwiązania stosowane w szkołach w związku z dystrybucją tych dokumentów były różnorodne: materiały dystrybuował ankieter, pracownik placówki (wychowawca lub dyrektor szkoły) za pośrednictwem uczniów lub podczas zebrań z rodzicami. Wypełnione kwestionariusze odbierał ankieter (nie mogły być przechowywane w szkole), zapewniając poufność uzyskanych danych. Po uzyskaniu zgód rodziców, ankieter równolegle przystępował do realizacji badań kwestionariuszowych uczniów (audytoryjnych) oraz nauczycieli i dyrektorów (ankiety przeznaczone do samodzielnego wypełniania). W tym samym czasie psycholog przeprowadzał z uczniami badanie inteligencji. Po zebraniu kwestionariuszy, sprawdzeniu kompletności ich wypełnienia oraz zrealizowaniu badań uczniów, ankieter przygotowywał protokół zbiorczy z realizacji badania oraz przekazywał komplet materiałów badawczych firmie realizującej badanie.

Cały przebieg badania w szkołach był objęty kontrolą przeprowadzaną przez osoby niezależne od firmy badawczej. Kontrolę zrealizowano w 10 gimnazjach. Zadaniem kontrolera było sprawdzenie, czy badanie przeprowadzane jest według instrukcji przygotowanej przez zespół EWD. Sama kontrola odbywała się w momencie badania ankietowego uczniów. Wykazała ona pewne uchybienia w realizacji badania, które jednak nie wpływały negatywnie na jakość zebranych danych. Uwagi dotyczące realizacji badania w skontrolowanych szkołach były na bieżąco przekazywane firmie badawczej. Pomimo drobnych nieprawidłowości sposób przeprowadzenia badania w szkołach należy uznać za zgodny z założoną procedurą badania oraz obowiązującymi standardami.

Badanie zasadnicze, rozpoczęte na początku lutego 2010 roku, powinno było zakończyć się do końca marca 2010 roku. Niestety, z powodów trudności realizacyjnych, nie udało się firmie badawczej utrzymać tego terminu – ostatecznie badanie zostało zakończone pod koniec maja 2010 roku.

Największe problemy sprawiło zebranie zgód rodziców na uczestnictwo dzieci w badaniu. Niewystarczająca liczba zgód (poniżej 70% liczebności klasy) była najczęstszym powodem wykluczenia szkoły z badania, zwłaszcza w warstwach, w których znajdowały się najliczniejsze szkoły (co zwykle było tożsame z lokalizacją szkoły w średnim lub dużym mieście). Zdarzały się przypadki wielokrotnego wyboru szkoły rezerwowej do badania w danej warstwie. Miało

to miejsce, gdy kolejne szkoły rezerwowe z różnych przyczyn były wykluczane z badania. Opóźniało to znacznie realizację badania.

Opóźnienia wynikały także z organizacji roku szkolnego. W początkowym etapie opóźnienia w rekrutacji i przeprowadzeniu badań w szkołach powodowane były przerwą świąteczną. Natomiast w końcowym etapie badania szkoły niechętnie zgadzały się na udział w badaniu z powodu egzaminów gimnazjalnych (odbywały się one pod koniec kwietnia). Terminy badań w szkołach z powodu sesji egzaminacyjnej przesuwane były na maj.

Ostatnim czynnikiem wpływającym na opóźnienie badania była katastrofa samolotu w Smoleńsku 10 kwietnia 2010 roku. W czasie następującego po niej tygodnia, objętego żałobą narodową, nie przeprowadzano badań ani nie rekrutowano szkół.

Analiza wymienionych trudności wskazuje, że szybsze rozpoczęcie realizacji badania w szkołach mogłoby pozwolić na ukończenie go w terminie. Jednak największa przeszkoda w realizacji, odmowy rodziców, prawdopodobnie w taki sam sposób utrudniałyby sprawne przeprowadzanie badań. Niestety, nie znamy dokładnych przyczyn, dla których rodzice nie wyrażali zgody na udział swoich dzieci w badaniach, dlatego trudno rozważać możliwości zmniejszenia odsetka odmów w przyszłych badaniach. Wiemy jednak, że zdecydowanie częściej sytuacja tego typu miała miejsce w dużych szkołach znajdujących się w średnich i dużych miastach. Ta informacja może stanowić pomocną wskazówkę podczas realizacji następnych badań tego typu. Szkoły zlokalizowane w średnich i dużych miastach powinny zostać objęte szczególną uwagą.

1.4. Etap II

1.4.1. Próba badawcza oraz stopy realizacji w grupach respondentów

Celem II etapu badania był pomiar osiągnięć szkolnych przed rozpoczęciem intensywnych przygotowań do egzaminu gimnazjalnego. Próbę badawczą stanowili uczniowie z wylosowanych szkół i oddziałów, dla których w I etapie uzyskano zgody rodziców na uczestnictwo w badaniu. Liczebność próby na I i II drugim etapie powinna być zatem zbieżna. Tak nie jest, gdyż pewna liczba uczniów musiała zostać usunięta z zasadniczej próby z powodów losowych (zmiana szkoły, długotrwała choroba, kalectwo bądź śmierć). Ostatecznie, w próbie II etapu znalazło się 94% uczniów, stanowiących próbę I etapu (czyli podstawową próbę całego badania panelowego).

Podczas II etapu badania założono dwa progi realizacyjne. Pierwszy z nich dotyczył szkół, drugi uczniów. Założono, że minimalnym progiem realizacji badania jest 145 szkół, w których udało się przebadать minimum 80% uczniów objętych badaniem. W badaniu wykorzystano dwa rodzaje testów: czytania oraz matematyki. Ucznia uznawano za przebadanego, jeśli wypełnił oba narzędzia.

Tabela 5. Próba II etapu badań panelowych wraz z poziomami realizacji

grupa respondentów	procent próby z I etapu (liczebność próby II etapu)	zakładane stopy realizacji	procent w stosunku do próby II etapu (uzyskane w badaniu stopy realizacji)
uczniowie	94% (5688)	80%	90,3% (5134)
szkoły	100% (150)	97%	100% (150)

Badanie udało się zrealizować z bardzo wysokimi progami realizacyjnymi. W każdej ze szkół, zrealizowano badanie uczniowskie na poziomie minimum 80% – oznacza to, że wszystkie ze 150 szkół zostały zachowane w próbie badawczej. Realizacja w całej próbie uczniów wyniosła 90,3% w stosunku do próby II etapu badania.

1.4.2. Przebieg II etapu badania

Drugi etap badania odbył się w pierwszych miesiącach nauki uczniów z próby w trzeciej klasie gimnazjum. W ramach badania wypełniali oni testy umiejętności czytania i umiejętności matematycznych (każdy z testów miał dwie wersje, z częścią zadań kotwiczących). Samo badanie zostało podzielone na trzy fazy. W fazie pierwszej Wykonawca kontaktował się z dyrektorami badanych szkół w celu wyjaśnienia przebiegu badania testowego i ustalenia terminu testowania uczniów. Druga faza badania obejmowała testowanie osiągnięć szkolnych oraz zebranie informacji o wynikach uczniów na sprawdzianie po klasie szóstej szkoły podstawowej. Testowanie przeprowadzono w terminie od 19 września do 10 października 2011 roku. Badanie w szkołach zrealizowali ankieteryzy z doświadczeniem w prowadzeniu audytoryjnych badań ankietowych z uczniami. Wszyscy ankieteryzy uczestniczący w II etapie badania zostali przeszkoleni z procedur badania w trakcie jednodniowego szkolenia. Podczas drugiej fazy badania odbyło się testowanie uczniów, w tym testowe sesje uzupełniające. Równocześnie rozpoczęła się trzecia faza badania, która obejmowała kodowanie testów osiągnięć szkolnych i tworzenie bazy danych. Faza kodowania została rozpoczęta centralnym szkoleniem koderów i ustaleniem interpretacji klucza kodowego. Szkolenie zakończyło się próbnym kodowaniem testów. Wyniki próbnego kodowania pozwoliły na lepsze opisanie kategorii kodowych i poprawę kluczy odpowiedzi. Trzecia, ostatnia faza zakończyła się 18 listopada 2011 roku. W tym dniu firma badawcza przekazała CKE wszystkie wymagane materiały badawcze. Kontrola dokumentacji, materiałów badawczych i zbiorów danych nie wykazała błędów. Realizacja badania w terenie podlegała kontroli w zakresie zgodności z procedurami. Wykonana została przez zewnętrznych obserwatorów CKE, którzy byli obecni w szkole podczas badania uczniów. Obserwacja nie wykazała żadnych poważnych błędów.

1.5. Etap III

1.5.1. Próba badawcza oraz stopy realizacji w grupach respondentów

Trzeci etap badania objął cztery grupy respondentów, które uczestniczyły w nim w I etapie: uczniów, ich rodziców, nauczycieli wybranych przedmiotów oraz dyrektorów.

W przypadku uczniów podstawę próby badawczej III etapu stanowili uczniowie objęci badaniem we wcześniejszych etapach. Tak określona próba została zweryfikowana w szkołach – wykluczeni z niej zostali ci uczniowie, którzy zmienili szkoły lub z przyczyn losowych trwale nie byli w stanie uczestniczyć w badaniu (długotrwała choroba, śmierć).

Próbę rodziców uczniów trzecich klas gimnazjów stanowili ci rodzice, którzy zgodzili się na udział ich dzieci na I etapie badań, a następnie ich dzieci nie zostały wykluczone z badania podczas II etapu badań.

W grupie nauczycieli próbę stanowili ci, którzy uczyli w wylosowanych gimnazjach wybranych przedmiotów (język polski, język obcy, matematyka, fizyka, chemia, biologia, geografia, historia, WOS). Część z nich została objęta badaniem po raz kolejny, niektórzy po raz pierwszy (np. dlatego że rozpoczęli pracę w placówce po zakończeniu realizacji I etapu badania). Z tego względu przed przeprowadzeniem badania w szkołach nie można było dokładnie wskazać liczebności próby nauczycieli.

Jako próbę dyrektorów określono wszystkich nauczycieli pełniących tę funkcję w badanych szkołach.

Podczas III etapu założono progi realizacyjne osobno dla każdej z grup respondentów. By badanie zostało uznane za zrealizowane należycie, w każdej z grup objętych badaniem musiał zostać spełniony minimalny stopień realizacji:

- Stopa realizacji w przypadku szkół i uczniów: nie mniej niż 145 szkół wskazanych do badania. Badanie uznano za zrealizowane w danej szkole, jeżeli w każdej badanej klasie przebadano przynajmniej 80% uczniów stanowiących próbę III etapu badań. Ucznia uznano za zbadanego, gdy uczestniczył w obu częściach badania – badaniu testem Ravena i badaniu kwestionariuszowym.
- Stopa realizacji w przypadku rodziców uczniów: minimum 75% w każdej z badanych klas i minimum 80% w całej próbie badawczej.
- Stopa realizacji w przypadku nauczycieli: minimum 75% w każdej z badanych szkół i minimum 80% w całej próbie badawczej.
- Stopa realizacji w przypadku dyrektorów: 100% w całej próbie badawczej.

W przypadku III etapu badania nie udało się osiągnąć zakładanych stop realizacji w dwóch szkołach.

Tabela 6. Próba III etapu badań panelowych wraz z poziomami realizacji

grupa respondentów	procent próby z I etapu (liczebność próby III etapu)	zakładane stopy realizacji	procent w stosunku do próby III etapu (uzyskane w badaniu stopy realizacji)
uczniowie	94% (5691)	80%	94% (5081)
rodzice uczniów objętych badaniem	94% (5691)	80%	86% (4706)
nauczyciele	-	80%	89% (2630)
dyrektorzy	100% (150)	100%	100% (150)
szkoły	100% (150)	97%	99% (148)

Podsumowując, mimo że w dwóch szkołach nie osiągnięto minimalnych progów, stopę realizacji III etapu badania należy uznać za satysfakcjonującą. W próbie badawczej udało się utrzymać 148 ze 150 szkół oraz 94% uczniów, z którym rozpoczynano badanie. Uczniowie przebadanie w III etapie stanowili 84% początkowej liczebności próby. Odsetki te można uznać za sukces.

1.5.2. Przebieg III etapu badania

Trzeci etap badania zaplanowano jako ostatni w panelu. Obejmował on badanie kwestionariuszowe uczniów, rodziców, nauczycieli i dyrektorów badanych klas i szkół. Narzędzia badawcze wykorzystane w tym etapie badania to: kwestionariusze dla dyrektora, nauczycieli, rodziców i uczniów oraz wybrane serie testu matryc Ravena w wersji standard – forma klasyczna (TMS-K) i w wersji dla zaawansowanych (TMZ).

Badanie było podzielone na trzy fazy. W czasie pierwszej fazy, która trwała od marca do kwietnia 2012 roku, firma realizująca badanie kontaktowała się z dyrektorami szkół, by zapoznać ich z przebiegiem badania i ustalić jego terminarz. Druga faza obejmowała badanie kwestionariuszowe uczniów, rodziców, nauczycieli i dyrektorów badanych klas oraz badanie testowe inteligencji. Trwała ona od końca kwietnia do połowy czerwca 2012 roku. Ankieterzy kierowani do szkół podczas III etapu badania posiadali doświadczenie w przeprowadzaniu badań edukacyjnych. Badanie inteligencji przeprowadzały osoby posiadające dyplom magistra psychologii. Ankieterzy przeszli centralne szkolenie, dotyczące procedury badania. Kwestionariusze uczniów były przez nich wypełniane audytoryjnie, kwestionariusze rodziców i nauczycieli zostały przekazane respondentom do samodzielnego wypełnienia, a następnie przekazane z powrotem ankieterowi. Badanie dyrektora szkoły przeprowadzono za pomocą wywiadu kwestionariuszowego. Kwestionariusze zostały oznaczone odpowiednimi kodami tak, by umożliwić jednoznaczną identyfikację respondenta i połączenie zebranych danych z jego rekordem z wcześniejszych etapów badań. Procedury badawcze zapewniały poufność

zbieranych od uczniów, rodziców oraz nauczycieli danych. Ostatnią fazą było kodowanie kwestionariuszy i testu inteligencji oraz tworzenie bazy danych. Mogła ona rozpocząć się podczas fazy drugiej, trwała do drugiej połowy czerwca 2012 roku.

Kontrola terenowa przebiegu badania została przeprowadzona w 14 szkołach. Zostały one dobrane tak, by każdy ankieter przeprowadzający badanie został objęty kontrolą przynajmniej jeden raz. Przeprowadzone kontrole nie wykazały większych uchybień. Badanie w szkołach przebiegało zgodnie z założoną przez zespół EWD procedurą badawczą.

1.6. Zmienne zależne używane w analizach

Zmiennymi zależnymi w badaniu był poziom osiągnięć szkolnych uczniów w zakresie przedmiotów humanistycznych oraz matematyczno-przyrodniczych. Za ich wskaźniki przyjęto wyskalowane wyniki egzaminu gimnazjalnego. Pobrano je ze zbioru wyników wszystkich uczniów zdających egzamin gimnazjalny w 2012 roku oraz wszystkich uczniów zdających sprawdzian w 2009 i 2008 roku, wyskalowanych na potrzeby obliczania trzyletnich wskaźników EWD dla gimnazjów.

Skalowanie przeprowadzono w modelu 2PL (dla zadań punktowanych 0–1) oraz SGRM (dla zadań punktowanych na szerszych skalach) metodą *marginal maximum likelihood* (MML). Wykorzystano oprogramowanie MPlus 7. Za wskaźnik poziomu umiejętności uczniów przyjęto oszacowania *expected a posteriori* (EAP). Ze względu na trudności techniczne związane z estymacją modelu z uwzględnieniem wszystkich zdających egzamin w danym roku (około 400 tys. osób) nie rozpatrywano wykorzystania do skalowania modelu 3PL. Warto przy tym zauważyć, że choć wykorzystanie tego modelu może zapewnić lepsze dopasowanie do danych oraz poprawniejsze oszacowania błędów standardowych dla przewidywań umiejętności uczniów niż model 2PL, o tyle same oszacowania punktowe umiejętności uczniów uzyskane z modelu 2PL i 3PL niemal się od siebie nie różnią.

Przy skalowaniu wyników części matematyczno-przyrodniczej uwzględniono zadania z testu matematycznego i z testu przyrodniczego, a przy skalowaniu części humanistycznej zadania z testu z języka polskiego oraz z testu wiedzy o społeczeństwie i historii.

Dodatkowo wynik każdego testu wchodzącego w skład egzaminu gimnazjalnego (z języka polskiego, historii i wiedzy o społeczeństwie, matematyki, biologii, chemii, fizyki, geografii) wyskalowano oddzielnie. Modele dla poszczególnych dziedzin egzaminu gimnazjalnego zostały wykorzystane w analizach dotyczących efektu nauczyciela.

We wszystkich sześciu modelach skalowania założono prostą, jednoczynnikową strukturę konstruktu. Z testu z przedmiotów przyrodniczych usunięte zostały trzy zadania o wyjątkowo niskich wartościach parametru dyskryminacji ($< 0,2$): zadanie 1, zadanie 14

i zadanie 23. Zadania te nie były uwzględniane również w modelu, w którym skalowane były wyniki części matematyczno-przyrodniczej jako całości.

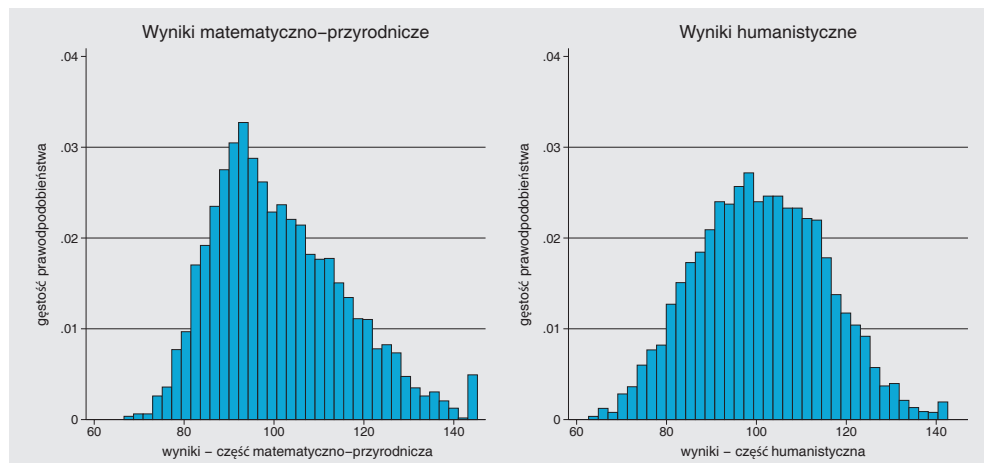
Ostatnim krokiem było przeniesienie wyskalowanych wyników na skalę o średniej 100 i odchyleniu standardowym 15. Wyniki sprawdzianu po klasie szóstej wykorzystane do analiz zostały poddane analogicznej procedurze skalowania, co wyniki egzaminu gimnazjalnego (z wykorzystaniem takiego samego schematu i modelu obliczeniowego).

Tabela 7. Statystyki opisowe wyników egzaminu gimnazjalnego w roku 2012 dla próby uczniów, których wyniki egzaminacyjne zostały połączone z informacjami z kwestionariuszy

wyniki uczniów	N	średnia	odch. std.	min	max	skośność	kurtosa	kwartył 1	mediana	kwartył 3
część matematyczno-przyrodnicza	5249	101,2	14,6	66,6	145,1	0,61	3,00	90,4	98,1	110,8
część humanistyczna	5248	101,1	14,3	62,7	142,8	0,11	2,65	90,8	100,8	111,3

W tabeli 7 znajdują się natomiast podstawowe statystyki opisowe z próby uczniów, których wyniki egzaminacyjne udało się połączyć z zebranymi w badaniu danymi.

Rysunek 1. Rozkłady wyników egzaminu gimnazjalnego w roku 2012 w próbie uczniów, których wyniki egzaminacyjne zostały połączone z danymi zebranymi w badaniu



Opisane powyżej procedury dotyczące przekształcania wyników egzaminacyjnych dokonane zostały na danych z kompletnej populacji uczniowskiej. Tego rodzaju informacje są zbierane na potrzeby projektu EWD. Przeskalowane wyniki zostały w następnym kroku przyłączone do danych uczniowskich zebranych w czasie trwania badania podłużnego. Udało się

przyłączyć wyniki dla 5249 uczniów, co stanowi blisko 87% zakładanej próby na pierwszym etapie badania. Należy jednak zauważyć, że w analizach będzie wykazywana mniejsza liczba uczniów – wynika to z braków danych, których liczba jest różna w przypadku różnych zmiennych.

Porównanie statystyk opisowych z informacjami uzyskanymi z całej populacji uczniów (tabele 9 i 10) wskazuje na dobrą reprezentację wyników populacji przez próbę. Do podobnych wniosków prowadzi nas także porównanie rozkładów wykresów (patrz rysunki 3 i 4).

1.7. Reprezentatywność próby

Przeprowadzono szereg analiz mających na celu weryfikację reprezentatywności próby. Ze względu na fakt, że w niniejszym badaniu wszelkie analizy i obliczenia przeprowadzono z wykorzystaniem danych uczniów o standardowym (niewydłużonym) toku kształcenia (czyli takich, którzy zdawali sprawdzian w roku 2009 i egzamin gimnazjalny w roku 2012), z analiz reprezentatywności wyłączeni zostali uczniowie o toku wydłużonym (tj. tacy, którzy egzamin gimnazjalny zdali w 2012 roku, jednak sprawdzian po klasie szóstej zdali w 2007 lub 2008 roku).

Badanie reprezentatywności przebiegło w dwóch krokach. W pierwszym porównano rozkłady wyników sprawdzianu oraz egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej w: (1) populacji wszystkich uczniów o standardowym toku kształcenia (trzy lata nauki w gimnazjum), (2) w populacji uczniów o standardowym toku kształcenia z wyłączeniem uczęszczających do typów szkół wykluczonych z operatu losowania (małych, przyszpitalnych, specjalnych, prywatnych, dla dorosłych), jak również w próbach (3) uczniów o standardowym toku kształcenia ze szkół wylosowanych oraz (4) ze szkół, w których badanie zostało zrealizowane. Ponadto przeanalizowano (5) wyniki szkół, które odmówiły udziału w badaniu¹. Było to konieczne, gdyż odmowa mogła wynikać z obawy przed ujawnieniem niskich umiejętności uczniów lub z innych przyczyn powiązanych z osiągnięciami szkolnymi uczniów, a te stanowią kluczową zmienną w całym badaniu. W drugim kroku porównano zróżnicowanie osiągnięć uczniów w wymienionych wcześniej grupach.

W analizach wykorzystano wyniki sprawdzianu oraz obu części egzaminu gimnazjalnego, pochodzące z bazy wyników egzaminacyjnych przygotowywanej na potrzeby wyliczania wskaźników EWD na podstawie danych pozyskanych z OKE. Informacje o szkołach pochodzą z Systemu Informacji Oświatowej². Wyniki surowe każdego z egzaminów zostały

¹ Przyjmowano, że szkoła odmówiła udziału w badaniu, jeżeli dyrektor szkoły nie wyraził zgody na badanie lub odsetek zgód rodziców uczniów z klas wylosowanych do badania był niższy niż 70 %.

² Informacje o Systemie Informacji Oświatowej znaleźć można na stronie <http://sio.men.gov.pl>.

przekształcone zgodnie z opisanym wcześniej schematem (wyskalowane z wykorzystaniem metod IRT oraz przeniesione na skalę o średniej 100 i odchyleniu standardowym 15).

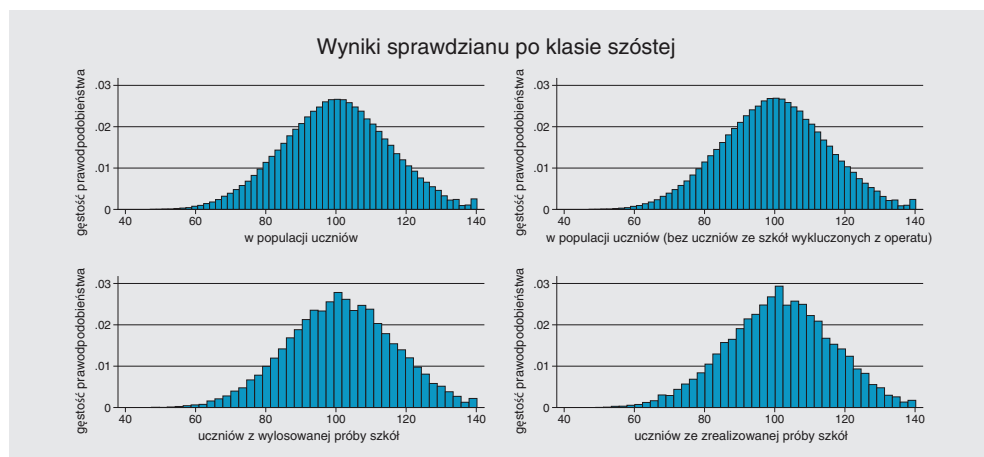
1.8. Wyniki egzaminacyjne

1.8.1. Sprawdzian po klasie szóstej

Na rysunku 2 znajdują się histogramy przedstawiające rozkłady wyników sprawdzianu po klasie szóstej uczniów, którzy rozpoczęli naukę w gimnazjach w roku 2009 i zdawali egzamin gimnazjalny w roku 2012. W górnej części znajdują się wykresy dla populacji uczniów wszystkich gimnazjów w Polsce oraz populacji uczniów gimnazjów z wyłączeniem uczęszczających do placówek wykluczonych z operatu losowania. W dolnej części znajdują się wyniki uczniów ze szkół z próby wylosowanej i faktycznie zrealizowanej.

Punkt odniesienia stanowią wyniki uczniów gimnazjów z wyłączeniem uczęszczających do placówek wykluczonych z operatu losowania (2). Widoczne jest, że rozkład wyników sprawdzianu w tej grupie nie odbiega od rozkładu w całej populacji (1), a wartości ich statystyk opisowych niemal nie różnią się (średnia odpowiednio: 99,81 i 100; odchylenie standardowe σ odpowiednio 14,9 i 15). Bardzo zbliżone są do siebie także rozkłady wyników uczniów ze szkół wylosowanych (3) i zrealizowanych (4). Jednakże w porównaniu z rozkładem w populacji pomniejszonej o uczniów ze szkół, które nie znalazły się w operacie (2), średni wynik w próbie zrealizowanej okazał się nieznacznie wyższy (o ok. 1,5 punktu). Statystyki opisowe wyników sprawdzianu uczniów we wskazanych grupach szkół umieszczone zostały w tabeli 8.

Rysunek 2. Rozkłady wyników sprawdzianu w roku 2009 w różnych grupach uczniów gimnazjów



Uwzględniono dane uczniów o niewydłużonym toku kształcenia.

Tabela 8. Statystyki opisowe dla wyników sprawdzianu w roku 2009

grupa	N	średnia	odch. std.	min	max	skośność	kurtოza	kwarty 1	mediana	kwarty 3
1	365 290	100,00	15,00	47,05	140,20	-0,026	-0,203	89,79	100,084	110,224
2	343 037	99,84	14,90	47,05	140,20	-0,019	-0,195	89,71	99,91	109,98
3	13 896	101,34	15,07	47,40	140,20	-0,060	-0,202	91,21	101,35	111,61
4	13 924	101,42	14,92	48,99	140,20	-0,097	-0,187	91,43	101,51	111,63

1 – populacja uczniów o niewydłużonym toku kształcenia,

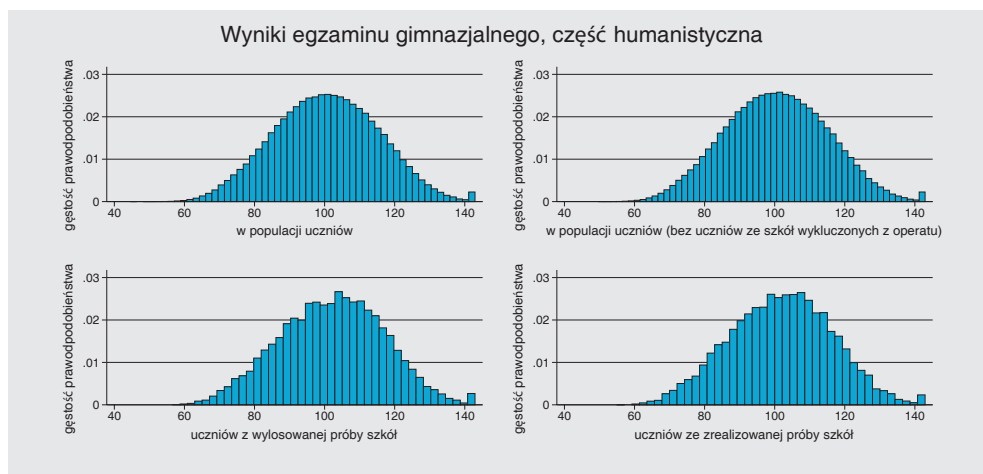
2 – populacja uczniów o niewydłużonym toku kształcenia (z wyłączeniem uczniów z typów szkół wykluczonych z operatu),

3 – uczniowie o niewydłużonym toku kształcenia z wylosowanej próby szkół,

4 – uczniowie o niewydłużonym toku kształcenia ze zrealizowanej próby szkół.

1.8.2. Egzamin gimnazjalny, część humanistyczna

Rysunek 3. Rozkłady wyników egzaminu gimnazjalnego w części humanistycznej w różnych grupach uczniów (rok 2012)



Uwzględniono dane uczniów o niewydłużonym toku kształcenia.

Rysunek 3 przedstawia histogramy wyników humanistycznej części egzaminu gimnazjalnego we wskazanych wcześniej grupach uczniów. Tak jak w przypadku wykresów dla sprawdzianu, w górnej części znajdują się wyniki w populacji wszystkich uczniów (1) oraz w populacji uczniów pomniejszonej o tych, którzy uczęszczali do szkół wykluczonych z operatu losowania (2). W dolnej części znajdują się wykresy wyników uczniów ze szkół z wylosowanej (3) i zrealizowanej próby (4). Ponownie nie ma znaczących różnic pomiędzy rozkładami wyników między grupami (1) i (2) oraz (3) i (4). Widać jednakże, że w porównaniu do grupy odniesienia – populacji uczniów z wyłączeniem

tych, którzy uczęszczali do szkół wykluczonych z operatu losowania (2) – średnie wyniki w zrealizowanej próbie są nieco wyższe (o około 1 punkt), a rozkład nieznacznie przesunięty w stronę wyników wysokich (patrz: minimum, kwartyle). Statystyki opisowe znajdują się w tabeli 9.

Tabela 9. Statystyki opisowe dla wyników egzaminu gimnazjalnego w części humanistycznej (rok 2012)

grupa	N	średnia	odch. std.	min	max	skośność	kurtoza	kwartył 1	mediana	kwartył 3
1	374 204	100,60	14,82	44,61	142,93	0,037	-0,316	90,10	100,57	111,05
2	348 209	100,56	14,59	49,76	142,93	0,058	-0,301	90,19	100,47	110,80
3	14 092	102,10	14,80	56,64	142,93	-0,021	-0,315	91,62	102,57	112,64
4	14 111	101,86	14,52	54,97	142,93	-0,006	-0,273	91,66	102,16	112,05

1 – populacja uczniów o niewydłużonym toku kształcenia,

2 – populacja uczniów o niewydłużonym toku kształcenia (z wyłączeniem uczniów z typów szkół wykluczonych z operatu),

3 – uczniowie o niewydłużonym toku kształcenia z wylosowanej próby szkół,

4 – uczniowie o niewydłużonym toku kształcenia ze zrealizowanej próby szkół.

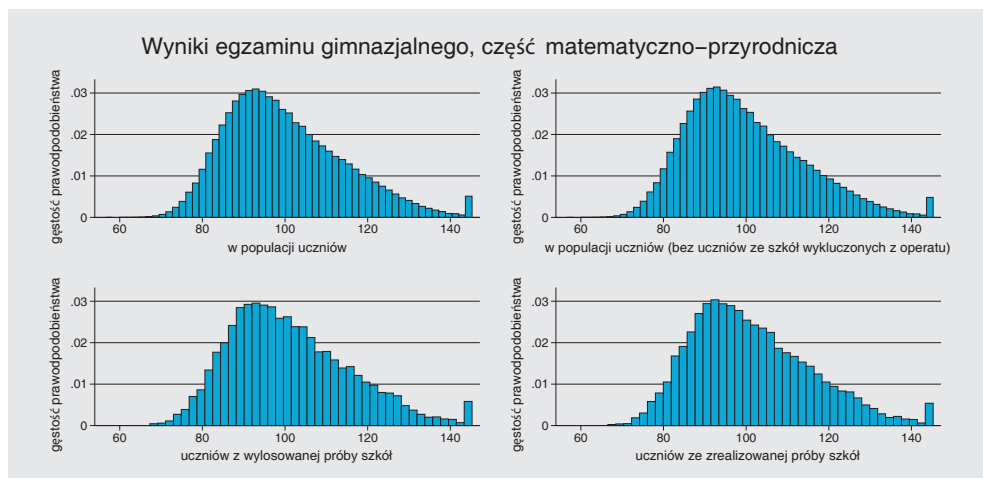
1.8.3. Egzamin gimnazjalny, część matematyczno-przyrodnicza

Na rysunku 4 przedstawiony został rozkład wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej. Podobnie jak w przypadku poprzednich wykresów, w górnej części znajdują się wyniki w populacji wszystkich uczniów (1) oraz w populacji uczniów pomniejszonej o tych, którzy uczęszczali do szkół wykluczonych z operatu losowania (2). W dolnej części znajdują się wykresy wyników uczniów ze szkół z wylosowanej (3) i zrealizowanej próby (4).

Zarówno rozkłady, jak i statystyki opisowe (tabela 10) dla danych z populacji uczniów o niewydłużonym toku kształcenia (1) i populacji tych uczniów z wyłączeniem uczęszczających do szkół wykluczonych z operatu (2) są bardzo zbliżone. Ponownie bardzo zbliżone są także rozkłady wyników uczniów ze szkół wylosowanych (3) i zrealizowanych (4). Jednakże przeciętne wyniki w grupach (3) i (4) są nieco wyższe niż w grupie (2) (średnie odpowiednio: 101,16 i 101,24), a ich rozkłady nieco mniej wypiętrzone (kurtozy odpowiednio: 0,118; 0,108) i nieznacznie przesunięte w prawo (patrz: kwartyle).

Podsumowując, zarówno w wylosowanej, jak i zrealizowanej próbie szkół przeciętne wyniki sprawdzianu oraz egzaminu gimnazjalnego okazały się nieco wyższe niż w populacji szkół uwzględnionych w operacie losowania. Rozkłady okazały się nieznacznie przesunięte w stronę wyników wysokich. Zauważone różnice nie są jednak duże (w przypadku miar tendencji centralnej – około 1/10 – 1/15 odchylenia standardowego). Wskazuje to, że próba dobrze odzwierciedla populację, z której była losowana.

Rysunek 4. Rozkłady wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej w różnych grupach uczniów (rok 2012)



Uwzględniono dane uczniów o niewydłużonym toku kształcenia.

Tabela 10. Statystyki opisowe dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej (rok 2012)

grupa	N	średnia	odch. std.	min	max	skośność	kurtოza	kwartył 1	mediana	kwartył 3
1	369 485	100,00	14,31	56,65	145,26	0,663	0,180	89,44	97,75	108,85
2	347 873	99,75	14,16	56,65	145,26	0,675	0,225	89,34	97,53	108,45
3	14 077	101,16	14,59	67,29	145,26	0,630	0,118	90,36	99,01	110,28
4	14 102	101,24	14,46	66,51	145,26	0,628	0,108	90,52	99,10	110,38

1 – populacja uczniów o niewydłużonym toku kształcenia,

2 – populacja uczniów o niewydłużonym toku kształcenia (z wyłączeniem uczniów z typów szkół wykluczonych z operatu),

3 – uczniowie o niewydłużonym toku kształcenia z wylosowanej próby szkół,

4 – uczniowie o niewydłużonym toku kształcenia ze zrealizowanej próby szkół.

1.8.4. Wyniki egzaminacyjne w szkołach, które odmówiły uczestnictwa w badaniu

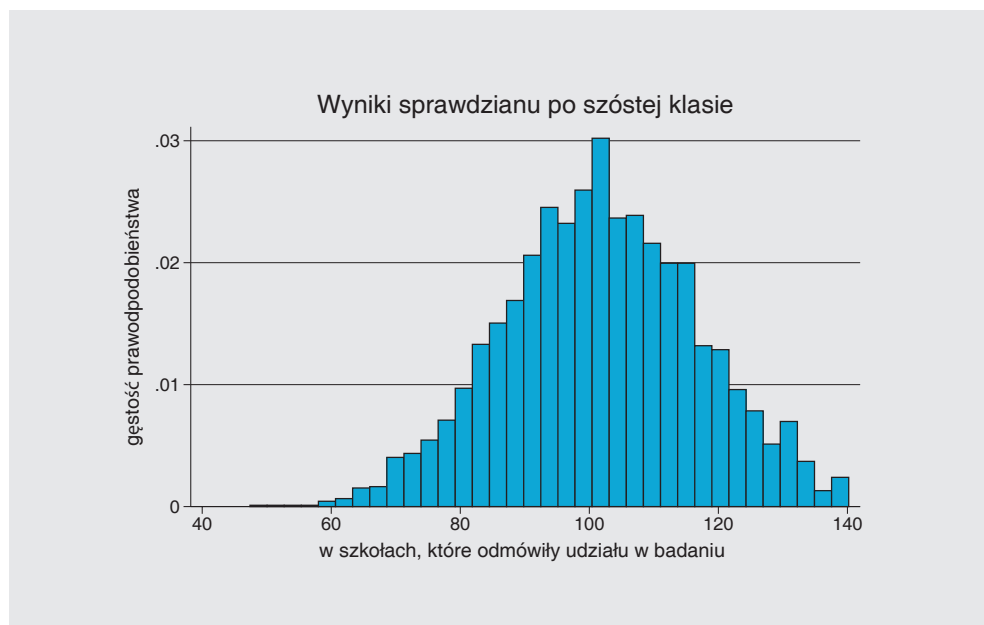
Sprawdzono rozkłady wyników egzaminów zewnętrznych w szkołach, które odmówiły uczestnictwa w badaniu. Podyktowane było to przypuszczeniem, że szkoły mogły odmawiać ze względu na niskie osiągnięcia szkolne uczniów lub z innych przyczyn powiązanych z umiejętnościami uczniów. Na rysunkach 5, 6 oraz 7 znajdują się rozkłady wyników odpowiednio sprawdzianu, egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej

uczniów tych szkół. Uwzględnione zostały tylko wyniki uczniów o niewydłużonym toku kształcenia.

Rozkład wyników sprawdzianu po klasie szóstej wydaje się nie różnić w sposób znaczący od rozkładów z prób wylosowanej (3) i zrealizowanej (4). Statystyki opisowe (tabela 11) we wszystkich trzech grupach są bardzo zbliżone, a ich rozkłady nieco przesunięte w stronę wyników wysokich w porównaniu do grupy odniesienia – populacji z wyłączeniem uczniów uczęszczających do szkół wykluczonych z operatu (2).

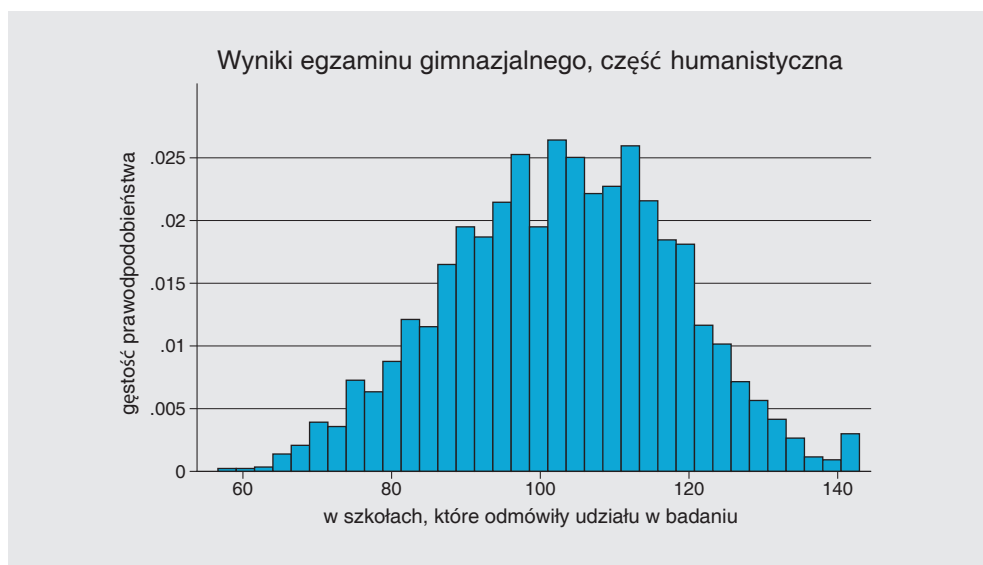
Do podobnych wniosków prowadzi analiza rozkładów wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej. Rozkład w szkołach utraconych niemal nie różni się do rozkładów w próbach wylosowanej (3) i zrealizowanej (4), widoczne jest jedynie nieco większe zróżnicowanie wyników (odchylenie standardowe odpowiednio: 15,53; 14,80; 14,52). Jednocześnie rozkład jest nieco przesunięty w stronę wyników wysokich w porównaniu do grupy odniesienia (2). Różnice w kształcie histogramu przypisać należy znacznie mniejszej liczbie obserwacji w grupie uczniów ze szkół, które nie wyraziły zgody na uczestnictwo w badaniu w porównaniu do pozostałych grup.

Rysunek 5. Rozkłady wyników sprawdzianu wśród uczniów ze szkół, które odmówiły uczestnictwa w badaniu



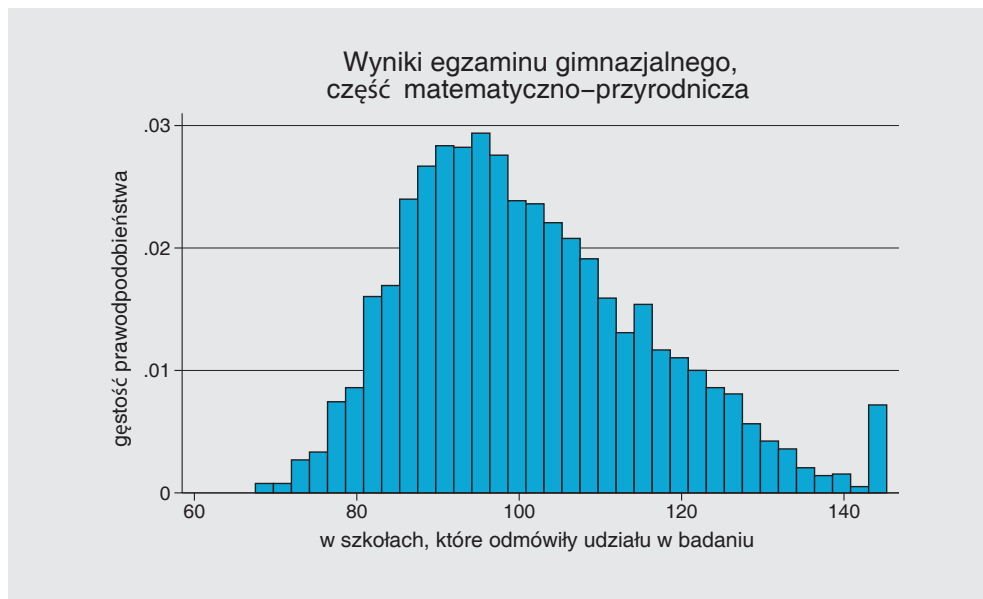
Uwzględniono dane uczniów o niewydłużonym toku kształcenia.

Rysunek 6. Rozkłady wyników egzaminu gimnazjalnego w części humanistycznej wśród uczniów ze szkół, które odmówiły uczestnictwa w badaniu



Uwzględniono dane uczniów o niewydużonym toku kształcenia.

Rysunek 7. Rozkłady wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej wśród uczniów ze szkół, które odmówiły uczestnictwa w badaniu



Uwzględniono dane uczniów o niewydużonym toku kształcenia.

Tabela 11. Statystyki opisowe dla wyników egzaminów zewnętrznych w gimnazjach, które odmówiły udziału w badaniu

egzamin	N	średnia	odch. std.	min	max	skośność	kurtoza	kwartył 1	mediana	kwartył 3
spr.	3459	101,77	15,07	47,40	140,20	-0,023	-0,197	91,70	101,68	112,02
EG hum	3514	102,04	15,53	48,12	151,40	0,214	0,933	92,03	103,94	110,23
EG mp	3509	101,37	15,09	67,51	145,26	0,629	0,0735	90,04	99,06	110,53

spr – sprawdzian

EG hum – egzamin gimnazjalny, część humanistyczna

EG mp – egzamin gimnazjalny, część matematyczno-przyrodnicza

W przypadku egzaminu gimnazjalnego w części matematyczno-przyrodniczej, wyniki uczniów ze szkół utraconych niemal nie różnią się od wyników uczniów ze szkół wylosowanych (3) i przebadanych (4) (patrz: mediana, kwartyle) i jednocześnie są nieco przesunięte w stronę wyników wysokich w porównaniu do populacji odniesienia (2). Różnice pojawiają się w przypadku trzeciego kwartyła, ponadto sam rozkład jest nieco mniej wypiętrzony.

Podsumowując, nie ma podstaw by sądzić, że odmowy szkół związane były z poziomem osiągnięć szkolnych uczniów. Stanowi to dodatkowy argument przemawiający za odzwierciedleniem w próbie populacji, z której była losowana.

1.8.5. Międzyszkolne i międzyoddziałowe zróżnicowanie wyników egzaminacyjnych

W celu weryfikacji, na ile międzyszkolne i międzyoddziałowe zróżnicowanie osiągnięć szkolnych w wylosowanej i zrealizowanej próbie odzwierciedla zróżnicowanie obserwowane w populacji, wyspecyfikowano serię trypoziomowych pustych modeli regresji (por. Domański i Pokropek, 2011; Snijders i Bosker, 2012). Zmienne zależne stanowiły opisane wcześniej wyskalowane wyniki sprawdzianu po klasie szóstej oraz części humanistycznej i matematyczno-przyrodniczej egzaminu gimnazjalnego. Modele zbudowano dla (1) populacji uczniów o niewydułżonym toku kształcenia (którzy zdali sprawdzian w roku 2009 i egzamin gimnazjalny w roku 2012), (2) populacji uczniów o niewydułżonym toku kształcenia z pominięciem uczniów uczęszczających do szkół wykluczonych z operatu losowania, (3) uczniów z wylosowanych szkół, (4) uczniów ze szkół, w których zrealizowano badanie. Wykorzystano oprogramowanie Stata 11.2, estymacja przeprowadzona została metodą *full maximum likelihood* (por. Snijders i Bosker, 2012). W tabeli 12 przedstawiono wartości współczynnika korelacji wewnątrzgrupowej (*intra-class correlation*, ICC), informującego, jaki odsetek zróżnicowania wyników sprawdzianu przypisać można podziałowi na szkoły oraz oddziały wewnątrz szkół (obliczone zostały na podstawie oszacowań

wariancji na poziomie uczniów, klas i szkół). Pełne wyniki modelowania zamieszczone zostały w aneksie.

Analiza wartości współczynników korelacji wewnątrzgrupowej wskazuje, że zróżnicowanie międzyszkolne wyników sprawdzianu³ jest nieco większe w całej populacji (1) niż w populacji z wyłączeniem uczniów uczęszczających do gimnazjów małych, niepublicznych, specjalnych oraz szkół dla dorosłych (2) (z tej próby dokonano losowania szkół). Oszacowania punktowe w całej populacji okazały się o niespełna 1,5% wyższe na poziomie szkół (odpowiednio: 11,07% i 8,63%). Na poziomie klas różniły się nieznacznie. Analizując oszacowania w próbach wylosowanej i zrealizowanej, należy wziąć pod uwagę mniejszą ich precyzję w porównaniu do oszacowań uzyskanych dla grup (1) i (2), widoczną w szerokości przedziałów ufności. Oszacowania punktowe w próbie wylosowanej (3) okazały się wyższe niż w populacji odniesienia (2), zarówno na poziomie szkół, jak i oddziałów. Jednakże próba zrealizowana (3) odznacza się wartościami korelacji wewnątrzgrupowej bardziej zbliżonymi do grupy odniesienia (2). Na poziomie oddziałów są one wyższe od obserwowanych w grupie (2), nawet gdy uwzględnimy ich oszacowanie z pewnym błędem (przedziały ufności), jednak na poziomie szkół, gdy uwzględnimy popełniany błąd, możemy uznać je za zbliżone.

W tabeli 13 przedstawiono ICC dla poziomu szkół oraz oddziałów wewnątrz szkół z modeli, w których zmienną zależną stanowiły wyniki egzaminu gimnazjalnego w części humanistycznej. Podobnie jak w modelach dla sprawdzianu, wykluczenie z operatu uczniów ze szkół małych, niepublicznych, specjalnych i dla dorosłych (2) związane było ze zmniejszeniem zróżnicowania na poziomie szkół w porównaniu do całej populacji (1), choć na poziomie klas oszacowania pozostały zbliżone. Wylosowana próba (3) odznaczała się większym zróżnicowaniem międzyszkolnym i międzyoddziałowym w porównaniu do grupy odniesienia (2) (odpowiednio: 13,24%; 14,3% oraz 9,85%; 11,56%). Podobnie było w przypadku próby zrealizowanej, jednak uwzględniając błąd oszacowań, przyjąć możemy, że wartości na poziomie szkół były zbliżone; na poziomie klas okazały się nieco wyższe. Jednocześnie w próbie zrealizowanej (4) zróżnicowanie międzyszkolne było niższe niż w całej populacji (1), choć między oddziałami wewnątrz szkół – nieco wyższe (odpowiednio: 11,26%; 16,43% oraz 13,73%; 11,09%).

³ Należy pamiętać, że analizujemy międzyszkolne i międzyoddziałowe zróżnicowanie wyników sprawdzianu uczniów w gimnazjach, w których rozpoczęli naukę, a nie w szkołach podstawowych, które ukończyli.

Tabela 12. Wyniki trzypoziomowych pustych modeli regresji dla zmiennej zależnej: sprawdzian – współczynniki korelacji wewnątrzgrupowej (ICC) dla poziomu szkół oraz oddziałów

grupa	poziom szkół			poziom oddziałów		
	ICC	przedział ufności 95%		ICC	przedział ufności 95%	
1	11,07%	10,60%	11,55%	9,69%	9,46%	9,91%
2	8,63%	8,22%	9,06%	9,84%	9,61%	10,07%
3	12,08%	9,54%	15,17%	12,98%	11,86%	14,08%
4	10,47%	8,14%	13,35%	12,19%	11,06%	13,31%

1 – populacja uczniów o niewydłużonym toku kształcenia

2 – populacja uczniów o niewydłużonym toku kształcenia (z wyłączeniem uczniów z typów szkół wykluczonych z operatu)

3 – uczniowie o niewydłużonym toku kształcenia z wylosowanej próby szkół

4 – uczniowie o niewydłużonym toku kształcenia ze zrealizowanej próby szkół

Tabela 13. Wyniki trzypoziomowych pustych modeli regresji dla zmiennej zależnej: egzamin gimnazjalny, część humanistyczna – współczynniki korelacji wewnątrzgrupowej (ICC) dla poziomu szkół oraz oddziałów

grupa	poziom szkół			poziom oddziałów		
	ICC	przedział ufności 95%		ICC	przedział ufności 95%	
1	16,43%	15,83%	17,05%	11,09%	10,88%	11,30%
2	9,85%	9,40%	10,33%	11,56%	11,32%	11,80%
3	13,24%	10,52%	16,48%	14,30%	13,18%	15,36%
4	11,26%	8,79%	14,29%	13,73%	12,58%	14,84%

1 – populacja uczniów o niewydłużonym toku kształcenia,

2 – populacja uczniów o niewydłużonym toku kształcenia (z wyłączeniem uczniów z typów szkół wykluczonych z operatu),

3 – uczniowie o niewydłużonym toku kształcenia z wylosowanej próby szkół,

4 – uczniowie o niewydłużonym toku kształcenia ze zrealizowanej próby szkół.

Tabela 14. Wyniki trzypoziomowych pustych modeli regresji dla zmiennej zależnej: egzamin gimnazjalny, część matematyczno-przyrodnicza – współczynniki korelacji wewnątrzgrupowej (ICC) dla poziomu szkół oraz oddziałów

grupa	poziom szkół			poziom oddziałów		
	ICC	przedział ufności 95%		ICC	przedział ufności 95%	
1	12,76%	12,25%	13,29%	11,35%	11,13%	11,57%
2	9,88%	9,42%	10,35%	11,61%	11,38%	11,85%
3	12,96%	10,24%	16,24%	14,83%	13,72%	15,87%
4	12,03%	9,44%	15,17%	13,70%	12,59%	14,76%

1 – populacja uczniów o niewydłużonym toku kształcenia,

2 – populacja uczniów o niewydłużonym toku kształcenia (z wyłączeniem uczniów z typów szkół wykluczonych z operatu),

3 – uczniowie o niewydłużonym toku kształcenia z wylosowanej próby szkół,

4 – uczniowie o niewydłużonym toku kształcenia ze zrealizowanej próby szkół.

W tabeli 14 przedstawiono wartości ICC na poziomie szkół oraz oddziałów wewnątrz szkół z modeli, w których zmienną zależną stanowiły wyniki egzaminu gimnazjalnego w części matematyczno-przyrodniczej. Wzorec uzyskanych wyników okazał się podobny do wzorca uzyskanego dla sprawdzianu i humanistycznej części egzaminu gimnazjalnego. Ponownie w grupie (2) zróżnicowanie międzyszkolne było niższe niż w całej populacji (1) przy zbliżonych wartościach na poziomie oddziałów. W próbie wylosowanej (3) zróżnicowanie międzyszkolne i międzyoddziałowe było nieco większe niż w grupie odniesienia (2). W próbie zrealizowanej (4), w porównaniu do grupy odniesienia (2), wartości korelacji wewnątrzgrupowej na poziomie szkół przyjęły wartości nieco wyższe, lecz jeżeli uwzględnimy błąd popełniany przy szacowaniu, możemy je uznać za zbliżone. Korelacje na poziomie klas są jednak widocznie wyższe.

Ponadto zróżnicowanie międzyoddziałowe w całej populacji (1) okazało się mniejsze niż w próbie zrealizowanej (4) (odpowiednio: 11,13%; 13,70%), natomiast międzyszkolne – zbliżone (odpowiednio: 12,76% i 12,03%).

Analiza zróżnicowania wyników trzech egzaminów zewnętrznych pokazała zatem, że wylosowana próba (3) była nieco bardziej zróżnicowana na poziomie szkół i klas niż populacja, z której pochodzi (2). Właściwości zrealizowanej próby (4) są jednak korzystniejsze – odsetek wariancji wyjaśniany przez przynależność do szkoły można uznać za zbliżony do obserwowanego w grupie odniesienia (2), choć pozostaje on nieco większy dla oddziałów wewnątrz szkół. Jednocześnie zróżnicowanie międzyszkolne w zrealizowanej próbie na początku badania było zbliżone do zróżnicowania w całej populacji uczniów (1), natomiast międzyoddziałowe – większe. Na zakończenie gimnazjum zróżnicowanie międzyszkolne było już mniejsze niż w całej populacji (1) przy niezmiennym wzorcu na poziomie oddziałów.

Podsumowując, przeprowadzone analizy miały odpowiedzieć na pytanie o jakość wylosowanej i zrealizowanej próby. Analiza rozkładu wyników sprawdzianu oraz egzaminu gimnazjalnego wykazała, że w obu próbach są one tylko nieznacznie przesunięte w stronę wyników wysokich. Pomyślnie zakończyła się także analiza zróżnicowania międzyszkolnego i międzyoddziałowego wyników egzaminów zewnętrznych. W zrealizowanej próbie (4) odsetek wariancji wyników egzaminów wyjaśniany przynależnością do klas jest zbliżony do odsetka w populacji, z której próba pochodzi (2), choć na poziomie klas jest wyższy. Pamiętać jednak należy, że podczas analizy reprezentatywności uwzględniono wyniki uczniów ze wszystkich klas w szkole, natomiast w badaniu uczestniczyli wyłącznie uczniowie nie więcej niż dwóch klas. Mogło to spowodować niedoszacowania wariancji międzyoddziałowej w analizach przedstawionych w dalszych częściach tego raportu. Tym niemniej analizy wskazują, że pobrana próba była wysokiej jakości. Świadczy to o tym, że odzwierciedla ona w zadowalającym stopniu populację, z której pochodzi.

1.9. Wagi stosowane w analizach

1.9.1. Wagi szkół, klas i uczniów

Wagi w analizach przedstawionych w niniejszej publikacji zastosowano z kilku powodów. Po pierwsze, w celu uwzględnienia różnic w prawdopodobieństwie znalezienia się poszczególnych jednostek badanych w próbie. Szkoły do próby były losowane z prawdopodobieństwem wprost proporcjonalnym do liczby uczniów w klasach pierwszych. Jednak szkoły na wsiach są mniejsze. Z tego powodu w próbie znalazło się więcej szkół zlokalizowanych w miastach, niż wynikałoby to z udziału szkół miejskich w liczbie szkół ogółem. Ponadto, wagi wykorzystano w celu lepszego ozwierciedlenia relacji zachodzących w populacji. Ze względu na wielopoziomowy charakter losowania (szkoły i klasy) oraz możliwość nie wzięcia udziału w badaniu przez uczniów pomimo wylosowania (np. ze względu na nieobecność w szkole w dniu ankietowania), na potrzeby analiz przygotowano kilka grup wag. Pierwsza grupa to wzajemnie uzupełniające się wagi szkół, klas i uczniów. Drugą grupę stanowiła waga będąca iloczynem wag szkół, klas oraz uczniów. Trzecią grupę – 76 wag replikacyjnych stosowanych w modelach jednopoziomowych, do właściwego oszacowania zmienności w danych pobranych, w ramach schematu losowania grupowego.

Wagi obliczono w taki sposób, by informowały o tym ile jednostek z populacji badawczej reprezentuje dana jednostka w próbie.

Na poziomie szkół informują one, jaką liczbę szkół podobnych pod względem wielkości miejscowości, w której znajduje się placówka oraz liczby oddziałów w szkole, prezentuje wylosowany podmiot. Jeśli wylosowanych zostało pięć jednooddziałowych szkół wiejskich, to wagi na tym poziomie informują, ile tego typu szkół reprezentuje każda z nich. Wagi te stosowane są w analizach trzypoziomowych na poziomie szkół.

Wagi na poziomie klas informują, jaką liczbę klas z wylosowanej szkoły reprezentują wylosowane klasy. Jeśli w szkole było 5 klas pierwszych, to dla każdej z wylosowanych dwóch klas waga ta przyjmie wartość 2,5. W analizach trzypoziomowych wagi te stosowane są na poziomie klas. Wagi uczniów wprowadzają korektę ze względu na poziom realizacji badania w danej klasie. Informują zatem, jaką liczbę uczniów danego oddziału reprezentują uczniowie przebadani. Jeśli w klasie na III etapie badania było 25 osób stanowiących próbę podstawową (osoby niewykluczone z badania), a ostatecznie narzędzia badawcze uzupełnione zostały przez 20 osób, to każdemu uczniowi w tej klasie przyporządkowana zostanie waga 1,25. Ze względu na wieloetapowy charakter badania oraz konieczność dołączenia do zbioru informacji o wynikach uzyskanych przez uczniów na sprawdzianie szóstoklasisty oraz na egzaminie gimnazjalnym, ostatecznie korekty ze względu na realizację na poziomie klas obliczone zostały jedynie dla uczniów zaklasyfikowanych do badania i jednocześnie takich, dla których przyłączono wyniki z obu egzaminów. W sytuacji, gdy uczeń był badany

na którymkolwiek etapie, jednak nie udało się dla niego połączyć danych z wynikiem sprawdzianu i egzaminu, liczebność efektywnie przebadanych respondentów w klasie była pomniejszana, natomiast waga uczniów w klasie – obliczana z uwzględnieniem niższego stopnia realizacji. Wartość tej wagi została ograniczona do liczby 3, podobnie jak np. w badaniach PISA (*Programme for International Student Assessment*, OECD, 2009), w celu ograniczenia wpływu przypadków odstających. Waga ta jest wykorzystywana w analizach wielopoziomowych na najniższym poziomie analizy (indywidualnym).

Iloczyn wagi szkoły, klasy i ucznia to indywidualna waga ostateczna. Na jej podstawie, opierając się na macierzy Hadamarda z korektą Faya (0,5), przygotowany został zestaw 76 wag replikacyjnych.

1.9.2. Przypadki szczególne

W dwóch szkołach, w których badano dwa oddziały na poziomie klasy pierwszej, jeden z nich został rozwiązany. Wagi dla tych szkół pozostawiono bez zmian. Wagi dla oddziałów podwojono, jednak nie więcej niż do wartości 3. Ze względu na okoliczność zlikwidowania oddziałów pomiędzy pierwszym a drugim rokiem nauki przyjęto, że uczniowie, którzy przeszli z tych oddziałów do drugiego badanego oddziału, zostaną wliczeni do ogólnej liczby uczniów w tej klasie. Na podstawie tej liczby wylicza się korektę. Liczba ta jest podstawą procentowania przy wprowadzaniu korekty ze względu na poziom realizacji badania w danej klasie.

1.9.3. Wagi nauczycieli

W analizach, w których nauczyciele traktowani są jako grono pedagogiczne danej szkoły, przyjmuje się dla nich łącznie wagi odpowiednich szkół. W przypadku przyłączania konkretnych nauczycieli do klas, w których uczyli, stosowane są wagi dla klas lub (w przypadku analiz dwupoziomowych) iloczyn wagi szkoły i wagi klasy.

1.9.4. Modele podstawowe stosowane w analizach

Po przygotowaniu zbioru danych wraz z wagami przystąpiono do opracowania zestawu modeli podstawowych. Działanie to miało dwa zasadnicze cele. Pierwszy to sprawdzenie, jak na danych dla przebadanych 150 gimnazjów wyglądałyby modele (dwupoziomowe), analogiczne do estymowanych na danych populacyjnych (modeli służących obliczaniu edukacyjnej wartości dodanej gimnazjów). Realizacja tego celu daje podstawy do porównania modeli stosowanych przy obliczaniu wskaźników EWD (dla wszystkich gimnazjów w Polsce) z modelami estymowanymi na danych dla 150 gimnazjów z próby. Porównanie to pośrednio wskazuje na reprezentatywność wybranych szkół. Celami szczegółowymi było przetestowanie modeli dwupoziomowych dla wszystkich sześciu części egzaminu gimnazjalnego (matematycznej,

przyrodniczej, języka polskiego, WOS-u i historii, matematyczno-przyrodniczej i humanistycznej) na podstawie danych egzaminacyjnych. Celem drugim – nakierowanym przede wszystkim na usprawnienie prac zespołu opracowującego raport z badań – było (1) opracowanie modeli trzypoziomowych analogicznych do modeli, w których szacowane są wskaźniki EWD oraz (2) przygotowanie podstawowego zestawu zmiennych kontrolowanych (wiek, inteligencja, wykształcenie rodziców, ich status zawodowy, zasobność gospodarstwa domowego i sytuacja rodzinna), które mogłyby być wykorzystywane w modelach bardziej zaawansowanych, w których testowane będą postawione hipotezy. Analizy te, mimo, że za wykorzystaniem poszczególnych zmiennych stoją przesłanki teoretyczne, miały przede wszystkim charakter eksploracyjny, zwłaszcza pod względem wyboru potencjalnych predyktorów osiągnięć szkolnych.

1.10. Analiza modeli dwupoziomowych analogicznych do populacyjnych modeli EWD

Dla realizacji pierwszego z celów przygotowano i przeprowadzono wieloetapową procedurę. Dla każdej z sześciu osobno wyskalowanych części egzaminu powtórzone zostały następujące etapy:

1. Określenie stopnia wielomianu sprawdzianu zapewniającego najlepsze dopasowanie modelu do danych.
 - a. Ze względu na brak możliwości obliczenia testu istotności zmiany chi-kwadrat na podstawie wartości testu dla modelu na danych ważonych określenie stopnia wielomianu odbywało się na danych bez wag. Wprowadzanie kolejnych potęg wyniku sprawdzianu następowało do momentu, w którym dany model nie był istotnie lepszy od poprzedniego. Wykorzystano test ilorazu wiarygodności (*likelihood-ratio test*).
 - b. Procedura była powtarzana na danych ważonych, jednak przy określaniu, która potęga sprawdzianu zapewnia wystarczające dopasowanie do danych, kierowano się poziomem istotności kolejnych wprowadzanych do modelu zmiennych. Jeśli w danym modelu nowo wprowadzona zmienna nie była istotnie związana z wynikiem danej części egzaminu ($p > 0,05$), to ustalono, że maksymalną potęgą sprawdzianu wprowadzaną do modelu będzie ta z poprzedniego modelu.
 - c. Następnie porównano wyniki modeli dla danych ważonych i nieważonych.
2. Wprowadzenie do modelu zmiennej płęć.

Tabela 15. Porównanie modeli populacyjnych i modeli dla danych z próby badawczej (ważonych i nieważonych)

	model część polonistyczno-humanistyczna				model część polonistyczna				model część humanistyczna			
	populacyjny	próba nieważony	próba ważony		populacyjny	próba nieważony	próba ważony		populacyjny	próba nieważony	próba ważony	
stała	100,6	100,9	100,9		100,1	100,6	100,5		102,0	102,0	102,0	
wynik sprawdzianu (skala 0;1)	10,8	11,8	12,0		11,0	11,2	11,2		11,5	11,5	11,3	
druga potęga wyniku sprawdzianu	-0,3	-0,4	-0,4		-1,3	-1,7	-1,6		0,1	0,2	-0,5	
trzecia potęga wyniku sprawdzianu		-0,3	-0,4		-0,4	-0,3	-0,4		-0,7	-0,5		
czwarta potęga wyniku sprawdzianu					0,1	0,2	0,2					
piąta potęga wyniku sprawdzianu					0,0							
pleć: kobieta		1,4	1,7		4,8	4,5	4,9		-1,7	-2,0	-2,0	
dysleksja na egz. gimnazjalnym	1,2	1,4	1,8		0,8	1,4	1,7		1,4	0,0	1,5	
dysleksja na sprawdzianie	-4,2	-3,2	-3,2		-4,4	-3,7	-3,6		-3,2	-1,4	-2,0	
dysleksja na obu egzaminach	0,6				1,1				0,0			
wariancja												
poziom szkoły	8,6	6,6	6,9		8,7	6,9	8,2		8,3	6,9	6,5	
poziom ucznia	80,5	77,7	79,2		86,7	84,0	84,6		100,4	97,1	98,3	

	model część matematyczno-przyrodnicza				model część matematyczna				model część przyrodnicza			
	populacyjny	próba nieważony	próba ważony		populacyjny	próba nieważony	próba ważony		populacyjny	próba nieważony	próba ważony	
stała	102,3	102,0	102,0		100,3	100,5	100,6		100,8	101,0	100,6	
wynik sprawdzianu (skala 0;1)	11,9	11,4	11,5		12,5	13,2	12,0		10,9	11,2	11,2	
druga potęga wyniku sprawdzianu	-1,0	-0,2	-0,3		2,0	2,0	2,0		0,9	0,8	0,8	
trzecia potęga wyniku sprawdzianu	-0,6	-0,3	-0,4		-0,6	-1,1	-0,6		-0,3	-0,2	-0,3	
czwarta potęga wyniku sprawdzianu					-0,2	-0,2	-0,2					
piąta potęga wyniku sprawdzianu					-1,2	-1,0	-0,9		-0,9	-1,1	-0,6	
pleć: kobieta	-1,3	-1,5	-1,4		1,5	1,4	1,9		1,5	0,0	1,7	
dysleksja na egz. gimnazjalnym	-3,5	1,6	2,0		-3,4				-3,0	-1,6	-2,0	
dysleksja na sprawdzianie					0,2	-3,6	-4,5		0,0			
dysleksja na obu egzaminach	0,5	-3,1	-4,1									
wariancja												
poziom szkoły	6,8	5,7	6,0		8,1	8,2	9,2		7,3	5,6	5,2	
poziom ucznia	87,9	87,1	86,3		84,4	84,4	83,0		101,6	97,7	98,9	

3. Wprowadzenie do modelu zmiennych: dysleksja⁴ na sprawdzianie, dysleksja na egzaminie gimnazjalnym oraz ich interakcji.
 - a. Ze względu na wielkość próby, wszystkie zmienne dotyczące posiadania opinii o dysleksji nie oddziaływały istotnie w modelach, do których były wprowadzane łącznie. Z tego względu dla każdej części egzaminu przetestowano modele z różną konfiguracją zmiennych opisujących dysleksję.
 - b. Wybór modelu najlepszego dla zmiennych ważonych i zmiennych nieważonych odbył się na podstawie kryteriów analogicznych do zastosowanych podczas wyboru stopnia wielomianu sprawdzianu.

Współczynniki z tak wyłonionych modeli były porównywane z analogicznymi modelami EWD oszacowanymi na danych populacyjnych. W efekcie przeprowadzonych analiz można stwierdzić, że modele szacowane na danych populacyjnych oraz na danych ze 150 wylosowanych do badań gimnazjów nie różnią się od siebie. Pośrednio wskazuje to na reprezentatywność badanej próby szkół.

1.11. Analiza modeli trzypoziomowych i testowanie oddziaływania zmiennych dodatkowych

Cel drugi procesu szacowania modeli podstawowych osiągnięty został na podstawie następujących procedur analitycznych. Dla oszacowania współczynników i parametrów modeli trzypoziomowych estymowanych analogicznie do modeli EWD, przeprowadzono procedurę podobną, jak miało to miejsce dla modeli dwupoziomowych. Poniżej opisano kolejne kroki.

- a. Określenie stopnia wielomianu odpowiadającego najlepszemu dopasowaniu zmiennych do danych przeprowadzono na danych nieważonych i na danych ważonych (zgodnie z kryteriami wyboru zastosowanymi w modelach dwupoziomowych).
- b. Porównanie wyników modeli dla danych ważonych i nieważonych.
- c. Wprowadzenie do modelu zmiennej płeć.
- d. Wprowadzenie do modelu zmiennych: dysleksja na sprawdzianie, dysleksja na egzaminie gimnazjalnym oraz ich interakcji (zgodnie z procedurą przyjętą dla modeli dwupoziomowych).

⁴ Ścisłej – posiadanie zaświadczenia o specyficznych trudnościach w uczeniu się wydanego przez poradnię psychologiczno-pedagogiczną. W tekście będziemy się odwoływać do tej zmiennej skrótowo, stosując termin „dysleksja”.

Modele dwupoziomowe analizowane były z wykorzystaniem oprogramowania Stata 11. W modelach z wagami wykorzystywany był odpowiedni rodzaj skalowania wag. Ze względu na brak możliwości wprowadzenia skalowania wag w modelach trzypoziomowych w programie Stata 11, do testowania modeli wykorzystany został program HLM 6.

W następnym kroku testowano znaczenie dodatkowych zmiennych na poziomie ucznia w modelach dwupoziomowych i trzypoziomowych. Poniżej przedstawiono etapy procesu testowania istotności wpływu dodatkowych zmiennych indywidualnych w modelu dwupoziomowym.

- a. Po wstępnej analizie szeregu zmiennych, procedurze testowania znaczenia dla osiągnięć szkolnych w poszczególnych dziedzinach (częściach egzaminu gimnazjalnego) poddano następujące zmienne: poziom inteligencji na pierwszym etapie badania (w pierwszej klasie) mierzony testem matryc Ravena, wyższy z poziomów wykształcenia rodziców mierzony liczbą lat nauki, wyższy ze statusów społeczno-ekonomicznych rodziców mierzonych na skali *International Socio-Economic Index* (ISEI), wyższy z prestiżu zawodów rodziców mierzony skalą *Standard International Occupational Prestige Scale* (SIOPS), wiek w miesiącach, analizowany łącznie z informacją o wcześniejszym lub późniejszym, niż wynikałoby to z roku urodzenia, podjęciu nauki w gimnazjum, zasobność gospodarstwa domowego i jego wyposażenie w różnego rodzaju kategorie dóbr (mierzona za pomocą skal będących efektem analizy danych o wyposażeniu gospodarstwa domowego), liczba dzieci, wielkość rodziny, kolejność urodzenia badanego ucznia, struktura rodziny (rodzina pełna, rodzina niepełna, rodzina dwu-, trzypokoleniowa).
- b. Do modelu analogicznego do modelu EWD (wyłonionego w pierwszym kroku specyfikacji) dołączane były poszczególne zmienne w kolejności wymienionej powyżej. Analizy prowadzono jedynie na danych ważonych.
- c. W przypadku poziomu inteligencji najpierw testowano, czy ma on znaczenie dla wyników w danej części egzaminu bez kontroli zmiennych zawartych w modelach EWD. Następnie wprowadzano ją łącznie z tymi zmiennymi i sprawdzano, czy pozostaje istotnym statystycznie predyktorem ($p < 0,05$), oraz jak zmieniły się współczynniki regresji.
- d. Kolejne zmienne testowane były w czterech krokach:
- e. Do modelu wprowadzana była sama testowana zmienna (analogicznie do poziomu inteligencji).
- f. Następnie wprowadzana była łącznie ze zmiennymi indywidualnymi, które w poprzednich krokach były istotnymi predyktorami, lecz bez zmiennych z modeli analogicznych do modeli EWD.

- g. W trzecim kroku wprowadzano jedynie testowaną zmienną do modelu ze zmiennymi z modeli EWD.
- h. Ostatecznie testowano znaczenie zmiennej w modelu ze „zmiennymi EWD” oraz tymi zmiennymi indywidualnymi, które w poprzednich krokach były istotnymi predyktorami.
- i. Zmienną pozostawiano w modelu, jeśli przy kontroli wszystkich pozostałych zmiennych pozostawała istotnym statystycznie predyktorem.
- j. Trzy zmienne dotyczące wieku ucznia były testowane łącznie (jednocześnie).

Oszacowania współczynników i parametrów z modeli dwupoziomowych estymowanych z wykorzystaniem programów Stata 11 i HLM 6 przedstawione zostały w aneksie. Podsumowując zawarte tam informacje, widzimy, iż na podstawowym poziomie nie ma znaczących różnic w oszacowaniach parametrów między tymi dwoma programami.

Procedura testowania modeli trzypoziomowych ze zmiennymi indywidualnymi była podobna do procedury dla modeli dwupoziomowych. Testowano w nich także te same zmienne. Najpierw do modelu ze zmiennymi analogicznymi do modelu EWD (wyłonięgo w pierwszym kroku specyfikacji) dołączano poszczególne zmienne. Opracowane modele trzypoziomowe przedstawiono w aneksie.

Zmienną z modelu analogicznego do EWD (dwu- lub trzypoziomowego) pozostawiano w analizie, nawet jeśli przestawała mieć istotne znaczenie dla wyników egzaminu gimnazjalnego po włączeniu zmiennych indywidualnych. W kilku przypadkach podjęto decyzję o pozostawieniu zmiennych indywidualnych w modelach ostatecznych, mimo że nie wpływały istotnie na zmienną zależną. Dotyczyło to sytuacji, w których wartości poziomu istotności były bliskie 0,05.

Analiza opracowanych modeli podstawowych pozwala stwierdzić, że wartości parametrów dwupoziomowych modeli estymowanych nie różnią się od wartości z analogicznych modeli populacyjnych, używanych do wyliczania wskaźników EWD. Innymi słowy przyjąć można, że wyniki z pobranej próby są reprezentatywne dla populacji. Ponadto analiza modeli podstawowych pozwoliła wyłonić zbiór zmiennych istotnie powiązanych z wynikami obu części egzaminu gimnazjalnego przy kontroli wyniku sprawdzianu, tak w modelach dwu- jak i trzypoziomowych.

Literatura

Domański, H. i Pokropek, A. (2011). *Podziały terytorialne, globalizacja a nierówności społeczne: wprowadzenie do modeli wielopoziomowych*. Warszawa: Wydawnictwo IFiS PAN.

Ganzeboom, H. B. G., De Graaf, P. M. i Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56.
doi:10.1016/0049-089X(92)90017-B

Jaworowska, A. i Szustrowa, T. (1991a). *Podręcznik do testu matryc Ravena. Wersja Standard (1956). Polska standaryzacja 1989 (5;11–15;11)*. Warszawa: Pracownia Testów PTP.

Jaworowska, A. i Szustrowa, T. (1991b). *Podręcznik do Testu Matryc Ravena. Wersja dla Zaawansowanych. Polska standaryzacja*. Warszawa: Pracownia Testów PTP.

OECD. (2009). *PISA 2009 Main Study Data Management Manual*.

Snijders, T. A. B. i Bosker, R. J. (2012). *Multilevel analysis: an introduction to basic and advanced multilevel modeling (2nd ed.)*. Los Angeles: Sage.

Rozdział 2

Trafność testów egzaminacyjnych

Zgodnie z przyjętą koncepcją badania trafności metody EWD analizy należy zacząć od sprawdzenia jakości testów egzaminacyjnych, których wyniki są wykorzystywane do szacowania gimnazjalnych wskaźników EWD. Jeśli pomiar wyników nauczania nie będzie wystarczająco dobry, nawet najbardziej wyrafinowane metody statystyczne nie dostarczą wartościowych miar. Dlatego tak ważne jest zbadanie, czy egzaminy zewnętrzne mają takie właściwości, które pozwalają z powodzeniem wykorzystywać je do wyliczania trafnej miary EWD. W literaturze przedmiotu poświęconej znaczeniu jakości danych wykorzystywanych do modelowania EWD zwraca się uwagę na różne aspekty, z których kilka wydaje się kluczowych (Chudowsky i in., 2010; OECD, 2008; por. także: Pokropek, 2013). Przede wszystkim stawia się pytanie o to, czy wykorzystane do modelowania EWD testy są dobrą miarą tego, na ile szkoła zrealizowała najważniejsze cele kształcenia. Tylko w takiej sytuacji mogą one bowiem stanowić podstawę do wyliczania miar efektywności kształcenia. Dodatkowo zwraca się uwagę na to, że testy wykorzystane do szacowania danego wskaźnika EWD powinny mierzyć ten sam konstrukt, by móc interpretować uzyskane wskaźniki jako miarę względnego przyrostu określonej umiejętności. Podkreśla się także to, że pomiar osiągnięć szkolnych musi być wystarczająco rzetelny, a jego wyniki nie powinny być obciążone działaniem czynników niezwiązanych z poziomem badanej umiejętności. Ponadto skala, na której wyrażane są wyniki pomiaru, powinna mieć wystarczające właściwości, by mogła być wykorzystana w modelowaniu EWD. Wymienione postulaty wobec jakości miar osiągnięć zostaną kolejno poddane analizie w tym rozdziale. Spełnienie ich otwiera drogę do szacowania trafnej miary EWD.

Dodatkowo niektórzy badacze zwracają uwagę na znaczenie porównywalności warunków, w których przeprowadzany jest pomiar osiągnięć oraz motywacji uczniów do rozwiązywania testów (Reckase, 2008). Polskie dane wypadają w tym świetle dobrze, gdyż egzaminy zewnętrzne są przeprowadzane wedle ściśle określonych, wystandaryzowanych procedur. Egzamin gimnazjalny jest dla uczniów tzw. testem wysokiej stawki (testem doniosłym), mającym znaczenie dla ich dalszych losów edukacyjnych. Natomiast sprawdzian w klasie szóstej szkoły podstawowej jest dla niemal wszystkich uczniów pierwszym poważnym egzaminem. Choć dla zdecydowanej większości zdających jego wyniki nie niosą ze sobą istotnych konsekwencji, ze strony szkoły występuje duży jednak nacisk na osiągnięcie w nim wysokich wyników. W związku z tym motywacja uczniów do rozwiązania tych egzaminów powinna być wysoka i podobna w różnych szkołach.

W analizach przeprowadzonych na potrzeby tego rozdziału wykorzystano dane ze sprawdzianu w klasie szóstej szkoły podstawowej przeprowadzonego w 2009 roku oraz egzaminu

gimnazjalnego zrealizowanego w 2012 roku. Dane te pozwalają wyznaczyć jednoroczne wskaźniki EWD dla kohorty uczniów objętych badaniem, którego wyniki podsumowuje ta publikacja. Analogiczne analizy dla wszystkich sesji egzaminacyjnych są oczywiście potrzebne, w szczególności dlatego, że właściwości egzaminów z kolejnych lat nie są stabilne. Jednak tak szerokie zagadnienie nie mieści się w ramach niniejszego studium.

2.1. Rzetelność i błędy pomiaru

Rozważania dotyczące jakości testów egzaminacyjnych zaczniemy od analizy rzetelności tych narzędzi, bowiem zadowalająca rzetelność jest warunkiem koniecznym dla możliwości wykorzystania wyników testowych do szacowania EWD. Rzetelność pomiaru zdaje sprawę z tego, w jakim stopniu wynik uzyskany (obserwowany) jest uzależniony od losowych czynników, niezwiązanych z poziomem cechy, którą test mierzy. Jest to więc kluczowa właściwość testu, informująca o precyzji, z jaką dokonywany jest pomiar. Jeśli rzetelność testu jest niska, niewielka jest też dokładność dokonywanego nim pomiaru. O rzetelności możemy również myśleć w kategoriach miary powtarzalności wyników pomiaru. Gdyby na przykład uczniowie przystąpili do napisania równoważnego testu w niewielkim odstępie czasu, na ile zbliżone wyniki otrzymaliby w obu testach? Jeśli testy byłyby równoważne i charakteryzowałyby się wysoką rzetelnością, rezultaty byłyby bardzo zbliżone. Jeśli niska – wyniki tych samych uczniów uzyskane w obu testach zapewne znacząco by od siebie odbiegały. Testy o niskiej rzetelności nie powinny więc służyć wyznaczaniu ani wyników indywidualnych, na podstawie których podejmowane są ważne dla uczniów decyzje, ani wskaźników mających na celu ocenę pracy szkoły (nie tylko miar EWD, ale również np. średniego wyniku egzaminacyjnego). Wystarczająco wysoką rzetelność pomiaru uznajemy za warunek konieczny dla uzyskania wiarygodnych miar jakości nauczania.

Badacze zainteresowani właściwościami psychometrycznymi polskich egzaminów zewnętrznych podsumowywali w swoich publikacjach wybrane sesje egzaminacyjne (patrz np.: Dolata, Jasińska i Modzelewski, 2012; Pokropek i Żółtak, 2012; Pokropek, 2011). Nas w szczególności interesują te egzaminy, do których przystąpili badani uczniowie. Analiza rzetelności egzaminu gimnazjalnego przeprowadzonego w 2012 roku została opisana w pracy Artura Pokropka i Tomasza Żółtaka (2012) poświęconej znaczeniu zmiany formuły tegoż egzaminu dla szacowania wskaźników EWD. Ponieważ uwzględnia ona wszystkie interesujące nas wskaźniki rzetelności, przytoczymy tu wyniki tych analiz, uzupełniając je o analogiczne miary wyliczone dla sprawdzianu szóstoklasisty z 2009 roku¹.

¹ Analizy te zostały przeprowadzone na danych populacyjnych, które są wykorzystywane do modelowania EWD (połączone wyniki sprawdzianu i egzaminu gimnazjalnego).

Tabela 1. Rzetelność egzaminów wykorzystywanych do wyliczania EWD dla gimnazjów

	wskaźnik rzetelności	sprawdzian 2009	egzamin gimnazjalny 2012					
			GH	GH-P	GH-H	GMP	GM-M	GM-P
liczba zadań		35	56	27	29	49	23	26
bez sumowania podpunktów i kryteriów	alfa	0,859	0,905	0,848	0,826	0,883	0,841	0,729
	IRT	0,891	0,915	0,865	0,848	0,912	0,877	0,776
liczba zadań		25	47	22	25	47	---	24
z sumowaniem podpunktów i kryteriów	alfa	0,757	0,879	0,757	0,822	0,882	---	0,724
	IRT	0,833	0,907	0,838	0,846	0,912	---	0,775

Źródło: dane w tabeli dot. egzaminu gimnazjalnego pochodzą z: (Pokropek i Żółtak, 2012). Dane dot. sprawdzianu – obliczenia własne. oznaczenia dla egzaminu gimnazjalnego: GH – część humanistyczna łącznie, GH-P – język polski, GH-H – historia i WOS, GMP – część matematyczno-przyrodnicza łącznie, GM-M – matematyka, GM-P – przyroda

W powyższej tabeli podano dwa wskaźniki rzetelności: klasyczną, powszechnie znaną miarę alfa Cronbacha oraz oszacowanie rzetelności na podstawie wyników modelu IRT. Alfa Cronbacha jest często stosowaną metodą szacowania dolnej granicy rzetelności danego pomiaru w ramach klasycznej teorii testu. Z tego też powodu została przez nas zaprezentowana. Nie jest to jednak metoda optymalna w sytuacji, w której w teście znajdują się zadania oceniane na skalach o różnej długości (z czym mamy do czynienia w niemal każdym egzaminie zewnętrznym). W takim wypadku wartość wskaźnika rzetelności może być niedoszacowana (Novick i Lewis, 1967). Świadomość tego faktu doprowadziła do przyjęcia konwencji wyliczania rzetelności dla testu, w którym zadania otwarte podzielono na części odpowiadające różnym kryteriom i każdą z takich części traktuje się jak osobne zadanie. Podobnie traktuje się zadania składające się z kilku podpunktów ocenianych niezależnie, choć nawiązujących do tego samego materiału, problemu. Traktowanie jednak tak utworzonych części jako osobnych zadań może prowadzić do przeszacowania wartości wskaźnika rzetelności ze względu na złamanie założenia o lokalnej (warunkowej) niezależności od siebie poszczególnych zadań i sztuczne zwiększenie długości testu.

Z tego względu w powyższej tabeli zaprezentowano wskaźniki rzetelności wyliczone dla poszczególnych testów w dwóch wariantach: w pierwszym każdy oceniany niezależnie przez egzaminatorów podpunkt czy kryterium było traktowane jak odrębne zadanie (co w tabeli nazwano: bez sumowania wiązek), w drugim zsumowano punkty zdobyte dla każdej powiązanej grupy zadań, tworząc w ten sposób zadania o dłuższej skali. Egzamin gimnazjalny

z matematyki składał się z niezależnych zadań kodowanych na skali 01, stąd podano wyniki tylko dla pierwszego rozwiązania.

Obok wskaźnika alfa Cronbacha podano alternatywną miarę rzetelności, która wykazuje się większą odpornością na różne długości skal poszczególnych zadań w teście. Jest to oszacowanie rzetelności na podstawie wyników egzaminów wyskalowanych modelami IRT. Uogólniając nieco definicję rzetelności stosowaną w klasycznej teorii testu, możemy zdefiniować rzetelność jako siłę związku liniowego (kwadrat współczynnika korelacji liniowej) między prawdziwymi wartościami mierzonej cechy a wynikami testu. Prawdziwe wartości mierzonej cechy nie są oczywiście znane, ale aby oszacować wartość takiego kwadratu współczynnika korelacji wystarczy wiedza o ich zróżnicowaniu (wariancji), nie musimy przypisywać wartości cechy poszczególnym uczniom.

Całkowitą wariancję badanej cechy możemy zdekomponować na wariancję średnich warunkowych i średnią wariancji warunkowych, w grupach wyróżnionych ze względu na obserwowany profil odpowiedzi (tj. punktację uzyskaną przez ucznia za poszczególne zadania – oznaczaną tu X):

$$D^2(\theta) = D^2[E(\theta|X)] + E[D^2(\theta|X)] \quad (1)$$

Zauważmy przy tym, że jeśli do szacowania poziomu umiejętności uczniów wykorzystywane są estymatory *expected a posteriori* (EAP), to jako oszacowanie poziomu umiejętności ucznia wykorzystywane jest właśnie przewidywanie wartości oczekiwanej rozkładu *a posteriori*:

$$\hat{\theta}_{EAP} = E(\theta|X) \quad (2)$$

Z kolei w roli analogicznej do błędu standardowego oszacowań wykorzystywane jest przewidywane odchylenie standardowe tego rozkładu:

$$BS(\hat{\theta}_{EAP}) = D(\theta|X) \quad (3)$$

Obie wymienione wartości: przewidywanie wartości oczekiwanej i odchylenia standardowego rozkładu *a posteriori* zwracane są dla każdego ucznia po wyestymowaniu modelu, możemy więc bez problemu policzyć wariancję tych pierwszych: $D^2(\hat{\theta}_{EAP})$ i średni kwadrat tych drugich: $E[BS^2(\hat{\theta}_{EAP})]$.

W związku z tym rzetelność możemy oszacować, dzieląc wariancję przewidywania cechy ukrytej (za pomocą estymatorów EAP) przez sumę tej wariancji i średniej wariancji rozkładów *a posteriori* (por. de la Torre i Patz, 2005 s. 305):

$$r = \frac{D^2[E(\theta|X)]}{D^2[E(\theta|X)] + E[D^2(\theta|X)]} = \frac{D^2(\hat{\theta}_{EAP})}{D^2(\hat{\theta}_{EAP}) + E[BS^2(\hat{\theta}_{EAP})]} \quad (5)$$

Do wyliczenia przedstawionych tu oszacowań rzetelności wykorzystane zostały wyniki z modeli IRT użytych do wyskalowania wyników egzaminu gimnazjalnego z 2012 roku i sprawdzianu z 2009 roku na potrzeby oszacowania wskaźników EWD dla gimnazjów (a więc również na potrzeby niniejszej publikacji). Zastosowany sposób skalowania został opisany w rozdziale dotyczącym organizacji i metodologii badania (część *Zmienne zależne wykorzystywane w analizach*).

Wyniki zaprezentowane w tabeli 1 pokazują dość spójny obraz. Wskaźniki alfa Cronbacha dają trochę niższe oszacowania rzetelności niż te wyliczone na podstawie modeli IRT. Rzetelność oszacowana dla testów, w których zsumowano punkty za kryteria i podpunkty dla poszczególnych zadań, tworząc zadania o dłuższych skalach, jest także trochę niższa niż ta wyliczona bez sumowania punktów. Dodatkowo alfa Cronbacha daje bardziej różniące się rezultaty dla obu sposobów liczenia rzetelności. Rzetelność wyliczona na podstawie modeli IRT, dająca bardziej stabilne wyniki, jest dla prawie wszystkich egzaminów wyższa niż 0,8 i tylko dla egzaminu gimnazjalnego z przedmiotów przyrodniczych wynosi około 0,78.

W zależności od celu pomiaru przyjmuje się różne progi dopuszczalnej minimalnej rzetelności narzędzia. Testy wykorzystywane w celach badawczych do szacowania zależności populacyjnych lub tzw. testy niskiej stawki mogą charakteryzować się nieco niższą rzetelnością ze względu na mniejsze konsekwencje popełnienia błędu w oszacowaniu poziomu danej cechy. Testy, których wyniki decydują o losach osób je piszących, powinny mieć jak najwyższą rzetelność. Przyjmuje się, że powinna ona wynosić co najmniej 0,9. Dla wykorzystania wyników testu do wyliczenia miar zagregowanych (np. dla szkół czy oddziałów), kryterium dopuszczalnej granicy rzetelności może zostać trochę obniżone. Dlatego można stwierdzić, że choć z punktu widzenia indywidualnego wyniku ucznia, w szczególności w przypadku egzaminu gimnazjalnego będącego podstawą rekrutacji do szkoły ponadgimnazjalnej, rzetelność poszczególnych testów egzaminacyjnych jest nieco za niska, to dla miar zagregowanych, takich jak EWD czy średni wynik egzaminacyjny szkoły, rzetelność analizowanych testów jest wystarczająca (ok. 0,8), by mogły być podstawą ich wyliczania.

W kontekście analiz opisywanych w dalszej części tego rozdziału warto poruszyć jeszcze jedną kwestię związaną z rzetelnością testów. Skutkiem niedoskonałej rzetelności testów

może być zaniżanie² oszacowań siły związku pomiędzy badanymi cechami, jako że użyte w analizie wyniki testów zawierają w sobie dodatkowe komponenty błędów losowych, niezwiązanych z badanymi cechami. Wynika z tego, że związki pomiędzy wynikami egzaminów a zewnętrznymi kryteriami trafności opisywane w dalszej części rozdziału mogą być w istocie nieco większe niż raportowane. Należy przy tym zastrzec, że po pierwsze, z punktu widzenia trafności metody EWD w istocie interesują nas związki pomiędzy wynikami egzaminów (oszacowaniami umiejętności uczniów), a nie cechami ukrytymi, stojącymi za tymi wynikami, gdyż modele EWD wyliczane są właśnie na podstawie takich oszacowań³. Po drugie, przeprowadzone w ograniczonym zakresie analizy danych z badania z wykorzystaniem modeli strukturalnych pokazują, że siła związków latentnych pomiędzy analizowanymi cechami niewiele przewyższa siłę zależności między oszacowaniami. Jednocześnie siła zależności latentnych jest wyraźnie niższa, niż wynikałoby to z zastosowania prostej korekty na rzetelność (*correction for attenuation*), w przypadku współczynnika korelacji liniowej polegającej na podzieleniu jego wartości przez pierwiastek z iloczynu rzetelności obu analizowanych konstruktów.

Precyzja pomiaru jest warunkiem koniecznym do zbudowania miar efektywności kształcenia, jednak nie jest warunkiem wystarczającym. Dla jakości wskaźników EWD większe znaczenie mogą mieć błędy nielosowe mogące prowadzić do obciążenia miar zagregowanych (Linn, 2008). Problem błędów nielosowych badany z perspektywy pozycji testowych można rozważać przez pryzmat tzw. analizy zróżnicowania funkcjonowania zadań (Grudniewska i Kondratek, 2012; Pokropek, 2013). Jeśli zakładamy, że zadania, z których składa się test, mierzą ten sam konstrukt, to ewentualny wpływ charakterystyk ucznia na prawdopodobieństwo poprawnego rozwiązania każdego z tych zadań powinien być w całości zapośredniczany przez wpływ tych charakterystyk na poziom badanej umiejętności. Inaczej mówiąc, jeśli dysponujemy informacją o poziomie umiejętności ucznia (w praktyce oszacowaniem jego poziomu umiejętności), to informacja o cechach ucznia, jak np. płeć czy status społeczno-ekonomiczny rodziny, nie powinny wnosić już żadnych dodatkowych informacji, użytecznych do przewidywania szans poprawnego rozwiązania zadania. Wpływ takich cech na wyniki powinien być dla wszystkich zadań taki sam i wynikać ze średnich różnic w poziomie umiejętności, występujących w ramach całej badanej zbiorowości. Występowanie w teście zadań, które wyłamują się z takiej ogólnej zależności,

² Czasem stosowane jest też określenie „rozrzędzanie”, jako bardziej dosłowne tłumaczenie angielskiego terminu *attenuation*.

³ Można wyobrazić sobie wyliczanie wskaźników EWD za pomocą modeli strukturalnych, jednak z racji bardzo dużej złożoności obliczeniowej jest to rozwiązanie obecnie niemożliwe do wdrożenia w przypadku danych populacyjnych.

funkcjonują inaczej niż inne zadania w teście, jest zjawiskiem niepożądanym i konstruując testy należy starać się takie zadania usuwać.

Niestety, procedury tworzenia testów egzaminacyjnych nie przewidują badania stronniczości zadań ze względu na cechy ucznia na etapie konstrukcji ostatecznej wersji arkusza, ani weryfikacji skali problemu po przeprowadzonej sesji egzaminacyjnej. Analizy dla wybranych testów egzaminacyjnych prowadzone przez badaczy zainteresowanych jakością tych narzędzi sugerują, że pewnym problemem może być znaczna liczba zadań na egzaminie gimnazjalnym stronniczych ze względu na płeć ucznia (Grudniewska i Kondratek, 2012; Pokropek, 2013). Pochodzenie ucznia lub wielkość miejscowości, w której znajduje się szkoła, ma z kolei mniejsze znaczenie (Pokropek, 2013). Jednak analizy te są niekompletne z tego powodu, że dotyczą tylko niektórych egzaminów przeprowadzanych w konkretnych latach.

Oddzielny problem stanowi kwestia ewentualnej stronniczości całych arkuszy egzaminacyjnych. Stronniczy test można wyobrazić sobie jako złożony w przeważającej mierze z zadań, które faworyzują pewną grupę, np. zamieszkałych na wsiach względem mieszkających w miastach. W takiej sytuacji szkoły, w których uczy się większa proporcja uczniów, dla których test był przychylny, zyskiwałyby na wartości miary EWD nie dlatego, że uczą efektywniej, ale ze względu na stronniczość testu. Niestety, w praktyce stwierdzenie, czy dany test jest stronniczy, jest rzeczą trudną i często kontrowersyjną. Jeśli test składa się z zadań „jednorodnie stronniczych”, procedury statystyczne służące wyszukiwaniu zróżnicowania funkcjonowania zadań nie znajdują tu zastosowania. Pomocne jest oczywiście zbadanie związków pomiędzy mierzonym przez test konstruktem a interesującymi nas charakterystykami uczniów. Niestety określenie kryteriów oceny, jaka siła i kierunek tych związków mogą zostać uznane za naturalne, a jakie za wskazujące na niepożądaną stronniczość, jest już kwestią dyskusyjną i nawet w środowisku specjalistów od dziedziny, jaką mierzy dany test, mogą występować znaczące różnice opinii. Koniecznym uzupełnieniem jest więc analiza treściowa zadań wchodzących w skład testu, np. w celu sprawdzenia, czy zachowana została równowaga w zakresie kontekstów zadań. Oczywiście dobrą metodą unikania stronniczości testów jest zaplanowanie zrównoważonego składu testu w zakresie kontekstu zadań już na etapie przygotowywania planu testu. Choć kwestia ewentualnej stronniczości egzaminów jest istotna z punktu widzenia własności wyliczanych na ich podstawie wskaźników EWD, to niestety, z przedstawionych wyżej powodów określenie, czy polskie egzaminy są stronnicze, wykracza dalece poza ramy tego rozdziału. Należy jednak dodać, że dopóki porównujemy szkoły nie różniące się znacząco między sobą np. cechami środowiska takimi jak wielkość miejscowości, ewentualne nielosowe błędy pomiaru nie będą podważały trafności wyciąganych na podstawie wskaźników EWD wniosków.

Kolejnym wartym rozpatrzenia problemem związanym ze znaczeniem dokładności pomiaru dla wskaźników EWD jest zależność błędu pomiaru od poziomu badanej cechy. Klasyczna teoria testu, posługując się jedną miarą rzetelności, przyjmuje założenie, że precyzja pomiaru jest identyczna dla każdego ucznia niezależnie od poziomu jego umiejętności. Założenie to nie jest jednak w pełni poprawne. W modelach IRT przyjmuje się, że precyzja pomiaru cechy ucznia jest uzależniona nie tylko od liczby rozwiązanych przez niego zadań i ich jakości pomiarowej, ale także od tego, jak dobrze pod względem trudności dobrano zadania, które przed uczniem postawiono (De Ayala, 2009). Dane zadanie dostarczy bowiem najwięcej informacji potrzebnej do oszacowania poziomu cechy ucznia, jeśli jego trudność będzie dobrana tak, by uczeń miał 50% szansy na poradzenie sobie z nim. Dla przykładu, jeśli w teście znajduje się niewiele zadań bardzo trudnych, poziom umiejętności uczniów bardzo zdolnych będzie oszacowany na podstawie niewielkiej ilości informacji, czyli z mniejszą precyzją. Zdecydowana większość testów wiadomości i umiejętności jest skonstruowana w taki sposób, że z najmniejszym błędem szacuje wyniki uczniów o zbliżonym do przeciętnego poziomie danej cechy, a na krańcach skali precyzja pomiaru jest mniejsza. Wynika to z rozkładu trudności zadań w stosunku do rozkładu umiejętności badanej populacji (niewielka liczba zadań skrajnie łatwych i skrajnie trudnych). Egzamin zewnętrznego nie odbiegają znacząco od tej reguły. Jakie mogą być konsekwencje tego dla miar zagregowanych? Wskaźniki dla szkół, których znaczna część uczniów osiąga skrajnie wysokie lub skrajnie niskie wyniki, mogą być obciążone większym błędem niż te, które są szacowane dla szkół o przeciętnych wynikach (por. Chudowsky i in., 2010). Tak jak w przypadku większości statystyk: na przypadki odstające należy patrzeć z większą dozą nieufności.

2.2. Reprezentatywność treści nauczania

O testach osiągnięć szkolnych bardzo często mówi się, że są niekompletną miarą wyników kształcenia ze względu na to, że w każdym teście znajdują się zadania mierzące tylko niewielką część nauczanych przez szkoły treści i umiejętności. Dzieje się tak ze względu na ograniczoną długość testu (praktycznie niemożliwe jest, by nawet w serii testów, którą może rozwiązać jeden uczeń, pojawiły się pytania dotyczące wszystkich treści objętych programem nauczania) oraz jego formę (test pisemny składający się przede wszystkim z zadań zamkniętych i zadań wymagających udzielenia krótkiej odpowiedzi), która sprzyja pomiarowi pewnych określonych kompetencji, a uniemożliwia pomiar innych. Ponadto testy obejmujące pomiarem wszystkich lub większość uczniów w kraju nie mierzą (przynajmniej wprost) takich ważnych wyników kształcenia jak na przykład motywacja, wytrwałość w rozwiązywaniu problemów,

ciekawość poznawcza, umiejętność współpracy z innymi czy pozytywne nastawienie do nauki. Nie oznacza to jednak, że testy osiągnięć mierzą umiejętność rozwiązywania tylko tych zadań, które zostały w nich zawarte – jeśli tylko są dobrze skonstruowane.

Pomiar osiągnięć szkolnych opiera się na założeniu, że reakcja ucznia na zadanie (poprawne lub błędne jego rozwiązanie) jest przejawem cechy ukrytej: badanej umiejętności. Zakładamy, że cecha ta jest ciągła i nie daje się zmierzyć bezpośrednio. Oznacza to, że poprawne lub niepoprawne rozwiązanie pojedynczego zadania jest bardzo niedoskonałym wskaźnikiem poziomu umiejętności ucznia. Aby móc go lepiej określić, musimy wykorzystać większą liczbę zadań. Oczywiście, co do wszystkich tych zadań musimy móc założyć, że mierzą one pewną wspólną, ogólniejszą umiejętność, wpływającą na szanse poprawnego rozwiązania każdego z nich. Należy przy tym zaznaczyć, że założenie o istnieniu takiej cechy daje się weryfikować empirycznie. Jeśli istnieje, w wynikach egzaminów powinniśmy zaobserwować pewnego rodzaju zależności statystyczne pomiędzy poszczególnymi zadaniami (por. Jakubowski i Pokropek, 2009).

Interesującą nas cechą, dla przykładu, może być znajomość języka obcego. Umiejętność tę możemy sobie wyobrazić jako cechę rozciągającą się od kompletnej nieznajomości języka do biegłego posługiwania się nim w najróżniejszych sytuacjach. To, czy badana osoba poprawnie rozwiąże zadanie, czy to wymagające znajomości określonych pojęć, czy struktur gramatycznych, zależy od jej biegłości w tym języku. Możemy jednak z dużym prawdopodobieństwem założyć, że jeśli dana osoba poradziła sobie z zadaniem wymagającym napisania złożonego tekstu, poradziłaby sobie z prostym zadaniem wymagającym przedstawienia się osobie nieznajomej. Na podstawie dowodów opanowania bardziej złożonych umiejętności, możemy wyciągnąć wniosek o posiadaniu mniej złożonych kompetencji (nawet bez weryfikowania tego zadaniami w teście). Podobnie, jeśli osoba wykazała się umiejętnością napisania listu formalnego, możemy się spodziewać, że potrafiłaby napisać list do przyjaciela.

Przykład ten miał pomóc wyrobić intuicję, że test może mierzyć zdolność do poradzenia sobie z pewnymi problemami, nawet jeśli nie zawiera zadań ściśle się do nich odwołujących. Nie oznacza to jednak, że twórcy narzędzia mogą mieć zupełną dowolność w doborze zadań. Wręcz przeciwnie. Tylko przemyślany plan testu (zawierający specyfikację tego, jakie zadania – mierzące jakie treści i umiejętności – powinny wejść w jego skład) daje szansę na to, że test będzie reprezentatywny dla danej dziedziny. Oznacza to, że będzie on pokrywał zakres treści najważniejszych z punktu widzenia pomiaru badanej cechy oraz że dobrane zadania będą na tyle różnorodne, że nie będą zbyt wąsko mężyły badanej umiejętności. Jeśli w teście umiejętności posługiwania się językiem obcym znajdowałyby się tylko zadania na tłumaczenie słów, nie można byłoby twierdzić, że

mierzy on kompetencję tworzenia z tych pojęć odpowiednich struktur umożliwiających komunikację.

Testy osiągnięć zawsze obejmują więc najważniejsze (z punktu widzenia konstruktorów narzędzia), ale nadal wybrane treści kształcenia. Skoro tak, to powstała na ich podstawie miara EWD także będzie odzwierciedlać efektywność nauczania w zakresie mierzonym przez wykorzystane do jej wyliczenia testy. Jeśli pomiar testem nie pokryje najważniejszych z punktu widzenia programu nauczania treści kształcenia, miara EWD także będzie ograniczoną, a w pewnych sytuacjach może nawet mylącą informacją na temat efektywności kształcenia. Skoro bowiem wskaźniki EWD mają być miarą efektywności pracy szkoły, to do ich wyliczania powinny zostać wykorzystane testy, które koncentrują się na pomiarze tego, co należy do zadań szkoły, czyli na pomiarze realizacji celów kształcenia. Jeśli na przykład wymaga się, by szkoły rozwijały myślenie naukowe, a testy wykorzystane do wyznaczenia wskaźników EWD sprawdzałyby tylko umiejętność odtwarzania faktów, ich wyniki nie byłyby dobrą podstawą dla miar efektywności nauczania w określonym przez wymagania zakresie. Faworyzowałyby szkoły, które przede wszystkim koncentrują się na utrwalaniu faktów i pojęć. Należy mieć jednak świadomość, że ograniczenie to stosuje się do każdej miary jakości kształcenia wykorzystującej wyniki danych testów, nie tylko miary EWD. Niemniej trzeba mieć na uwadze fakt, że wskaźniki EWD są czułe na wykorzystane do ich wyliczenia wyniki testów, a wybór określonej miary osiągnięć może mieć większe znaczenie dla wartości wyliczonych wskaźników niż wybór konkretnego modelu, na podstawie którego są wyznaczane (Lockwood i in., 2007).

W polskim systemie edukacji nie mamy dużego wyboru miar wyników kształcenia, które mogą być wykorzystane do wyliczenia wskaźników EWD. Nie zwalnia to nas jednak z obowiązku oceny jakości wykorzystywanych wyników egzaminacyjnych dla potrzeb szacowania wskaźników efektywności pracy szkół. Zadbanie o reprezentatywność mierzonych umiejętności szczegółowych, ważnych z punktu widzenia celów kształcenia na danym etapie edukacji, spoczywa na Centralnej Komisji Egzaminacyjnej, która jest odpowiedzialna za ostateczny kształt egzaminów zewnętrznych. Dokumentem regulującym zakres treściowy pomiaru dla egzaminu gimnazjalnego od 2012 roku jest podstawa programowa kształcenia ogólnego. W „Informatorze o egzaminie gimnazjalnym” (Centralna Komisja Egzaminacyjna, 2010) znajdujemy deklarację, że „egzamin będzie sprawdzał opanowanie przez uczniów wiadomości i umiejętności określonych w wymaganiach ogólnych i szczegółowych zawartych w podstawie programowej kształcenia ogólnego”. W dokumentach opublikowanych przez Centralną Komisję Egzaminacyjną po zrealizowanej sesji egzaminacyjnej informację tę uzupełniono dość ogólną charakterystyką skal pomiarowych: „Egzamin gimnazjalny w nowej formule umożliwia bowiem określenie,

w jakim stopniu uczeń opanował kluczowe dla dalszego rozwoju umiejętności złożone takie jak m.in. krytyczne myślenie, precyzyjne rozumowanie oraz analizowanie przebiegu zjawisk. Umiejętności te stanowią rdzeń kształcenia gimnazjalnego; dla każdego przedmiotu zostały one określone w aktualnie obowiązującej podstawie programowej w postaci tzw. wymagań ogólnych. Umiejętności te mierzy się za pomocą zadań zakotwiczonych w konkretnych zapisach podstawy programowej. Uzyskano w ten sposób pełną spójność wymagań egzaminacyjnych z celami procesu kształcenia, a zarazem postawiono w centrum uwagi najważniejsze cele tego procesu” (Centralna Komisja Egzaminacyjna, 2012). W przypadku sprawdzianu realizowanego w 2009 roku, czyli przed wprowadzeniem zmian programowych, zakres treści i umiejętności objętych pomiarem, regulują publicznie dostępne standardy wymagań egzaminacyjnych oraz podstawa programowa kształcenia ogólnego (Informator z aneksem dla uczniów ze specyficznymi trudnościami w uczeniu się, 2009).

Przedstawione powyżej założenia pomiaru pozwalają wierzyć, że egzaminy zewnętrzne są dobrą miarą tego, na ile szkoła zrealizowała stawiane przed nią cele. Analiza tego, na ile w praktyce treść egzaminów odpowiada sformułowanym założeniom, powinna polegać na krytycznej ocenie – przez ekspertów z zakresu dydaktyki – planów testów, precyzujących, jaki odsetek zadań w teście będzie mierzył konkretne umiejętności i treści. Dodatkowo powinna ona objąć ocenę skali odstępstw od planów testów ostatecznych narzędzi pomiarowych⁴. Dokumentacja z takich analiz nie jest publicznie dostępna, co utrudnia ocenę, czy konkretne testy nie pomijają pewnych kluczowych umiejętności, które są wyszczególnione w podstawie programowej. Prace nad rozwojem metodologii szacowania wskaźników EWD byłyby łatwiejsze i potencjalnie bardziej owocne, gdyby kolejne sesje egzaminów zewnętrznych doczekały się stosownego opracowania w postaci raportów technicznych opisujących ich właściwości psychometryczne. Niektóre założenia dotyczące narzędzi mogą bowiem zostać adekwatnie opisane przede wszystkim przez osoby konstruujące dany test i są trudne do odtworzenia post factum. Mimo swoich niedostatków w postaci odpowiedniej dokumentacji, egzaminy nie powstają jednak w sposób przypadkowy. Procedury ich tworzenia są przestrzegane dzięki wiedzy i doświadczeniu pracowników centralnej i okręgowych komisji egzaminacyjnych, którzy odpowiadają za ich jakość. Pozwala to mieć nadzieję, że polskie egzaminy są dobrą miarą tego, na ile szkoła zrealizowała stawiane przed nią przez państwo cele, i w konsekwencji stanowią dobrą podstawę dla wskaźników EWD.

⁴ Odstępstwa takie mogą się zdarzyć np. w przypadku zastąpienia zadania o bardzo słabych właściwościach psychometrycznych innym, lepszym pomiarowo, choć odwołującym się do innych umiejętności.

2.3. Zewnętrzne kryteria trafności egzaminów gimnazjalnych

Utrudniona weryfikacja zgodności treściowej egzaminów z celem pomiaru sprawia, że analiza związków egzaminów z zewnętrznymi kryteriami stanowi jeszcze bardziej wartościowy wkład w ocenę trafności egzaminów. Umożliwia ona weryfikację, na ile testy, na podstawie których liczone są wskaźniki EWD, pozwalają mierzyć coś więcej niż tylko skuteczność rozwiązania danego zestawu zadań. Idąc dalej, możemy zapytać, na ile wyniki egzaminów są powiązane z szerzej ujętymi efektami kształcenia i czy dzięki temu mogą stanowić podstawę bardziej kompleksowej oceny pracy szkoły. We współczesnym myśleniu o trafności pomiaru wychodzi się poza jej kryterialne znaczenie, które koncentruje się na tym, w jakim stopniu test mierzy to, co w założeniu miał mierzyć. Coraz częściej sięga się po modele definiujące trafność w kategoriach użyteczności wykorzystania wyników testu w określonym celu (Skorupiński, 2013). Zgodnie z tym ujęciem możemy zapytać, czy wynikami egzaminów zewnętrznych możemy się sensownie posługiwać jako miarą wyników kształcenia.

Przeprowadzone badanie umożliwiło weryfikację tego, czy wyniki polskich egzaminów są związane z innymi miarami podobnych konstruktów (ukrytych dyspozycji umysłowych, umiejętności) oraz zmiennymi opisującymi czynniki, co do których możemy oczekiwać, że są powiązane z umiejętnościami, które z założenia mają mierzyć egzaminy. Pokażemy także, na ile wyniki testu są skorelowane z szerzej ujętymi efektami kształcenia. Odniesiemy się do sześciu kryteriów: poziomu inteligencji, ocen szkolnych, bezradności intelektualnej, wzorca motywacji, dwóch wymiarów integracji uczniów ze środowiskiem szkolnym: integracji emocjonalnej i motywacyjnej.

Poziom inteligencji uczniów był mierzony testem matrycy Ravena w wersji standardowej w pierwszej klasie gimnazjum⁵. Ponieważ znaczeniu inteligencji dla osiągnięć szkolnych poświęcony jest w tej publikacji osobny rozdział, wnioski ani szczegółowy opis narzędzia nie będą tu powtarzane. Oceny szkolne na koniec pierwszego semestru trzeciej klasy gimnazjum były zbierane za pomocą pytań w ankiecie uczniowskiej. W niniejszej analizie zostały wykorzystane informacje o ocenach z języka polskiego i matematyki. Poziom bezradności intelektualnej był mierzony w trzeciej klasie gimnazjum za pomocą standardowej skali bezradności intelektualnej opisanej w pracy Grzegorza Sędko (1995). Na tym samym etapie badania przeprowadzono pomiar motywacji, wykorzystując wstępną adaptację skali ASRS. Sposób konstrukcji wskaźników poziomu

⁵ W badaniu powtórzono pomiar testem matrycy Ravena także w III klasie gimnazjum, ale wyniki te nie będą tu wykorzystywane.

bezradności intelektualnej i wzorców motywacji oraz opis wykorzystanych narzędzi został przedstawiony w aneksie do rozdziału 6. Poziom integracji emocjonalnej i motywacyjnej był mierzony w trzeciej klasie gimnazjum za pomocą kwestionariusza integracji uczniów (KIU). Narzędzie to jest adaptacją szwajcarskiego kwestionariusza *Fragebogen zur Erfassung von Dimensionen der Integration von Schülern FDI 4–6* (Haeberlin, Moser i Bless, 1989), służącego do badania funkcjonowania szkolnego uczniów. Polska adaptacja narzędzia została opracowana przez Grzegorza Szumskiego. Interesujące nas, z punktu widzenia analizowanego problemu, wymiary szkolnego funkcjonowania, to 1) integracja emocjonalna wyrażająca nastawienie uczniów do szkoły, to czy uczniowie lubią do niej chodzić, czy szkoła kojarzy im się z czymś przyjemnym, ciekawym oraz 2) integracja motywacyjna, która odnosi się do obrazu siebie jako ucznia za pomocą samooceny własnych szkolnych kompetencji. Sposób konstrukcji wskaźników dla tych skal został opisany w aneksie do niniejszego rozdziału.

W poniższej tabeli przedstawiono siłę związków wymienionych kryteriów z czterema wynikami egzaminów gimnazjalnych: łącznym wynikiem części matematyczno-przyrodniczej, łącznym wynikiem części humanistycznej, wynikiem egzaminu z matematyki i języka polskiego. Dla zmiennych ciągłych (mierzonych na skali interwałowej) podano współczynnik korelacji liniowej i jego kwadrat (współczynnik determinacji r^2), mówiący o tym, jaką część zmienności wyników egzaminacyjnych możemy wyjaśnić przez dane kryterium. Podano także błąd standardowy dla oszacowania korelacji. Dla zmiennych porządkowych (oceny szkolne) i nominalnych (wzorce motywacji) siłę związku wyrażono współczynnikiem η^2 , mówiącym o tym, jaką część wariancji zmiennej zależnej możemy wyjaśnić przez przynależność do poszczególnych kategorii zmiennej niezależnej (jest to miara o analogicznej interpretacji jak r^2). Dodatkowo podano wartości współczynników regresji pozwalające określić, jakie były średnie wyniki egzaminów wśród uczniów o określonej wartości zmiennej niezależnej. Współczynniki istotne statystycznie na poziomie istotności $p < 0,05$ zaznaczono pogrubionym drukiem. Prezentowane parametry zostały wyliczone za pomocą modeli regresji estymowanych w programie Mplus 7 metodą *pseudo maximum likelihood*, z uwzględnieniem dwustopniowego schematu doboru próby i nierównych prawdopodobieństw doboru.

Pierwszym z rozpatrywanych kryteriów, z którym według teorii i licznych badań powiązane są osiągnięcia szkolne, jest inteligencja. Wyniki przeprowadzonych analiz potwierdziły istnienie tego związku. Jego siła jest zbliżona do zaobserwowanej w innych badaniach, a także zgodnie z doniesieniami innych prac, jest trochę silniejsza dla przedmiotów matematyczno-przyrodniczych niż humanistycznych (patrz przegląd literatury w rozdziale 4).

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 2. Siła związków między wynikami egzaminów gimnazjalnych a zewnętrznymi kryteriami

kryterium	współczynnik	GMP	GH	GM-M	GH-P
test matryc Ravena	r^2	0,404	0,315	0,383	0,284
	r	0,636	0,561	0,619	0,533
	s.e.	0,013	0,013	0,014	0,013
oceny na koniec I semestru III klasy z języka polskiego (kat. odniesienia = niedostateczny)	η^2	0,323	0,419	0,288	0,407
	s.e.	0,019	0,021	0,018	0,020
	b1	88,34	85,05	89,00	84,02
	b2	2,27	3,43	2,41	4,81
	s.e.	1,38	1,48	1,43	1,77
	b3	8,81	12,14	8,49	13,86
	s.e.	1,45	1,57	1,52	1,81
	b4	17,98	22,70	17,17	23,87
	s.e.	1,57	1,56	1,62	1,76
	b5	27,11	30,60	25,41	30,36
	s.e.	1,69	1,75	1,68	1,84
	b6	22,20	25,80	20,36	24,81
	s.e.	3,14	3,01	2,85	2,91
	η^2	0,530	0,413	0,514	0,353
oceny na koniec I semestru III klasy z matematyki (kat. odniesienia = niedostateczny)	s.e.	0,018	0,017	0,018	0,017
	b1	85,60	84,16	86,63	84,30
	b2	4,12	5,90	3,37	6,86
	s.e.	0,78	0,95	0,80	1,15
	b3	12,68	15,68	11,89	16,13
	s.e.	0,89	1,14	0,90	1,26
	b4	23,97	25,34	23,31	24,87
	s.e.	1,01	1,21	1,01	1,32
	b5	34,14	30,66	32,44	28,31
	s.e.	1,23	1,27	1,13	1,41
	b6	34,98	29,24	31,16	25,24
	s.e.	3,07	2,75	2,66	2,53

kryterium	współczynnik	GMP	GH	GM-M	GH-P
bezzadność intelektualna – język polski	r^2	0,008	0,057	0,004	0,069
	r	-0,090	-0,238	-0,064	-0,262
	bs	0,019	0,018	0,018	0,018
bezzadność intelektualna – matematyka	r^2	0,135	0,068	0,136	0,047
	r	-0,367	-0,260	-0,369	-0,217
	s.e.	0,018	0,017	0,017	0,017
wzorec motywacji (kat. odniesienia = motywacja o przeciętnej sile)	η^2	0,033	0,033	0,029	0,034
	s.e.	0,007	0,006	0,006	0,006
	b1	98,91	98,74	99,24	98,84
	b2 (niska)	0,75	0,10	0,71	-0,64
	s.e.	0,90	1,02	0,89	1,06
	b3 (wysoka)	1,88	2,55	1,63	3,09
	s.e.	0,71	0,75	0,72	0,79
	b4 (zewnątrzna)	-0,48	-0,22	-0,68	0,18
	s.e.	0,64	0,71	0,64	0,69
	b5 (wewnętrzna)	7,68	7,50	6,97	7,17
	s.e.	0,95	0,77	0,89	0,70
integracja emocjonalna	r^2	0,284	0,268	0,261	0,220
	r	0,533	0,518	0,511	0,469
	bs	0,014	0,013	0,014	0,013
integracja motywacyjna	r^2	0,030	0,040	0,025	0,041
	r	0,172	0,199	0,157	0,203
	s.e.	0,020	0,019	0,020	0,019

oznaczenia: GMP – część matematyczno-przyrodnicza łącznie, GH – część humanistyczna łącznie, GM-M – matematyka, GH-P – język polski

W przypadku badania trafności testów osiągnięć szkolnych, jako zewnętrzne kryterium często wykorzystuje się oceny nauczycielskie lub wyniki innych testów osiągnięć (Anastasi i Urbina, 1999). Oceny nauczycielskie są często miarą, ze statystycznego punktu widzenia, mało rzetelną, zasadniczo nieporównywalną między grupami uczniów ocenianymi przez różnych nauczycieli (por. Jasińska, 2010). Mimo to ich wykorzystanie w badaniu trafności jest uzasadnione o tyle, że obejmują one szerokie spektrum umiejętności uczniów, które mają okazję ujawnić się w różnych okolicznościach (nie tylko podczas badania testowego). Dodatkowo są one w mniejszym stopniu uzależnione od chwilowej

dyspozycji ucznia. Z tego powodu zostały przez nas wykorzystane w przeprowadzonych analizach.

Wyniki analiz związków z ocenami szkolnymi przemawiają na rzecz trafności egzaminów zewnętrznych. Obserwujemy bowiem umiarkowanie silny, pozytywny związek wyników egzaminacyjnych z ocenami z przedmiotów, które są objęte danym egzaminem (ok. 40–50% wariancji wyników egzaminacyjnych można wyjaśnić za pomocą informacji o ocenie z danego przedmiotu na koniec I semestru trzeciej klasy). Dodatkowym potwierdzeniem jest silniejsza relacja dla odpowiadających sobie przedmiotów (np. między wynikiem egzaminu z matematyki a ocenami z tego przedmiotu) niż dla różnych przedmiotów (np. między wynikiem egzaminu z matematyki a ocenami z języka polskiego). Oczywiście te krzyżowe relacje także występują z uwagi na to, że osiągnięcia szkolne z matematyki i języka polskiego są ze sobą powiązane.

Za kolejne kryterium trafności egzaminów zewnętrznych, jako narzędzia pomiaru efektów kształcenia, posłuży związek z bezradnością intelektualną uczniów. Stan bezradności intelektualnej, według twórców narzędzia, może powstać wskutek niewłaściwego sposobu nauczania / uczenia się lub zaburzenia interakcji nauczyciel – uczeń (Sędek, 1995), jeśli prowadzą one do tego, że uczeń mimo długotrwałego wysiłku umysłowego nie jest w stanie zrozumieć danego materiału. Doświadczenia takie mogą prowadzić do zablokowania możliwości aktywnego myślenia problemowego i pogorszenia zdolności przyswajania nowych informacji do istniejących już struktur wiedzy. Stan taki stanowi istotę bezradności intelektualnej. Jest więc ona ważnym czynnikiem wpływającym na niskie osiągnięcia, ponieważ znacząco utrudnia lub nawet uniemożliwia uczenie się ze zrozumieniem, pozwalając jedynie na uczenie się reprodukcyjne, pamięciowe, które i tak może być utrudnione. Co ważne, przyjmuje się, że bezradność intelektualna nie jest ogólną cechą ucznia, ale stanem specyficznym dla danej dziedziny, przedmiotu szkolnego. W kontekście tak pojmowanej bezradności intelektualnej, jeśli egzaminy zewnętrzne są trafną miarą wyników kształcenia, mierzą coś więcej niż umiejętność reprodukcji wiadomości, oczekivalibyśmy po pierwsze negatywnego związku wyników egzaminacyjnych z poziomem bezradności. Po drugie, jako że bezradność intelektualna jest specyficzna dla dziedziny, spodziewaliśmy się silniejszych związków dla konstruktów odnoszących się do tych samych przedmiotów.

W przypadku bezradności intelektualnej z języka polskiego otrzymaliśmy relacje zgodne z przewidywaniami. Związki z wynikami egzaminów gimnazjalnych z języka polskiego i łącznymi wynikami z części humanistycznej są negatywne, choć niezbyt silne. Związki z wynikami egzaminów z matematyki, choć istotne i negatywne, są bardzo słabe. W przypadku bezradności intelektualnej z matematyki zanotowano jeszcze silniejsze,

negatywne związki z wynikami egzaminu z matematyki i łącznymi wynikami z przedmiotów matematyczno-przyrodniczych. Co ciekawe, zaobserwowano także znaczące (choć słabsze), ujemne korelacje z wynikami z egzaminów z przedmiotów humanistycznych i z języka polskiego. Oznacza to, że albo bezradność intelektualna nie jest tak ściśle specyficzna dla danego przedmiotu jak zakładali twórcy koncepcji tego konstruktu, albo egzamin z języka polskiego i przedmiotów humanistycznych mierzy także takie kompetencje, których rozwój utrudnia bezradność intelektualna z matematyki. Niemniej wzór zależności jest zgodny z oczekiwaniami i stanowi kolejny argument przemawiający za trafnością egzaminów gimnazjalnych jako miary efektów kształcenia.

Następnym rozpatrywanym przez nas kryterium jest motywacja do uczenia się⁶. Osiągnięcia szkolne wydają się być powiązane nie tylko z siłą motywacji, ale także jej jakością (Vansteenkiste, Sierens, Soenens, Luyckx i Lens, 2009). W szczególności zwraca się uwagę na znaczenie motywacji wewnętrznej i zewnętrznej (Ryan i Deci, 2000). Motywacja wewnętrzna dotyczy sytuacji, w której dane działania są podejmowane dla nich samych, a nie ze względu na jakiś instrumentalny cel. Osoba wykonuje jakąś aktywność dlatego, że jest nią zainteresowana, przynosi jej to satysfakcję, wzbudza zaangażowanie. Natomiast motywacja zewnętrzna dotyczy podejmowania działań w celu spełnienia wymagań zewnętrznych wobec jednostki, uniknięcia kar lub uzyskania konkretnych korzyści. Wiąże się ona z poczuciem presji i przymusu. Jeśli wyniki egzaminacyjne miałyby być trafną miarą efektów kształcenia, oczekivalibyśmy pozytywnych związków z motywacją wewnętrzną oraz raczej negatywnych związków (lub braku związku) z motywacją zewnętrzną. Chcielibyśmy bowiem, by wysokie wyniki egzaminacyjne częściej osiągalni uczniowie, którzy uczą się dlatego, że jest to dla nich ważne, wartościowe, a nie dlatego, że odczuwają zewnętrzny przymus i chcą uniknąć sankcji. Za tym oczekiwaniem stoi przeświadczenie, że szkoły mogą kształtować wzorzec motywacji, co nie jest sprzeczne z założeniami twórców koncepcji (Ryan i Deci, 2000). Traktujemy tu więc motywację jako alternatywne kryterium efektów kształcenia.

Wyniki przeprowadzonych analiz pokazały, że wzorzec motywacji wyjaśnia ok. 3% różnicowania wyników egzaminów gimnazjalnych, jednak dzieje się to głównie za sprawą odmienności dwóch spośród pięciu wyróżnionych wzorców. W stosunku do uczniów charakteryzujących się motywacją o przeciętnej sile, uczniowie o wysokiej motywacji uzyskali na egzaminie wyniki o 2–3 punkty wyższe, a efekt ten był podobny dla wszystkich przedmiotów. Natomiast uczniowie charakteryzujący się motywacją wewnętrzną otrzymali na egzaminach wyniki o około 7 punktów (ok. pół odchylenia standardowego) wyższe. Uczniowie o motywacji zewnętrznej oraz o niskiej motywacji nie różnili się statystycznie

⁶ W odróżnieniu od motywacji do rozwiązywania egzaminów, o której była mowa we wprowadzeniu do rozdziału.

wynikami egzaminów od uczniów o motywacji przeciętnej. Wyniki te pokazują, że egzaminy zewnętrzne pośrednio zdają też sprawę z tego, jaką motywacją kierują się uczniowie je zdający, a zaobserwowane związki są zgodne z naszymi oczekiwaniami, choć niezbyt silne.

Ostatnimi z rozpatrywanych kryteriów są: nastawienie gimnazjalisty do szkoły i postrzeganie siebie jako ucznia. Twierdzimy bowiem, że dobra szkoła powinna nie tylko dbać o jak najwyższe osiągnięcia mierzone egzaminami, ale równocześnie powinna rozwijać u uczniów takie umiejętności i postawy, które przełożą się na sprawne funkcjonowanie szkolne i pozytywne nastawienie do nauki i szkoły oraz wiarę we własne możliwości uczenia się. Przekonania takie mogą zwiększyć prawdopodobieństwo podjęcia nauki na kolejnych szczeblach edukacji oraz osiągania dalszych szkolnych sukcesów. Jeśli wyniki egzaminów gimnazjalnych pośrednio informują o szerzej ujętych efektach kształcenia, powinniśmy zaobserwować pozytywny ich związek z poziomem integracji emocjonalnej i motywacyjnej.

Integracja emocjonalna wyraża nastawienie uczniów do szkoły. Skala użyta w badaniu do jej pomiaru składała się z 16 pozycji zawierających stwierdzenia takie jak na przykład: *Lubię chodzić do szkoły*; *W szkole jest ciekawie*; *Cieszę się, że kiedyś nie będę już musiał chodzić do szkoły* (pozycja odwrócona); *Bez szkoły wszystko byłoby o wiele lepsze* (pozycja odwrócona); *Czasami chętnie bym poszedł na wagary* (pozycja odwrócona); *Wspaniale by było, gdybym nie musiał już chodzić do szkoły* (pozycja odwrócona). Przeprowadzone analizy pokazały dość silny związek wyników egzaminacyjnych z tym wymiarem funkcjonowania szkolnego. Współczynniki korelacji dla czterech rozpatrywanych wyników egzaminacyjnych wynoszą około 0,5.

Integracja motywacyjna, wyrażająca obraz siebie jako ucznia za pomocą samooceny własnych szkolnych kompetencji, była mierzona skalą, w skład której weszło 14 pozycji takich jak na przykład: *Nawet gdy bardzo się staram, nie umiem zrobić trudnych zadań* (pozycja odwrócona); *Kiedy nauczyciel tłumaczy nową lekcję, od razu wszystko rozumiem*; *Szybko się uczę*; *W szkole w wielu sprawach jestem mało zdolny* (pozycja odwrócona); *Wielu szkolnych zadań nie umiem zrobić* (pozycja odwrócona). Korelacja między tą skalą a wynikami egzaminacyjnymi jest słabsza niż dla integracji emocjonalnej, jednak nadal pozytywna i istotna statystycznie. Okazała się ona trochę silniejsza dla języka polskiego i łącznego wyniku z przedmiotów humanistycznych (ok. 0,2) i trochę słabsza dla matematyki i wyniku z przedmiotów matematyczno-przyrodniczych (ok. 0,16).

Zaprezentowane w tej części rozdziału wyniki analiz przemawiają na rzecz trafności egzaminów gimnazjalnych jako miary osiągnięć szkolnych. Potwierdzono istnienie związków wyników egzaminacyjnych z innymi konstruktami, z którymi – wedle oczekiwań – powinny być one związane. Dodatkowo wykazano ich związek z szerzej ujętymi efektami kształcenia takimi jak motywacja, funkcjonowanie szkolne czy brak bezradności intelektualnej.

Oznacza to, że wyniki egzaminów gimnazjalnych są miarą nie tylko tego, jak uczeń poradził sobie z rozwiązaniem danego zestawu zadań, ale zdają sprawę z ogólnego poziomu wiadomości i umiejętności z danej dziedziny przedmiotowej, a także pośrednio niosą za sobą informację o skuteczności realizacji szerzej rozumianych celów nauczania.

2.4. Spójność mierzonych konstruktów

Wskaźniki EWD służą ocenie względnego przyrostu umiejętności uczniów w szkołach między dwoma momentami pomiaru: przed rozpoczęciem nauki w danej szkole i w chwili jej zakończenia. Skoro mają one dostarczyć informacji na temat relatywnej zmiany poziomu umiejętności w tym okresie, oba testy wykorzystywane do pomiaru osiągnięć powinny mierzyć ten sam konstrukt (umiejętność). Oczywiście jest, że jeśli chcemy na przykład oszacować względny przyrost umiejętności czytania, powinniśmy porównać poziom tej właśnie umiejętności na początku i na końcu okresu, który nas interesuje. Najlepiej byłoby tego dokonać, wykorzystując dokładnie ten sam test. Wtedy mielibyśmy pewność, że w obu punktach czasowych mierzymy dokładnie tę samą umiejętność. Tu pojawia się jednak problem niedostosowania trudności testu do badanej populacji. Ta bowiem rozwija swoje umiejętności przez dany okres nauki (np. trzy lata spędzone w gimnazjum), więc test, który był odpowiednio trudny dla uczniów na początku tego okresu, może okazać się za łatwy na jego zakończenie (czego konsekwencją będzie nieprawidłowe oszacowanie wyników uczniów o najwyższym poziomie umiejętności). Rozwiązaniem tego problemu może być skonstruowanie dwóch testów mierzących tę samą umiejętność, ale odpowiednio różniących się trudnością oraz zastosowanie odpowiednich procedur umożliwiających przedstawienie wyników obu testów na jednej skali pomiarowej (tzw. skali pionowej, patrz np.: Kolen, Brennan i Kolen, 2004). Dzięki temu możliwe jest wyznaczenie bezwzględnego przyrostu umiejętności dla każdego badanego ucznia. Jest to niewątpliwie wartościowa informacja dla celów diagnozy indywidualnej ucznia, daje bowiem możliwość śledzenia tempa rozwoju poszczególnych umiejętności. Jednak z punktu widzenia modelowania EWD nie jest to informacja niezbędna, a nawet nie jest użyteczna.

Większość modeli EWD (w tym także te stosowane w Polsce) nie wymaga, by wyniki uczniów były wyrażone na tej samej skali, ponieważ wyliczana na ich podstawie wartość dodana mówi o względnym przyroście umiejętności uczniów uczących się w danej szkole w porównaniu do uczniów z innych szkół uwzględnionych w modelu. Dodatkowo, nawet gdybyśmy dysponowali zrównanymi pionowo wynikami testów, nie zmieniłoby to znacząco sposobu wyliczania wskaźników, dla których najważniejsza jest możliwość odniesień i porównań do pozostałych szkół w populacji (czyli właśnie

miara relatywna). Niemniej niektórzy badacze rozważają znaczenie wyrażenia obu pomiarów na jednej skali dla modelowania wartości dodanej, podkreślając zalety takiego rozwiązania (Chudowsky i in., 2010; Linn, 2008; Reckase, 2008), ale także wskazując na trudności modelowania i wrażliwość modeli na przyjętą metodę konstrukcji skal pionowych (Linn, 2008).

Wykorzystanie do modelowania EWD wyników wyrażonych na wspólnej skali ma tę zaletę, że spójność konstruktu jest zapewniona już na etapie tworzenia narzędzi i skal pomiarowych. Testy, które mają być zrównane pionowo, muszą być skonstruowane na podstawie spójnej koncepcji, muszą mierzyć tę samą umiejętność. Dodatkowo analizy statystyczne wykonywane na potrzeby zrównania wyników dają potwierdzenie empiryczne tego, że mamy do czynienia z tą samą cechą ukrytą łączącą u podstaw wyników dwóch pomiarów. Nie dysponując skalą pionową, lub nie planując jej konstruowania, możemy jednak zapewnić spójność konstruktów za pomocą wspólnej koncepcji skal pomiarowych oraz troski na etapie tworzenia narzędzia o dobór takich zadań, które dobrze wskaźnikują badaną umiejętność na danych poziomach nauczania.

W polskim systemie egzaminów zewnętrznych, na etapie planowania i konstruowania testów, nie zakłada się, że wyniki kolejnych egzaminów, do których przystępują ci sami uczniowie, będą wyrażane na wspólnej skali. Egzaminy zewnętrzne nie są także tworzone w celu wyznaczania na ich podstawie wskaźników EWD, więc ich twórcy nie muszą wykazywać się troską o zachowanie spójnej koncepcji testów obejmujących kolejne etapy kształcenia. Z punktu widzenia modelowania EWD niesie to za sobą ryzyko rozbieżności między umiejętnościami mierzonymi na egzaminach podsumowujących kolejne etapy kształcenia. Dlatego konieczna jest analiza tego, czy i które egzaminy mogą być wykorzystane do wyliczenia wskaźników EWD oraz jakie są tego konsekwencje dla interpretacji wyznaczonych wskaźników.

Im konstrukty mierzone za pomocą dwóch testów wykorzystanych do modelowania EWD będą do siebie bardziej zbliżone, tym EWD będzie miało czytelniejszą interpretację. Jeśli bowiem będziemy przewidywać poziom umiejętności czytania po zakończeniu nauki w danej szkole na podstawie uprzedniego poziomu umiejętności czytania, będziemy mogli myśleć o wyznaczonych wskaźnikach EWD jako o mierze względnego przyrostu danej umiejętności. Jeśli natomiast przewidywania tego dokonamy na podstawie na przykład informacji na temat inteligencji i statusu społeczno-ekonomicznego rodziny pochodzenia ucznia, wyliczone na podstawie tego wskaźniki EWD będą interpretowane jako miara wyników kształcenia zrelatywizowana do szkół, które uczą uczniów o takich samych zasobach (inteligencji i statusie społeczno-ekonomicznym). Wskaźniki te są podobne do siebie w takim sensie, że starają się wytrącić z wyniku końcowego to, co jest niezależne od szkoły: uprzedni poziom danej umiejętności lub zasoby ucznia mające duże znaczenie dla osiągnięć

szkolnych, pozostawiając przede wszystkim tę część, którą możemy przypisać pracy szkoły. Nie będą one jednak ani tożsame w sensie wyliczonych wartości ani ich interpretacji.

W przypadku gimnazjalnych modeli EWD do wyliczania wszystkich wskaźników wykorzystuje się tą samą miarę „na wejściu”: wyniki sprawdzianu w klasie szóstej szkoły podstawowej. Jest to test wiedzy ogólnej, mierzący wiadomości i umiejętności z różnych przedmiotów: języka polskiego, matematyki i przyrody. Test ten jest za krótki, by umożliwiał stworzenie wystarczająco rzetelnych podskal odpowiadających poszczególnym dziedzinom obecnym na egzaminie gimnazjalnym. Ponadto często zdarza się tak, że niektóre jego zadania odwołują się równocześnie do wiedzy z różnych przedmiotów i ich przydział do poszczególnych podskal mógłby być co najmniej problematyczny. Dlatego jest on wykorzystywany w całości we wszystkich modelach służących do wyznaczania gimnazjalnych wskaźników EWD. W konsekwencji, wyniki sprawdzianu są traktowane jako miara zasobów czy potencjału ucznia u progu nauki w gimnazjum. Zasoby te są związane z szansą na dobry wynik na egzaminie gimnazjalnym w dwojaki sposób. Po pierwsze, na sprawdzian składają się zadania mierzące umiejętności z przedmiotów obecnych także na egzaminie gimnazjalnym, tak więc jego wynik jest uzależniony częściowo od poziomu umiejętności, będącej przedmiotem zainteresowania z punktu widzenia danego wskaźnika EWD (np. wskaźnika matematyczno-przyrodniczego). Po drugie, wynik sprawdzianu jest nośnikiem informacji o cechach ucznia niezależnych od szkoły, a mających znaczenie dla jego osiągnięć, tak na tym, jak i kolejnych etapach kształcenia (patrz np. rozdziały książki omawiające znaczenie wieku i inteligencji oraz statusu społeczno-ekonomicznego rodziny).

Tak więc dla rozważanego problemu zasadności wykorzystania w wyliczaniu gimnazjalnych wskaźników EWD wyników sprawdzianu jako miary zasobów na progu nauki w gimnazjum kluczowe staje się pytanie o to, jak silnie wyniki egzaminu gimnazjalnego są powiązane z wynikami sprawdzianu. Gdyby związki te były bardzo słabe, oznaczałoby to, że wyniki na sprawdzianie w niewielkim stopniu pozwalają przewidywać wynik na egzaminie gimnazjalnym, a co za tym idzie, nie mogą być wartościową podstawą wyliczania wskaźników EWD. Im zaobserwowane związki będą silniejsze, tym większa pewność co do tego, że u podstawy wyników obu testów leżą zbliżone konstrukty. W tabeli 3. przedstawiono współczynniki korelacji liniowych między oszacowaniami poziomu umiejętności uczniów na sprawdzianie z 2009 roku i egzaminie gimnazjalnym z 2012 roku. Dodatkowo, dla zachowania punktu odniesienia, pokazano korelacje między oszacowaniami poziomu umiejętności dla poszczególnych części egzaminu gimnazjalnego. Korelacje policzone zostały w programie Mplus 7 metodą *pseudo maximum likelihood*, z uwzględnieniem trójstopniowego schematu doboru próby i nierównych prawdopodobieństw doboru.

Tabela 3. Korelacje między egzaminami wykorzystywanymi do wyliczania wskaźników EWD

egzamin	SPR	GMP	GM-M	GM-P	GH	GH-P	GH-H
sprawdzian (SPR)	1						
egzamin gimnazjalny – część matematyczno-przyrodnicza (GMP)	0,782	1					
egzamin gimnazjalny – matematyka (GM-M)	0,755	0,973	1				
egzamin gimnazjalny – przyroda (GM-P)	0,719	0,885	0,760	1			
egzamin gimnazjalny – część humanistyczna (GH)	0,769	0,776	0,731	0,757	1		
egzamin gimnazjalny – język polski (GH-P)	0,720	0,694	0,658	0,671	0,929	1	
egzamin gimnazjalny – historia i WOS (GH-H)	0,707	0,742	0,695	0,727	0,924	0,725	1

Korelacje między wynikami sprawdzianu a wynikami poszczególnych części egzaminu gimnazjalnego są wysokie i wynoszą ok. 0,71–0,78. Są one nieco silniejsze dla przedmiotów matematyczno-przyrodniczych niż humanistycznych, jednak różnice są znikome. Co ciekawe, mimo że sprawdzian jest testem wiedzy ogólnej, korelacje między nim a egzaminem gimnazjalnym z części matematyczno-przyrodniczej lub z matematyki są podobnej wielkości jak korelacje między egzaminem gimnazjalnym w części matematyczno-przyrodniczej a maturalnym wskaźnikiem matematyczno-przyrodniczym lub maturalnym wskaźnikiem z matematyki, a dla przedmiotów humanistycznych są nawet wyższe niż dla analogicznych par egzaminów wykorzystywanych do wyliczenia maturalnych wskaźników EWD (por.: Pokropek, 2013).

Można jednak przypuszczać, że tak wysokie korelacje między egzaminami, które nie mierzą z założenia dokładnie tych samych umiejętności (sprawdzian mierzy szerszy zakres umiejętności) są konsekwencją tego, że wyniki wszystkich tych testów uwarunkowane są przede wszystkim inteligencją i statusem społeczno-ekonomicznym rodziny ucznia. W celu zweryfikowania, czy istotnie mamy tu do czynienia wyłącznie z korelacją pozorną policzono korelacje cząstkowe między wynikami egzaminów po wytrąceniu znaczenia inteligencji i wskaźników statusu. Korelacje wyliczono, porównując wartości współczynników determinacji r^2 modeli, w których wyniki danej części lub testu wchodzącego w skład

egzaminu gimnazjalnego przewidywano za pomocą inteligencji i zmiennych opisujących status rodziny ucznia ze współczynnikami determinacji modeli, w których dodatkowo uwzględniono wyniki sprawdzianu. Inteligencja opisywana była wynikiem testu matrycy Ravena przeprowadzonego na I etapie badania (sposób konstrukcji wskaźnika został opisany w rozdziale poświęconym problemowi wieku i inteligencji dla wyników egzaminacyjnych i EWD, a wskaźniki statusu społeczno-ekonomicznego rodziny ucznia stanowiły wykształcenie rodziców wyrażone w latach nauki (najwyższe z obojga rodziców lub jedynie dostępne), międzynarodowy wskaźnik statusu społeczno-ekonomicznego (HISEI), międzynarodowy wskaźnik prestiżu zawodu (SIOPS) i wskaźnik zasobności domu ucznia. Wskaźniki te zostały dokładnie scharakteryzowane w rozdziale 5. Wszystkie modele estymowane były w programie Mplus 7 metodą *pseudo maximum likelihood*, z uwzględnieniem trójstopniowego schematu doboru próby i nierównych prawdopodobieństw doboru.

Tabela 4. Siła związku poszczególnych części egzaminu gimnazjalnego ze sprawdzianem bez i po wyłączeniu wpływu inteligencji i statusu społeczno-ekonomicznego rodziny ucznia

współczynnik	egzamin gimnazjalny					
	GMP	GM-M	GM-P	GH	GH-P	GH-H
kwadrat korelacji ze sprawdzianem	0,598	0,562	0,497	0,574	0,499	0,490
kwadrat korelacji cząstkowej ze sprawdzianem	0,201	0,184	0,181	0,220	0,194	0,186

Oznaczenia dla egzaminu gimnazjalnego: GMP – część matematyczno-przyrodnicza łącznie, GM-M – matematyka, GM-P – przyroda, GH – część humanistyczna łącznie, GH-P – język polski, GH-H – historia i WOS

Wyniki analiz wskazują, że blisko dwie trzecie siły związku pomiędzy wynikami sprawdzianu a wynikami poszczególnych części egzaminu gimnazjalnego stanowi za pośrednictwem przez sprawdzian wpływ inteligencji i czynników statusowych. Należy jednak zaznaczyć, że wariancja wspólna dla wyników sprawdzianu i egzaminu gimnazjalnego, ale niezwiązana z wpływem inteligencji i czynników statusowych stanowi 18–22% wariancji wyników poszczególnych części egzaminu gimnazjalnego, co należy uznać za znaczny odsetek. Twierdzenie, że związki pomiędzy wynikami sprawdzianu i egzaminu gimnazjalnego są w istocie jedynie korelacją pozorną, wynikającą z wpływu inteligencji i statusu na oba wyniki obu tych egzaminów, jest w świetle tych wyników nieuprawnione.

Zaprezentowane w tej części rozdziału wyniki pokazały, że sprawdzian mierzy na tyle bliski konstrukt, co poszczególne części egzaminu gimnazjalnego, że może być z powodzeniem wykorzystywany w gimnazjalnych modelach EWD jako miara uczniowskich zasobów „na wejściu”. Dopóki pomiar osiągnięć szkolnych na zakończenie szkoły podstawowej nie zmieni swojej formy tak, by umożliwiał wyznaczenie wyników dla odpowiadających egzaminowi gimnazjalnemu umiejętności, dopóty jedyną możliwością wyliczania gimnazjalnych wskaźników EWD będzie wykorzystanie wyniku łącznego ze sprawdzianu jako miary zasobów ucznia rozpoczynającego naukę w danej szkole. Powyższe analizy pokazały jednak, że jest to rozwiązanie uzasadnione.

2.5. Skalowanie wyników

W ogólności pomiar jakiejś cechy sprowadza się do przypisania obiektom liczb określających poziom, wielkość lub natężenie badanej właściwości. W taki sposób mierzymy na przykład obiekty ze świata fizycznego. Chcąc określić masę jakiś produktów, posługując się wagą, przypisujemy im ilości kilogramów. Dodatkowo możemy powiedzieć, że jeden produkt jest na przykład dwa razy cięższy od drugiego. To powszechne doświadczenie sprawia, że wiele osób zwykło sądzić, że aby zmierzyć poziom cechy ukrytej, jaką jest na przykład umiejętność czytania, wystarczy poszczególnym osobom rozwiązującym test czytania przypisać liczbę zdobytych punktów czy odsetek prawidłowo rozwiązanych zadań. W tym przypadku sprawa nie jest jednak taka prosta właśnie z tego powodu, że interesująca nas umiejętność nie jest cechą bezpośrednio obserwowalną, jak większość fizycznych właściwości obiektów. Ponadto tak przypisany wynik będzie uzależniony od rodzaju i liczby zadań w teście, jego trudności itp. Wobec tego, jeśli chcemy mierzyć cechę bezpośrednio nieobserwowalną, musimy wykorzystać jakiś model pomiarowy, który na podstawie obserwowanych danych (rozwiązanych zadań w teście) ujawni lub potwierdzi strukturę badanej cechy oraz zdefiniuje skalę pomiarową. Dlatego do liczenia wskaźników EWD nie wykorzystuje się surowej sumy punktów czy odsetka poprawnie rozwiązanych zadań, ale wyniki przekształca się na skalę o pożądanych właściwościach.

Modele EWD oparte na regresji liniowej (do których możemy zaliczyć także te stosowane w Polsce) wymagają po pierwsze tego, by wyniki pomiarów wykorzystanych do modelowania miały wystarczająco dużo wartości, by mogły być traktowane w analizach jako zmienna o ciągłym charakterze (OECD, 2008). Ponadto powinny one spełniać założenia o interwałowym charakterze skali, na której są wyrażone (Ballou, 2009; Reardon i Raudenbush, 2009). Drugie z tych założeń oznacza, że wzrost o jedną jednostkę na skali oznacza równoważny przyrost wiadomości i umiejętności w każdym zakresie skali. Jest to bardzo silne i trudne do spełnienia założenie, jednak różne sposoby wyznaczania wyników mogą być mniej lub bardziej narażone na odstępstwa od interwałowego charakteru skali, a tym samym mogą powodować mniejsze lub większe ryzyko obarczenia błędem wskaźników EWD.

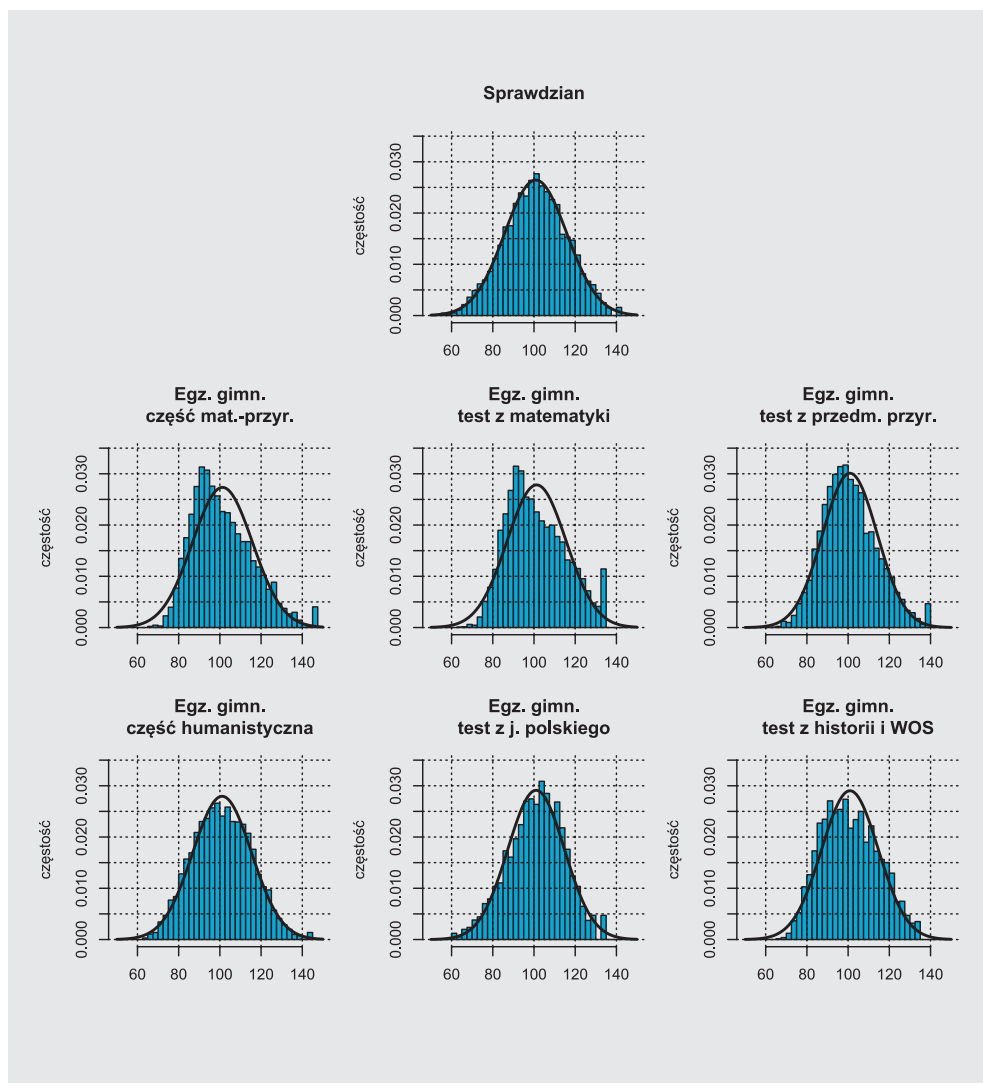
Niektórzy badacze uważają, że wykorzystanie modeli IRT do wyznaczenia wyników zapewnia interwałowy charakter skal, jednak nie jest to założenie niekontrowersyjne (Ballou, 2009; Pokropek, 2013). Podkreśla się, że w szczególności skale powstałe w wyniku zastosowania modelu Rascha i modeli jednoparametrycznych mają teoretyczne podstawy co do tego, by mogły być uznane za interwałowe. Ważne jest przy tym, by model IRT był dobrze dopasowany do danych. Generalnie należy jednak stwierdzić, że założenie o interwałowym charakterze skali jest dużo mniej problematyczne w odniesieniu do wyników wyskalowanych za pomocą modeli IRT niż w odniesieniu do wyników surowych, wyrażonych w centylach czy przeliczonych na odsetek zdobytych punktów.

Wyniki egzaminacyjne wykorzystywane do wyliczania gimnazjalnych wskaźników EWD są skalowane dwuparametrycznym modelem IRT (oraz modelem SGRM dla zadań punktowanych na szerszych skalach). Zadania, które mają bardzo niskie wartości parametru dyskryminacji (poniżej 0,2), są usuwane z testu⁷ na etapie skalowania i wyliczania wyników, jako że wykazują niewielki związek z mierzoną cechą i podważają założenie o jednowymiarowości testu. Te czynności pozwalają poprawić właściwości skal, na których wyrażane są wyniki egzaminacyjne, jednak nie zastąpią przemyślanej procedury tworzenia narzędzi pomiarowych tak, by spełniały one założenia teorii, która powinna być przyjęta z góry. W związku z tym należy poddać weryfikacji jakość powstałych skal pomiarowych. Dokonamy tego, porównując zgodność rozkładów wyskalowanych wyników z rozkładem normalnym, jako że przyjęło się założenie, że właśnie taki rozkład mają umiejętności mierzone testami egzaminacyjnymi (por.: Pokropek, 2013; Reckase, 2008). Znaczące odstępstwa będziemy uznawać za świadczące na niekorzyść skali pomiarowej, zaprzeczające interwałowemu charakterowi skali.

Formalne testy statystyczne badające odstępstwa od rozkładu normalnego są bardzo czułe na nawet niewielkie odchylenia, jeśli analizy są wykonywane na dużych próbach i dlatego ich wyniki mogą być mało informatywne. Przyjrzyjmy się zatem przedstawionym poniżej wykresom pokazującym histogramy wyskalowanych wyników wraz z naniesioną krzywą rozkładu normalnego. Najmniej zastrzeżeń, co do zgodności z rozkładem normalnym, można mieć w przypadku wyników sprawdzianu w 2009 roku. Wyniki egzaminów gimnazjalnych, może za wyjątkiem egzaminu z części humanistycznej i egzaminu z języka polskiego, budzą pewne wątpliwości. W wynikach egzaminów z matematyki i przedmiotów przyrodniczych, a także łącznego wyniku z części matematyczno-przyrodniczej uwidacznia się efekt sufitowy. Dodatkowo rozkłady wyników egzaminów z części matematyczno-przyrodniczej oraz egzaminu z matematyki są delikatnie prawostronnie skośne. Warto jednak podkreślić fakt, że zaobserwowany wzór zgodności lub jej braku z rozkładem normalnym nie jest powtarzalny w kolejnych sesjach egzaminacyjnych (np.: Pokropek, 2013).

⁷ Spośród wszystkich egzaminów wykorzystywanych do wyznaczenia jednorocznych wskaźników EWD za rok 2012 usunięto łącznie trzy zadania. Szczegóły znajdują się w rozdziale 1.

Rysunek 1. Rozkłady wyskalowanych wyników egzaminacyjnych wykorzystywanych do wyliczenia gimnazjalnych wskaźników EWD (dane z próby badawczej)



Wyniki te sugerują konieczność zachowania większej ostrożności w formułowaniu wniosków na podstawie wskaźników dla przedmiotów matematyczno-przyrodniczych z 2012 roku. Nie są one jednak na tyle duże, by dyskredytowały te oszacowania. Natomiast z punktu widzenia modelowania EWD wysoce pożądane byłoby, by egzaminy zewnętrzne miały bardziej stabilne właściwości w kolejnych latach, a rozkłady ich wyników były jak najbliższe rozkładowi normalnemu.

2.6. Podsumowanie

Przedstawione w tym rozdziale analizy wykazały, że egzaminy zewnętrzne wykorzystane do konstrukcji gimnazjalnych wskaźników EWD mają wystarczająco dobre właściwości, by mogły stanowić podstawę wyliczania trafnej miary EWD. Rzetelność testów egzaminacyjnych okazała się wystarczająca dla tworzenia miar zagregowanych. Związki z zewnętrznymi kryteriami przemawiają za tym, że niosą one za sobą informację również o skuteczności realizacji szerzej rozumianych celów nauczania. Ponadto, mimo iż sprawdzian jest testem wiedzy ogólnej, mierzy on na tyle bliski konstrukt co poszczególne części egzaminu gimnazjalnego, że może być z powodzeniem wykorzystywany w modelach EWD jako miara uczniowskich zasobów „na wejściu”. Skale wykorzystane do konstrukcji gimnazjalnych wskaźników EWD mają wystarczająco dobre właściwości, wykazano jedynie drobne odstępstwa od uznanych wzorców.

Jednocześnie warto zaznaczyć, że poprawa pewnych własności polskich egzaminów zewnętrznych mogłaby przyczynić się do zwiększenia jakości wyliczanych wskaźników EWD. Rekomendacje w tym zakresie sprowadzają się do zwiększenia troski o jakość egzaminów na etapie ich planowania i konstruowania; wykorzystania w pełni możliwości badań pilotażowych w celu diagnozy zadań niespełniających założeń przyjętego modelu oraz zadań stronniczych, a także dostosowania trudności egzaminów tak, by ich wyniki miały stabilne rozkłady w kolejnych latach, które zgodnie z założeniem o badanej umiejętności będą zbliżone swoim kształtem do rozkładu normalnego. Ponadto dokumentacja podsumowująca proces tworzenia oraz właściwości każdego testu egzaminacyjnego powinna być publicznie dostępna niedługo po zakończonej sesji. Zwyczaj taki ułatwiłby badaczom korzystanie z wyników egzaminacyjnych w celach naukowych, a dla szerszego grona odbiorców mógłby stać się czynnikiem uwiarygodniającym wykorzystywane narzędzia. Poprawa testów egzaminacyjnych we wspomnianych zakresach podniosłaby nie tylko jakość wskaźników EWD, ale każdej miary opartej na tych testach: średniego wyniku szkoły, a w szczególności wyników indywidualnych uczniów.

Literatura

Anastasi, A. i Urbina, A. (1999). *Testy psychologiczne*. Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.

Ballou, D. (2009). Test Scaling and Value-Added Measurement. *Education Finance and Policy*, 4(4), 351–383.

Centralna Komisja Egzaminacyjna (2010). *Informator o egzaminie gimnazjalnym od roku szkolnego 2011/2012*. Warszawa: Centralna Komisja Egzaminacyjna.

Centralna Komisja Egzaminacyjna (2012). Wstępne informacje o wynikach egzaminu gimnazjalnego w 2012 r. Retrieved August 7, 2013, from http://www.cke.edu.pl/files/file/kom_na_str_cke.pdf.

Chudowsky, N., Koenig, J. A. i Braun, H. I., (red.) (2010). *Getting value out of value-added report of a workshop*. Washington: National Academies Press. Pobrano z: <http://site.ebrary.com/id/10370371>

De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford Press.

de la Torre, J. i Patz, R. J. (2005). Making the most of what we have: A practical application of multidimensional item theory in test scoring. *Journal of educational and Behavioral Statistics*, 30, 295–311.

Dolata, R., Jasińska, A. i 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.

Grudniewska, M. i Kondrątek, B. (2012). Zróżnicowane funkcjonowanie zadań w egzaminach zewnętrznych w zależności od płci na przykładzie części matematyczno-przyrodniczej egzaminu gimnazjalnego. [w:] Niemierko, B. i Szmigel, M. K. (red.), *Regionalne i lokalne diagnozy edukacyjne*. Grupa Tomami.

Haeblerlin, U., Moser, U., Bless, G. (1989). Fragebogen zur Erfassung von Dimensionen der Integration von Schülern (FDI 4–6) – Questionnaire for recording dimensions of integration of pupils. *Psychologie in Erziehung und Unterricht: Zeitschrift für Forschung und Praxis*, 36(H. 1), 17–26.

Informator z aneksem dla uczniów ze specyficznymi trudnościami w uczeniu się. Sprawdzian w klasie szóstej szkoły podstawowej przeprowadzony od roku szkolnego 2009/2010. (2009). Warszawa: Centralna Komisja Egzaminacyjna.

Jakubowski, M. i Pokropek, A. (2009). *Badając egzaminy: podejście ilościowe w badaniach edukacyjnych*. Warszawa: Centralna Komisja Egzaminacyjna.

Jasińska, A. (2010). Pomiar gotowości szkolnej za pomocą skali quasi-obszewacyjnej. [w:] Niemierko, B., Szmigel, M. K. (red.), *Teraźniejszość i przyszłość oceniania szkolnego* (s. 415–424). Kraków: Grupa Tomami.

Kolen, M. J., Brennan, R. L. i Kolen, M. J. (2004). *Test equating, scaling, and linking: methods and practices*. New York: Springer.

- Linn, R. (2008). Measurement issues associated with value-added methods. [w:] *Paper prepared for a Workshop Held by the Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation and Educational Accountability*. Washington: DC.
- Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V.-N. i Martinez, J. F. (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. *Journal of Educational Measurement*, 44(1), 47–67.
- Novick, M. R. i Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32(1), 1–13.
- OECD (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paris: OECD.
- Pokropek, A. (2011). Matura z języka polskiego. Wybrane problemy psychometryczne. [w:] Niemierko, B., Szmigiel, M. K. (red.), *Ewaluacja w edukacji: koncepcje, metody, perspektywy*. Kraków: Grupa Tomami.
- Pokropek, A. (2013). Trafność testów egzaminacyjnych. [w:] M. Karwowski (Red.), *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne*. Warszawa: Wydawnictwo Instytutu Filozofii i Socjologii PAN.
- Pokropek, A. i Żółtak, T. (2012). Nowe modele jednorocznej EWD. [w:] Niemierko, B. i Szmigiel, M. K. (red.), *Regionalne i lokalne diagnozy edukacyjne*. Grupa Tomami.
- Reardon, S. F. i Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4), 492–519.
- Reckase, M. (2008). Measurement Issues Associated with Value-Added Methods. [w:] *Paper prepared for a Workshop Held by the Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation and Educational Accountability*. Washington: DC.
- Ryan, R. M. i Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67. doi:10.1006/ceps.1999.1020
- Sędek, G. (1995). *Bezradność intelektualna w szkole*. Warszawa: Wydawnictwo Instytutu Psychologii.
- Skorupiński, P. M. (2013). Modele trafności pomiaru. [w:] Karwowski, M. (red.), *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne*. Warszawa: Wydawnictwo Instytutu Filozofii i Socjologii PAN.
- Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K. i Lens, W. (2009). Motivational Profiles from a Self-Determination Perspective: The Quality of Motivation Matters. *Journal of Educational Psychology*, 101(3), 671–688.

Rozdział 3

Gimnazjalne modele edukacyjnej wartości dodanej

W niniejszym rozdziale przedstawione zostaną podstawowe problemy związane z modelowaniem statystycznym EWD. Na wstępie omówiona zostanie kwestia uwzględniania w modelach służących do wyliczania wskaźników EWD zmiennych kontrolnych. Rozwiązania przyjęte przy konstrukcji polskich wskaźników dla gimnazjów przedstawione zostaną w odniesieniu do rozwiązań wykorzystywanych w Anglii i USA. W dalszej części rozdziału przedstawiony zostanie dokładniej sposób wyliczania wskaźników EWD gimnazjów, zarówno jednorocznych jak i trzyletnich. Omówione zostaną też zmiany, jakie zaszły w tym zakresie od początku projektu.

3.1. Zmienne kontrolne w modelach EWD

Na osiągnięcia uzyskiwane przez uczniów na egzaminach wpływają różne czynniki. Zaletą wskaźników EWD jest to, że pozwalają one kontrolować te z nich, które w ten sam sposób wpływają na wyniki egzaminu na wejściu i na wyjściu. Jednocześnie znaczna liczba czynników wciąż może mieć wpływ na postępy uzyskiwane przez uczniów w toku dalszej nauki, niezależnie od ich wpływu na uprzednie osiągnięcia.

Jaskrawym przykładem mogą tu być uczniowie ze środowisk imigranckich, którzy nie opanowali biegle języka wykładowego. Ta cecha będzie bardzo utrudniać im zdobywanie dalszej wiedzy i mamy mocne podstawy by sądzić, że pod koniec danego etapu kształcenia ich poziom umiejętności i wyniki egzaminów, będą w związku z tym niższe niż innych uczniów, którzy mieli te same wyniki egzaminów na wejściu. Z polskiej perspektywy przykład ten może się oczywiście wydawać nieco egzotyczny, jednak dobrze pokazuje, że mogą istnieć czynniki, które mają wyraźny wpływ na postępy uczniów w nauce, a nie są bezpośrednio zależne od działań podejmowanych przez szkołę. Jeśli chcemy, aby wskaźniki EWD pokazywały efektywność pracy szkół, takie czynniki powinny być dodatkowo kontrolowane tak, aby różnice pomiędzy uczniami uczącymi się w różnych szkołach nie miały wpływu na wartości wyliczanych dla tych szkół wskaźników EWD. Jeśli wiemy, że pewni uczniowie będą osiągać większe (lub mniejsze) postępy w nauce, to powinniśmy odpowiednio modyfikować przewidywane dla nich wyniki na wyjściu – punkt odniesienia, do którego będą porównywani. Osiągnięciu tego celu służy dołączenie do modelu regresji dodatkowych zmiennych opisujących charakterystyki uczniów oraz środowiska, w jakim działa szkoła. Poniżej przedstawione zostaną rozwiązania, jakie przyjęto w systemie angielskim oraz w USA, na terenie *Dallas Independent School District*.

W rozwiązaniu angielskim zakres dodatkowych czynników uwzględnionych w modelu obejmował: płeć, wiek, posiadanie prawa do bezpłatnych posiłków, przynależność do grup

etnicznych (wyróżniono 19) oraz jej interakcję z prawem do bezpłatnych posiłków, informacje o specjalnych potrzebach edukacyjnych oraz o tym, czy język angielski jest pierwszym językiem ucznia, informacje o zmianie szkoły w trakcie toku kształcenia, wskaźnik ubóstwa dla miejsca zamieszkania ucznia (IDACI) oraz średnie wyniki uczniów szkoły na wejściu (Evans, 2008).

Warto przy tym zauważyć, że w początkowym okresie rozwoju angielskiego systemu EWD (w latach 2002–2005), wskaźniki wyliczane były bez uwzględnienia wymienionych powyżej zmiennych kontrolnych. Wprowadzenie ich do modeli wymagało rozwinięcia systemu informacji oświatowej i nastąpiło w 2006 r., a jednym z powodów tej zmiany była krytyka własności wskaźników wyliczanych bez kontroli dodatkowych czynników (Evans, 2008). Mimo to w 2012 r. postanowiono w Anglii odstąpić od wyliczania wskaźników EWD z uwzględnieniem zmiennych kontekstowych (RAISEonline, 2011). Pewien wpływ na tę decyzję ma zapewne przykładanie w ramach systemu angielskiego dużej wagi dla użyteczności wskaźników efektywności pracy szkół dla rodziców (wspomaganie wyboru szkoły). Model uwzględniający zmienne kontekstowe, jako bardziej skomplikowany, czynił mało przejrzystą procedurę wyliczania wskaźników. Dodatkowo podnoszono też wątpliwości odnośnie do zbyt szerokiego zakresu kontrolowanych w modelu czynników (Goldstein i Leckie, 2009).

Nieco bardziej skomplikowane rozwiązanie zastosowano w Dallas, gdzie wskaźniki EWD wyliczane są w ramach dwustopniowej procedury, w której najpierw przeprowadza się oddzielnie regresje wyników na wejściu, wyników na wyjściu oraz odsetka opuszczonych w ostatnim roku godzin lekcyjnych ze względu na kontrolowane czynniki z poziomu indywidualnego (określane jako *fairness variables*). Obejmują one: ograniczoną zdolność posługiwania się językiem angielskim, rasę/bycie Latynosem, płeć oraz posiadanie uprawnień do tańszych lub darmowych posiłków (oraz interakcje tych zmiennych). Dodatkowo kontrolowane są też – na podstawie danych spisowych¹ – wskaźniki zamożności dla regionu zamieszkania ucznia. W drugim kroku reszty z regresji opisującej wynik na wyjściu przewidywane są za pomocą reszt z regresji opisującej wynik na wejściu, reszt z regresji opisującej odsetek opuszczonych godzin lekcyjnych oraz czynników kontekstowych mierzonych na poziomie szkoły, które obejmują mobilność uczniów (częstość przychodzenia do szkoły nowych uczniów i/lub odchodzenia uczniów ze szkoły), zatłoczenie, oraz średnie zmiennych kontrolowanych na poziomie indywidualnym (McCaffrey, Lockwood, Koretz i Hamilton, 2003; Webster, Mendero, Orsak i Weerasinghe, 1998).

¹ Warto pamiętać, że w USA badania spisowe prowadzone są częściej niż w Polsce.

Generalnie przyjmuje się, że uwzględnienie w modelach zmiennych kontrolnych pozytywnie wpływa na własności wyliczanych wskaźników EWD. Można tu jednak sformułować dwa zastrzeżenia. Po pierwsze, zakres kontrolowanych zmiennych powinien być dostosowany do podstawowego celu, jakiemu mają służyć wskaźniki. Po drugie, w sytuacji nierosowego przydziału uczniów do szkół – co ma miejsce właściwie w każdym systemie szkolnym, choć z różnym nasileniem – część wpływu wynikającego z działań szkół może być w procesie estymacji modelu przypisywana oddziaływaniu zmiennych kontrolnych. Poszerzanie zakresu czynników kontrolowanych w modelu z jednej strony chroni więc przed „niesprawiedliwym” ocenianiem szkół, jednak z drugiej może utrudniać rozróżnianie od siebie szkół pracujących efektywnie i nieefektywnie (Ballou, Sanders i Wright, 2004; McCaffrey i in., 2003; McCaffrey i in., 2004; OECD, 2008; Raudenbush i Willms, 1995).

W przypadku pierwszego z wymienionych problemów wskazuje się na różnice w ocenie szkoły z perspektywy rodzica wybierającego szkołę dla swojego dziecka i z perspektywy ewaluatora, chcącego ocenić efektywność pracy szkoły. Różnice te przyjęło się w literaturze opisywać za pomocą wprowadzonego przez Stephena Raudenbusha i J. Douglasa Willmsa (1995) rozróżnienia na dwa rodzaje efektów (wskaźników), które można wyliczać w odniesieniu do szkół. Tak zwane efekty typu A uwzględniają jako zmienne kontrolne tylko charakterystyki uczniów, ale pomijają charakterystyki środowiska, w jakim działa szkoła (np. średni poziom wyników uczniów przychodzących do szkoły albo jej lokalizację), podczas gdy efekty typu B uwzględniają również te drugie. Z perspektywy rodzica interesujące są efekty typu A, gdyż chce on wybrać szkołę, w której może spodziewać się najwyższych wyników na wyjściu, bez względu na to, czy miałyby być one wynikiem działań samej szkoły, czy też faktu, że po prostu pracuje ona w bardzo korzystnym środowisku. Jeśli jednak chcemy na podstawie EWD oceniać jakość pracy szkoły, powinniśmy oczywiście wytrącać ze wskaźników również wpływ czynników kontekstowych, a więc wyliczać efekty typu B. Należy jednak zaznaczyć, że uzyskanie wiarygodnych wskaźników typu B jest trudne. W literaturze poświęconej EWD podkreśla się, że nawet gdy w modelach uwzględniany jest szeroki zakres zmiennych kontrolnych, nie należy na podstawie wskaźników EWD formułować mocnych interpretacji przyczynowych (Raudenbush, 2004; Reardon i Raudenbush, 2008; Rubin, Stuart i Zanutto, 2004).

Na koniec należy wspomnieć, że istnieje również inny sposób wytrącania ze wskaźników EWD wpływu charakterystyk uczniów, jednak jest on możliwy do zastosowania tylko, gdy jednocześnie spełnione są trzy warunki: 1) wskaźniki wyliczane są w odniesieniu do nauczycieli (a nie szkół); 2) testowanie poziomu umiejętności uczniów odbywa się co roku; 3) przydział nauczycieli do uczniów zmienia się co roku. W takiej sytuacji możliwe jest wykorzystanie do wyliczania wskaźników EWD modeli charakterystycznych dla amerykańskiego systemu TVAAS (*Tennessee Value-Added Assessment System*) i jego „skomercjalizowanej” wersji EVAAS (*Education*

Value-Added Assessment System). W tym przypadku wytrącenie wpływu czynników niezależnych od działań nauczycieli staje się możliwe, dzięki uwzględnieniu w jednym modelu wyników egzaminów z wielu lat oraz corocznym zmianom zespołów nauczycielskich uczących poszczególnych uczniów, pomimo nieuwzględniania w modelu żadnych zmiennych kontrolnych (McCaffrey i in., 2003, Wright, Horn i Sanders, 1997). Rozwiązanie to nie znajduje jednak zastosowania w przypadku wskaźników wyliczanych na poziomie szkół, gdyż zdecydowana większość uczniów uczęszcza na danym poziomie kształcenia cały czas do tej samej szkoły. Modele wykorzystywane do wyliczania EWD na poziomie szkół w ramach metodologii TVAAS/EVAAS są więc de facto równoważne modelom regresji uwzględniającym tylko pomiar na wejściu i na wyjściu.

3.2. Specyfika polskich modeli gimnazjalnych EWD

Gimnazjalne wskaźniki EWD wyliczane są w dwóch podstawowych wariantach: jako wskaźniki jednoroczne i trzyletnie. Warianty te różnią się nieco metodami wyliczania (zostaną one omówione w dalszej części rozdziału), ale za ważniejszą różnicę należy uznać ich przeznaczenie i sposób udostępniania odbiorcom.

Wskaźniki trzyletnie publikowane są w ogólnodostępnym serwisie internetowym², z myślą o bardzo szerokiej grupie odbiorców, od nauczycieli i dyrektorów, poprzez organy zarządzające i nadzór pedagogiczny, po rodziców uczniów. W związku z tym zdecydowano się na wyliczanie wskaźników w oparciu o wyniki trzech kolejnych sesji egzaminu gimnazjalnego tak, by były one bardziej stabilne, zdawały sprawę z przeciętnej efektywności pracy szkoły w dłuższym okresie i nie prowokowały do wyciągania zbyt pochopnych wniosków z różnic, jakie mogą występować pomiędzy pojedynczymi latami. Do prezentacji wskaźników wykorzystuje się przy tym nowatorską (w kontekście EWD) metodę, w której za pomocą elipsy jednocześnie pokazywana jest informacja o EWD i średniej wyników egzaminu na wyjściu w ramach szkoły, z uwzględnieniem niepewności oszacowania tych parametrów (Pokropek i Żółtak, 2012a).

Wskaźniki jednoroczne, historycznie powstałe jako pierwsze, przeznaczone są z kolei przede wszystkim na potrzeby ewaluacji wewnętrznej. Odbiorcom nie są przy tym udostępniane gotowe wskaźniki, wyliczone dla poszczególnych szkół (jak ma to miejsce w przypadku wskaźników trzyletnich), lecz aplikacja pozwalająca wyliczyć wartość EWD dla dowolnej grupy, określonej przez użytkownika na podstawie danych, wczytanych przez niego do tej aplikacji. EWD jest oczywiście wyliczana na podstawie modelu wyestymowanego na danych ogólnokrajowych, którego parametry, pozwalające ustalić przewidywany wynik egzaminu

² <http://gimnazjum.ewd.edu.pl>

na wyjściu, są zapisywane w przygotowywanej aplikacji. Pozwala to na prowadzenie bardziej dociekliwych analiz nad zróżnicowaniem efektywności kształcenia, jednak stawia użytkownikom dużo wyższe wymagania zarówno w zakresie technicznych umiejętności posługiwania się dostarczanym im narzędziem, jak i w zakresie interpretacji uzyskiwanych wyników. Grupa osób mogących korzystać ze wskaźników jednorocznych jest przy tym zawężona do tych, którzy mogą otrzymać z OKE dane z wynikami egzaminacyjnymi uczniów – w praktyce głównie nauczycieli i dyrektorów.

Nieco paradoksalnie, choć podstawowym zadaniem polskich wskaźników EWD jest ocena efektywności kształcenia szkół, uwzględnia się w nich jedynie bardzo ubogi zestaw zmiennych kontrolnych (taki sam zarówno w modelach jednorocznych, jak i w trzyletnich). Powoduje to zresztą konieczność przeprowadzenia badań stanowiących główną część tego raportu. Podstawowym problemem związanym z uwzględnianiem w modelu zmiennych kontrolnych jest słabość polskiego systemu informacji oświatowej. W praktyce niemożliwe jest dołączenie do wyników egzaminacyjnych żadnych dodatkowych informacji o uczniach ponad te, które przechowywane są w bazach danych okręgowych komisji egzaminacyjnych, które obejmują płeć, wiek (datę urodzenia) oraz informacje o tym czy uczeń na egzaminie posiadał zaświadczenie o dysleksji oraz czy był laureatem konkursu przedmiotowego. Zakres zmiennych uwzględnianych w modelach jest przy tym jeszcze nieco węższy, gdyż obejmuje tylko płeć i informacje o dysleksji (podczas egzaminu na wejściu, podczas egzaminu na wyjściu oraz interakcję tych dwóch zmiennych).

Włączenie do modeli płci oznacza, że jeśli w skali ogólnokrajowej występują różnice w średnich postępach dziewcząt i chłopców, to nie będą one mieć wpływu na wskaźniki EWD szkół – punktem odniesienia dla dziewcząt będą inne dziewczęta, a dla chłopców inni chłopcy. Analogiczna sytuacja występuje w przypadku informacji o posiadaniu zaświadczenia o dysleksji, z tym, że inny punkt odniesienia jest tu ustalany w ramach każdej z czterech grup: uczniów, którzy nie posiadali takiego zaświadczenia na żadnym z egzaminów, uczniów, którzy mieli je tylko na egzaminie na wejściu, uczniów, którzy posiadali je tylko na egzaminie na wyjściu i wreszcie uczniów, którzy posiadali je na obu egzaminach. Istnieją powody by przypuszczać, że posiadanie zaświadczenia o dysleksji nie jest zbyt dobrym wskaźnikiem występowania specyficznych trudności w nauce czytania i pisania (Dolata, 2007). Jest za to jasne, że wyniki egzaminów takich osób nie powinny być bezpośrednio porównywane z wynikami innych zdających, gdyż uczniowie posiadający zaświadczenie o dysleksji mają wydłużony czas pisania egzaminu i stosuje się wobec nich łagodniejsze kryteria oceny odnośnie do błędów ortograficznych.

Laureaci konkursów przedmiotowych również stanowią specyficzną grupę uczniów. Nie podchodzą oni do egzaminu (w przypadku egzaminu gimnazjalnego – do jednej z jego części, odpowiadającej tematyce konkursu, którego są laureatami), lecz mają przypisywane

maksymalne możliwe do uzyskania wyniki. Jednocześnie laureaci stanowią grupę niezwykle zróżnicowaną wewnątrz. Liczba konkursów dających możliwość zwolnienia z egzaminu jest znaczna, a do tego różna na terenie różnych kuratoriów. Oczywiście brak też standaryzacji co do trudności różnych konkursów. Sprawia to, że sensowność porównywania laureatów do innych laureatów jest mocno problematyczna. W związku z tym zdecydowano się nie uwzględniać w modelach informacji o byciu laureatem, uczniowie tacy traktowani są tak samo jak inni, którzy otrzymali z egzaminu maksymalną liczbę punktów.

Z punktu widzenia EWD najlepszym rozwiązaniem problemu związanego z laureatami byłoby, gdyby pisali oni egzaminy tak jak wszyscy uczniowie, a wyniki konkursów ujawniane były dopiero po napisaniu egzaminów. W ten sposób wyniki konkursów mogłyby być uwzględniane w rekrutacji, jednak w danych egzaminacyjnych dysponowalibyśmy również dobrym oszacowaniem bardziej ogólnych umiejętności laureatów, porównywalnym z wynikami innych uczniów. Niestety, nic nie wskazuje, aby w dającej się przewidzieć przyszłości taka zmiana miała nastąpić.

Wiek uczniów nie jest uwzględniany w modelach EWD z powodów, które można by określić jako historyczne. W pierwszym okresie prac nad rozwojem wskaźników informacje o dacie urodzenia uczniów (na podstawie której można wyliczyć ich wiek) nie była zbierana wraz z wynikami egzaminów, co oczywiście uniemożliwiało uwzględnienie tego czynnika. W późniejszym okresie modyfikacje metodologii skupiły się z kolei na skalowaniu wyników egzaminacyjnych i uwzględnieniu uczniów o wydłużonym toku kształcenia. W świetle ostatnich wyników badań (Dolata i Pokropek, 2012) kwestia uwzględnienia wieku w gimnazjalnych modelach EWD wydaje się ważna i z pewnością będzie rozważana.

3.3. Jednoroczne wskaźniki EWD

Jednoroczne gimnazjalne wskaźniki EWD, opublikowane zostały po raz pierwszy w 2007 roku. Początkowo ich wyliczanie umożliwiały odpowiednio przygotowane arkusze kalkulacyjne, udostępniane odbiorcom pod nazwą Kalkulator EWD. Od 2009 roku wydawana jest samodzielna aplikacja służąca do wyliczania wskaźników – Kalkulator EWD Plus, zastąpiona w 2012 roku przez Kalkulator EWD 100. Do 2011 roku dla każdego roku udostępniane były dwa wskaźniki: humanistyczny i matematyczno-przyrodniczy. W każdym z nich jako miara umiejętności uczniów na wejściu wykorzystywany był wynik sprawdzianu, a jako miara umiejętności uczniów na wyjściu wynik odpowiedniej części egzaminu gimnazjalnego. W 2012 roku, w związku ze zmianą formuły egzaminu gimnazjalnego, zakres wskaźników rozszerzono o cztery dodatkowe, które jako miary umiejętności na wyjściu wykorzystują wyniki poszczególnych testów, z których składa się nowy egzamin gimnazjalny. Są to wskaźniki dla: języka polskiego, historii

i WOS-u, matematyki oraz przedmiotów przyrodniczych. Wszystkie jako miarę umiejętności na wejściu wykorzystują wyniki sprawdzianu (Pokropek i Żółtak, 2012b).

Modele służące do wyliczania wskaźników jednorocznych są stosunkowo proste. W formie wykorzystywanej do 2011 roku pozwalały one uwzględnić w analizie jedynie uczniów, którzy uczyli się w gimnazjum przez 3 lata, a w modelu regresji wykorzystywane były nieprzekształcone (surowe) wyniki egzaminów (sumy punktów). Stosowane modele regresji miały postać:

$$Y_i = w(X_i) + b_p \text{płeć}_i + b_{dg} \text{dysl_g}_i + b_{ds} \text{dysl_s}_i + b_{dgs} (\text{dysl_g}_i \cdot \text{dysl_s}_i) + r_i \quad (1)$$

gdzie:

Y_i – wynik i -tego ucznia z danej części egzaminu gimnazjalnego,

X_i – wynik i -tego ucznia ze sprawdzianu,

$w(X)$ – funkcja opisująca zależność pomiędzy wynikami sprawdzianu a wynikami danej części egzaminu gimnazjalnego,

płeć_i – płeć i -tego ucznia (0 – mężczyzna, 1 – kobieta),

dysl_g_i – posiadanie przez i -tego ucznia zaświadczenia o dysleksji na egzaminie gimnazjalnym (0 – nie, 1 – tak),

dysl_s_i – posiadanie przez i -tego ucznia zaświadczenia o dysleksji na sprawdzianie (0 – nie, 1 – tak),

r_i – reszta regresji dla i -tego ucznia.

Funkcja $w(X)$ mogła przyjmować dwie formy – regresji dwoma kawałkami liniowej (rozwiązanie stosowane dla wskaźników z lat 2006–2008) lub wielomianu (od 2009 roku). Metoda wyznaczania stopnia takiego wielomianu opisana jest w publikacji poświęconej modyfikacjom, jakie wprowadzono w jednorocznych wskaźnikach gimnazjalnych w 2012 r. (Pokropek i Żółtak, 2012b).

Wskaźniki EWD są na podstawie takiego modelu wyliczane jako średnia z reszt regresji w ramach grupy uczniów, np. wszystkich uczniów danej szkoły. Analizy przeprowadzone w początkowym okresie prac nad rozwojem metodologii EWD w Polsce wykazały przy tym, że tak wyliczane wskaźniki nie różnią się wyraźnie od uzyskiwanych bardziej skomplikowanymi metodami, na podstawie wyników regresji mieszanych efektów (Jakubowski, 2007). Jednocześnie metoda oparta na prostym wyliczaniu średniej z reszt jest prostsza do zastosowania w Kalkulatorze EWD, bardziej zrozumiała dla odbiorców i, co najważniejsze, pozwala wyliczać EWD dla dowolnie zdefiniowanych grup uczniów, umożliwiając prowadzenie analiz wewnątrzszkolnego zróżnicowania efektywności kształcenia. Wskaźniki wyliczane na podstawie regresji mieszanych efektów (wielopoziomowej) nie mają tej ostatniej własności – wyniki z modelu, w którym założono określone pogrupowanie (z którym związany jest efekt losowy) mogą być w uprawniony sposób użyte tylko do wyliczania wskaźników EWD dla przyjętego rodzaju grup. Na przykład wyniki z modelu, który uwzględniał pogrupowanie uczniów na poziomie szkół, nie powinny być potem wykorzystywane do wyliczania wskaźników EWD na poziomie oddziałów czy w podziale na płeć w ramach jednej szkoły.

Zmiany wprowadzone w 2012 roku, oprócz rozszerzenia liczby wyliczanych wskaźników objęły zastąpienie w regresji surowych wyników egzaminacyjnych wynikami zrównanymi ekwikwantyloво oraz uwzględnienie uczniów, których tok kształcenia w gimnazjum uległ wydłużeniu o jeden rok (Pokropek i Żółtak, 2012b). Wykorzystanie wyników znormalizowanych ekwikwantyloво, wyrażonych na skali 100;15 miało na celu zarówno polepszenie własności tak przekształconych zmiennych w modelach regresji, jak i ułatwienie interpretacji wyników. Wadą wyników surowych polskich egzaminów jest bowiem zróżnicowany pomiędzy latami poziom trudności i rozkład wyników – normalizacja pozwala te problemy w znacznej mierze usunąć, co sprawia, że wyniki z różnych lat mogą być interpretowane z użyciem tych samych kryteriów. Dodatkowo skala 100;15 wykorzystywana była już od 2009 roku do prezentacji wyników egzaminów w ramach trzyletnich wskaźników gimnazjalnych.

Uwzględnienie uczniów o toku kształcenia wydłużonym o jeden rok, pozwala znacznie lepiej opisać efektywność pracy szkół, w których występuje znaczny odsetek uczniów nieotrzymujących promocji do następnej klasy. Odcina to też szkoły od możliwości niepromowania uczniów, z którymi pracowano wyjątkowo nieefektywnie, w celu wykluczenia ich z grupy, na podstawie której wyliczane będą wskaźniki EWD. Przyjęto rozwiązanie, w którym dla uczniów o wydłużonym o rok toku kształcenia wyznaczana jest oddzielna linia przewidywanego wyniku. Ma to uzasadnienie zarówno w specyfice tej grupy uczniów – rozsądne wydaje się stawianie im nieco innych oczekiwań niż uczniom o standardowym, trzyletnim toku kształcenia – jak i w fakcie, że pisali oni inną edycję sprawdzianu. Jednocześnie współczynniki związane z płcią i dysleksją wyliczane są w modelu łącznie dla wszystkich uczniów, bez podziału na tych o standardowym i wydłużonym toku kształcenia. Zmodyfikowane w ten sposób modele regresji wyraża się wzorem:

$$Y_{Ni} = w(X_{Ni}, tok_i) + b_p \text{płeć}_i + b_{dg} \text{dysl_g}_i + b_{ds} \text{dysl_s}_i + b_{dgs} (\text{dysl_g}_i \cdot \text{dysl_s}_i) + r_i \quad (2)$$

gdzie:

Y_{Ni} – znormalizowany ekwikwantyloво wynik i -tego ucznia z danego testu lub części egzaminu gimnazjalnego,

X_{Ni} – znormalizowany ekwikwantyloво wynik i -tego ucznia ze sprawdzianu,

tok_i – tok kształcenia i -tego ucznia (0-normalny, 1-wydłużony),

$w(X_i, tok_i)$ – wielomian opisujący zależność pomiędzy wynikami sprawdzianu a wynikami danego testu lub części egzaminu gimnazjalnego,

płeć_i – płeć i -tego ucznia (0 – mężczyzna, 1 – kobieta),

dysl_g_i – posiadanie przez i -tego ucznia zaświadczenia o dysleksji na egzaminie gimnazjalnym (0 – nie, 1 – tak),

dysl_s_i – posiadanie przez i -tego ucznia zaświadczenia o dysleksji na sprawdzianie (0 – nie, 1 – tak),

r_i – reszta regresji dla i -tego ucznia.

Funkcja $w(X_i, tok_i)$ jest z kolei postaci:

$$w(X_i, tok_i) = b_0 + b_1 X_i + \dots + b_k X_i^k + b_{w0} tok + b_{w1} X_i tok + \dots + b_{wk} X_i^k tok \quad (3)$$

a więc linia regresji dla uczniów o standardowym toku kształcenia³ opisywana jest wzorem:

$$w(X_i, tok_i = 0) = b_0 + b_1 X_i + \dots + b_k X_i^k \quad (4)$$

a dla uczniów o toku kształcenia wydłużonym o rok wzorem:

$$w(X_i, tok_i = 1) = (b_0 + b_{w0}) + (b_1 + b_{w1}) X_i + \dots + (b_k + b_{wk}) X_i^k \quad (5)$$

Więcej informacji na temat jednorocznych wskaźników gimnazjalnych można znaleźć w przytaczanych już publikacjach Macieja Jakubowskiego (2007) oraz Artura Pokropka i Tomasza Żółtaka (2012b).

3.4. Trzyletnie wskaźniki EWD

Wskaźniki trzyletnie wyliczane są wyłącznie na poziomie szkół i opublikowane zostały po raz pierwszy w 2009 roku. Choć nie pozwalają one prowadzić analiz wewnętrzzszkolnych, w serwisie internetowym, w którym są one prezentowane, udostępniane jest narzędzie pozwalające w łatwy sposób dokonywać porównań pomiędzy szkołami. Analogicznie jak w przypadku wskaźników jednorocznych obowiązujących do 2011 roku, przygotowywane zostały wskaźniki humanistyczne i matematyczno-przyrodnicze. Rozszerzenie zakresu wskaźników o cztery kolejne, odnoszące się do nowej struktury egzaminu gimnazjalnego, będzie możliwe dopiero w 2014 roku, gdy wszystkie trzy sesje egzaminacyjne objęte wskaźnikiem trzyletnim (2014–2012) będą przeprowadzone według nowej formuły.

Do 2011 roku w modelach wykorzystywane były znormalizowane ewkwikwantyloво wyniki sprawdzianu i egzaminu gimnazjalnego, uwzględniano przy tym wyłącznie wyniki uczniów, którzy uczyli się w gimnazjum przez trzy lata (Pokropek i Żółtak 2012a). Począwszy od wskaźników dla okresu 2012–2010 normalizację ekwkwantylową wyników egzaminów zastąpiono

³ Dokładnie dla chłopców, którzy nie posiadali zaświadczenia o dysleksji ani na egzaminie gimnazjalnym, ani na sprawdzianie.

skalowaniem IRT za pomocą modelu 2PL/GPCM, a w modelach regresji uwzględniono też uczniów, których tok kształcenia uległ wydłużeniu o jeden rok.

Trzyletnie gimnazjalne wskaźniki EWD wyliczane są jako Bayesowskie predykcje a posteriori (określane też mianem *best linear unbiased predictors*) na podstawie wyników modeli regresji mieszaných efektów (wielopoziomowej), w której efekt losowy dla stałej regresji związany jest z podziałem na szkoły. Są to takie same metody statystyczne jak wykorzystywane w Anglii i w Dallas (Evans, 2008; Webster i in., 1998). Krzywa przewidywanego wyniku danej części egzaminu gimnazjalnego na podstawie sprawdzianu wyznaczana jest oddzielnie dla każdej kombinacji roku zdawania sprawdzianu i roku zdawania egzaminu gimnazjalnego. Oznacza to, że w modelach, które nie uwzględniały uczniów o wydłużonym toku kształcenia wyznaczane były trzy różne krzywe przewidywania (po jednej dla każdego rocznika zdających), a w modelach uwzględniających uczniów o toku kształcenia wydłużonym o rok wyznaczane jest po sześć różnych krzywych przewidywania (dodatkowe trzy – po jednej dla każdego rocznika uczniów o wydłużonym toku kształcenia). Wartości współczynników związanych z płcią i dysleksją wyliczane są w modelu łącznie dla wszystkich uczniów. Wykorzystywane modele regresji opisuje wzór:

$$Y_{ji} = u_{0j} + w(X_{ji}, rok_{ji}, tok_{ji}) + b_p \text{płeć}_{ji} + b_{dg} \text{dysl_g}_{ji} + b_{ds} \text{dysl_s}_{ji} + b_{dgs} (\text{dysl_g}_{ji} \cdot \text{dysl_s}_{ji}) + \varepsilon_{ji} \quad (6)$$

gdzie:

Y_{ji} – wynik i -tego ucznia uczącego się w j -tej szkole z danej części egzaminu gimnazjalnego (do 2011 r. znormalizowany ekwikutantowo, od 2012 r. skalowany IRT 2PL/GPCM),

X_{ji} – wynik i -tego ucznia uczącego się w j -tej szkole ze sprawdzianu gimnazjalnego (do 2011 r. znormalizowany ekwikutantowo, od 2012 r. skalowany IRT 2PL/GPCM),

u_{0j} – EWD szkoły,

rok_{ji} – rok zdawania egzaminu gimnazjalnego przez i -tego ucznia uczącego się w j -tej szkole,

tok_{ji} – tok kształcenia i -tego ucznia uczącego się w j -tej szkole (0 – normalny, 1 – wydłużony),

$w(X_{ji}, rok_{ji}, tok_{ji})$ – wielomian opisujący zależność pomiędzy wynikami sprawdzianu a wynikami danej części egzaminu gimnazjalnego,

płeć_{ji} – płeć i -tego ucznia uczącego się w j -tej szkole (0 – mężczyzna, 1 – kobieta),

dysl_g_{ji} – posiadanie przez i -tego ucznia uczącego się w j -tej szkole zaświadczenia o dysleksji na egzaminie gimnazjalnym (0 – nie, 1 – tak),

dysl_s_{ji} – posiadanie przez i -tego ucznia uczącego się w j -tej szkole zaświadczenia o dysleksji na sprawdzianie (0 – nie, 1 – tak),

ε_{ji} – błąd indywidualny dla i -tego ucznia uczącego się w j -tej szkole.

Na podstawie wyników estymacji takiego modelu EWD szkoły szacuje się jako:

$$EWD_j = E(Y_{ji} - \hat{Y}_{ji}) \frac{D^2(u_0)}{D^2(u_0) + \frac{D^2(\varepsilon)}{n_j}} \quad (7)$$

gdzie:

$$\hat{Y}_{ji} = w(X_{jp}, rok_{jp}, tok_{ji}) + b_p płęć_{ji} + b_{dg} dysl_g_{ji} + b_{ds} dysl_s_{ji} + b_{dgs} (dysl_g_{ji} \cdot dysl_s_{ji}) \quad (8)$$

jest przewidywaniem na podstawie efektów stałych,

$D^2(u_j)$ – oszacowanie wariancji efektu losowego dla stałej regresji związanego z podziałem na szkoły (EWD),

$D^2(\epsilon)$ – oszacowanie wariancji błędów indywidualnych,

n_j – liczba uczniów w j -tej szkole; przy czym oszacowania wariancji $D^2(u_j)$ i $D^2(\epsilon)$ otrzymywane są w wyniku estymacji modelu regresji mieszanych efektów.

Tak szacowane wartości wskaźników EWD są ściągniętymi w kierunku zera średnimi z reszt regresji w ramach szkoły. Wielkość tego ściągnięcia jest tym większa, im mniej uczniów, na podstawie których szacowana jest wartość wskaźnika dla danej szkoły i im mniejsze jest zróżnicowanie wyników danej części egzaminu gimnazjalnego między szkołami. Warto też zauważyć, że ponieważ ściągnięcie w kierunku zera realizowane jest poprzez przemnożenie średniej reszt przez wartość mniejszą od jedności, to przesunięcie tak przekształconej wartości względem średniej reszt będzie tym większe, im większa była wartość bezwzględna średniej reszt w ramach szkoły.

Więcej informacji na temat sposobu wyliczania trzyletnich gimnazjalnych wskaźników EWD, w tym o sposobie prezentacji wyników w dwuwymiarowej przestrzeni, można znaleźć w dokumentacji technicznej, dostępnej na stronie internetowej EWD (Pokropek i Żółtak, 2012b). Należy przy tym zaznaczyć, że zapowiadane w tym dokumencie wprowadzenie w 2012 roku modeli uwzględniających zmianę efektywności nauczania w czasie nie doszło do skutku, w związku z negatywną oceną użyteczności takich bardziej skomplikowanych wskaźników dla odbiorców. Uzyskiwane z nich dodatkowe informacje zostały uznane za bardzo trudne do interpretacji w połączeniu z analizą zmiany wartości wskaźników (położenia elips), wobec czego komplikowanie modelu uznano za bezcelowe.

Literatura

Ballou, D., Sanders, W. i Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics* 29, 37–65.

Dolata, R. i Pokropek, A. (2012). Czy warto urodzić się w styczniu? Wiek biologiczny a wyniki egzaminacyjne. [w:] Niemierko, B. i Szmigiel, M. K. (red.), *Regionalne i lokalne diagnozy edukacyjne: XVIII Krajowa Konferencja Diagnostyki Edukacyjnej*, Wrocław, 21–23 września 2012 r. Kraków: Grupa Tomami.

Dolata, R. (2007). Krytyczna analiza metody edukacyjnej wartości dodanej. [w:] Dolata, R. (red.), *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania na podstawie egzaminów zewnętrznych*. Warszawa: Centralna Komisja Egzaminacyjna.

Evans, H. (2008). Value-added in English schools. Referat przedstawiony na National Conference on Value-Added Modeling, University of Wisconsin at Madison, 22–24 kwietnia. Pobrany 28 lutego 2013 z: http://www.wcer.wisc.edu/news/events/VAM%20Conference%20Final%20Papers/VAMinEnglishSchools_HEvens.pdf

Goldstein, H i, Leckie, G. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society. Series A*, 172, 835–851.

Jakubowski, M. (2007). Empiryczna analiza metod szacowania edukacyjnej wartości dodanej dla gimnazjum. [w:] Dolata, R. (red.), *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania na podstawie egzaminów zewnętrznych*. Warszawa: Centralna Komisja Egzaminacyjna.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M. i Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, Kalifornia: RAND.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., Louis, T. A. i Hamilton, L. S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29, 67–101.

OECD. (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paryż: OECD.

Pokropek, A. i Żółtak, T. (2012a). *Trzyletni wskaźnik egzaminacyjny. Dokumentacja techniczna. Wersja 2.0*. Pobrany 15 kwietnia 2013 z: <http://ewd.edu.pl/downloads/Dokumentacja%20techniczna%20v2.0.pdf>

Pokropek, A. i Żółtak, T. (2012b). Nowe modele jednorocznej EWD. [w:] Szmigiel, K., Niemierko, B. (red.), *Regionalne i lokalne diagnozy edukacyjne: XVIII Krajowa Konferencja Diagnostyki Edukacyjnej*, Wrocław, 21–23 września 2012 r. Kraków: Grupa Tomami.

RAISEonline. (2011). *RAISEonline proposed 2011/12 developments*. Pobrany 28 lutego 2013 z: <https://www.raiseonline.org/News.aspx?NewsID=303>

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29, 121–129.

Raudenbush, S. W. i Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307–335.

Reardon, S. F. i Raudenbush, S. W. (2008). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4, 492–519.

Rubin, D. B., Stuart, E. A. i Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, 29, 103–104.

Webster, W. J., Mendro, R. L., Orsak, T. H. i Weerasinghe, D. (1998). *An application of Hierarchical Linear Modeling to estimation of school and teacher effect*. Referat przedstawiony na dorocznym spotkaniu American Educational Research Association, San Diego, Kalifornia, 13–17 kwietnia.

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, 57–67.

Rozdział 4

Wiek i inteligencja ucznia

Kiedy zastanawiamy się nad tym, dlaczego jedni uczniowie uzyskują lepsze wyniki w nauce niż inni, często przychodzą nam na myśl cechy uczniów. Przyczyn zróżnicowania osiągnięć szkolnych upatrujemy na przykład w zdolnościach, motywacji, wytrwałości i zainteresowaniach. Różne charakterystyki uczniów pozwalają przewidywać osiągnięcia szkolne, choć siła ich związku z osiągnięciami jest bardzo zróżnicowana. W niniejszym rozdziale, w kontekście badania trafności metody EWD, skupimy się na dwóch takich cechach: wieku oraz inteligencji. Przeprowadzone analizy pozwolą ocenić trafność metody EWD w dwóch aspektach. Po pierwsze, oszacowanie siły efektów wieku i inteligencji w używanych modelach EWD pozwoli ocenić, na ile skutecznie modele te kontrolują ich znaczenie dla osiągnięć szkolnych. Jeżeli wskaźniki EWD dla szkół szacowane bez kontroli wieku i inteligencji uczniów byłyby znacząco skorelowane z tymi miarami, oznaczałoby to obciążenie metody EWD wynikające z niekompletności modelu wyjaśniającego wyniki egzaminacyjne. Im lepiej uwzględnienie w modelach uprzednich osiągnięć pozwala równocześnie kontrolować te cechy ucznia, tym większa trafność metody. Po drugie, dzięki podłużnemu pomiarowi inteligencji, będzie można oszacować ważny aspekt trafności kryterialnej metody EWD – skorelowanie przyrostu osiągnięć szkolnych z przyrostem inteligencji.

Wiek ucznia to jedna z nielicznych informacji gromadzonych, obok wyników uczniów, w ramach systemu egzaminów zewnętrznych. Dotychczas nie była ona włączana do modeli EWD, choć dwie pozostałe (płeć oraz posiadanie opinii o specyficznych trudnościach w uczeniu się) są w nich uwzględniane. Wiadomo jednak, że wiek ma znaczenie dla osiągnięć szkolnych (por. omówienia w dalszych częściach rozdziału). Powstaje więc pytanie o konsekwencje braku kontroli wieku w modelach EWD. Staje się ono szczególnie ważne w kontekście sposobu, w jaki wprowadzana jest reforma obniżenia wieku obowiązku szkolnego – dobrowolność posyłania sześciolatków do szkoły podczas okresu przejściowego reformy¹ powoduje znaczne zróżnicowanie wiekowe uczniów na jednym poziomie klasowym.

Z kolei inteligencja to jeden z najsilniejszych i najbardziej stabilnych predyktorów sukcesu życiowego, w tym sukcesu szkolnego. W niniejszym rozdziale spojrzymy na nią z kilku perspektyw. Po pierwsze, jako na zmienną, której wpływ należy kontrolować w modelach EWD ze względu na zróżnicowanie szkół pod względem poziomu inteligencji rekrutowanych uczniów.

¹ Reforma obniża wiek rozpoczęcia obowiązkowej nauki w szkole do lat sześciu (Ustawa z dnia 19 sierpnia 2011 r. o zmianie ustawy o systemie oświaty oraz niektórych innych ustaw, Dz.U. 2011 nr 205 poz. 1206). W okresie przejściowym reformy, tj. w roku szkolnym 2012/2013 oraz 2013/2014, objęcie obowiązkiem szkolnym dzieci sześciolatków następuje na wniosek ich rodziców (Ustawa z dnia 27 stycznia 2012 r. zmieniająca ustawę o zmianie ustawy o systemie oświaty oraz o zmianie niektórych innych ustaw, Dz.U. 2012 nr 0 poz. 176).

Dlatego zweryfikujemy, czy modele EWD w obecnym kształcie pozwalają w zadowalającym stopniu kontrolować jej znaczenie dla wyników egzaminacyjnych. Po drugie, przyjrzymy się jej jako zdolności poznawczej, której rozwój może zależeć od efektywności pracy szkoły. Istnienie tej zależności uważamy za jedno z kryteriów trafności wskaźników EWD. Będzie ono argumentem przemawiającym za tym, że wskaźniki EWD odzwierciedlają nie tylko to, jak dobrze szkoła przygotowała uczniów do egzaminu zewnętrznego, lecz są także przejawem takich procesów zachodzących w szkole, które przekładają się na rozwój ucznia.

Nasze dociekania przedstawimy w następującym porządku. Najpierw skupimy się na teoretycznych aspektach zależności między osiągnięciami szkolnymi a wiekiem oraz inteligencją. Następnie, na podstawie analiz teoretycznych postawione zostaną hipotezy, które poddamy empirycznej weryfikacji. Uzyskane wyniki stanowią będą podstawę do sformułowania wniosków dotyczących trafności wskaźników EWD oraz zaleceń dotyczących ich modelowania.

4.1. Wiek a osiągnięcia szkolne

Stwierdzenie, że gimnazjaliści uczęszczający do jednej klasy różnią się wiekiem biologicznym, może wydawać się zaskakujące. W Polsce wiek rozpoczęcia nauki szkolnej wyznacza się bowiem na rok kalendarzowy, w którym dziecko kończy 7 lat². Przywykliśmy więc sądzić, że w polskim systemie edukacji do klas na danym poziomie nauki chodzą uczniowie będący w tym samym wieku. Nic bardziej mylnego. Jeśli pomyślimy tylko o uczniach, którzy zostali posłani do szkoły w wieku zdefiniowanym przez prawo oświatowe, w jednej ławce szkolnej mogą zasiadać uczniowie różniący się wiekiem o rok (najstarszy uczeń będzie urodzony 1 stycznia, a najmłodszy 31 grudnia). Jeśli weźmiemy pod uwagę fakt, że start szkolny niektórych uczniów mógł zostać odroczony lub przyspieszony, maksymalne różnice zaczną sięgać kilku lat. Nieuzyskanie promocji do kolejnej klasy także skutkuje wzrostem zróżnicowania wieku uczniów będących w tej samej klasie.

Mamy więc do czynienia z czterema źródłami wariacji wieku biologicznego uczniów tych samych klas. Pierwszym jest przyjęta definicja wieku rozpoczęcia realizacji obowiązku szkolnego. W przypadku gimnazjalistów skutkuje ona zróżnicowaniem wieku sięgającym 12 miesięcy, które jest związane z tym, którego dnia w roku urodził się dany uczeń. Zwróćmy uwagę na to, że w kontekście analiz związków z osiągnięciami miesiąc urodzenia można uznać za losowy (to, w jakim miesiącu urodził się uczeń raczej nie zależało od jego umiejętności, genów, statusu rodziny

² Akty prawne regulujące tę kwestię to: Ustawa z dnia 7 września 1991 r. o systemie oświaty Dz.U. 1991 nr 95 poz. 425; Ustawa z dnia 19 sierpnia 2011 r. o zmianie ustawy o systemie oświaty oraz niektórych innych ustaw; Dz.U. 2011 nr 205 poz. 1206; Ustawa z dnia 27 stycznia 2012 r. zmieniająca ustawę o zmianie ustawy o systemie oświaty oraz o zmianie niektórych innych ustaw, Dz.U. 2012 nr 0 poz. 176.

pochodzenia itp.). Jeśli jednak skupimy się na uczniach na danym etapie kształcenia, stanowiących główną kohortę wiekową w danej klasie, musimy dopuścić możliwość zakłócenia tej losowości. Z badań prowadzonych w innych krajach wynika, że uczniom młodszym (urodzonym w ostatnich miesiącach roku) częściej odracza się obowiązek szkolny (Routley i de Lemos, 1993 za: Peck i Trimmer, 1994), a także są oni częściej narażeni na powtarzanie klasy (May i Kundert, 1995), więc tym samym nie uczęszczają do jednej klasy ze swoimi „rocznikowymi” rówieśnikami. Podobnie uczniowie urodzeni na początku roku szkolnego częściej mają przyspieszony start szkolny, a także mniejsze ryzyko nieuzyskania promocji do następnej klasy. W Polsce natomiast ustawa o systemie oświaty precyzuje, że uczniowie urodzeni w ostatnich czterech miesiącach roku nie mają możliwości wcześniejszego rozpoczęcia nauki szkolnej (Ustawa z dnia 7 września 1991 r. o systemie oświaty). Tak więc interpretując uzyskane wyniki analiz, powinniśmy dopuścić możliwość, że w głównej kohorcie wiekowej wyznaczonej przez dominujący rok urodzenia uczniów, możemy mieć do czynienia z niedoreprezentowaniem słabszych uczniów urodzonych pod koniec roku (czyli młodszych) oraz lepszych uczniów z początku roku (starszych). Niemniej zjawisko odraczania i przyspieszania startu szkolnego było sporadyczne w latach, w których badani uczniowie rozpoczynali naukę szkolną (co potwierdza analiza częstości urodzenia w poszczególnych miesiącach wśród uczniów młodszych i starszych niż główna kohorta wiekowa – patrz aneks do rozdziału). Dlatego interesująca nas zależność nie powinna być narażona z tego powodu na duże obciążenia.

Drugim źródłem zróżnicowania wieku uczniów jednej klasy jest wcześniejszy start szkolny. Aby dziecko wcześniej rozpoczęło naukę szkolną, potrzebny jest wniosek rodziców i zgoda dyrektora szkoły, a ta jest uzależniona od opinii poradni psychologiczno-pedagogicznej wydanej na podstawie diagnozy dojrzałości psychofizycznej dziecka (Ustawa z dnia 7 września 1991 r. o systemie oświaty). Tak więc decyzja o wcześniejszym posłaniu dziecka do szkoły jest uzależniona od czynników mających duże znaczenie dla jego przyszłych osiągnięć szkolnych. W naturalny sposób prowadzić więc może do związku osiągnięć szkolnych z faktem bycia młodszym w klasie niż główna kohorta wiekowa.

Podobnie odroczenie obowiązku szkolnego – kolejne źródło zróżnicowania wieku biologicznego w klasie – nie jest sprawą losową. Jest ono możliwe tylko w przypadkach „uzasadnionych ważnymi przyczynami” (Ustawa z dnia 7 września 1991 r. o systemie oświaty, Art. 16, ust. 3), które są związane z szansą na sukces szkolny. Uczniowie, którzy później rozpoczęli naukę szkolną, są starsi rocznikiem od swoich klasowych kolegów, jednak tyle samo lat spędzili na nauce szkolnej.

W innej sytuacji są uczniowie, którzy na którymś etapie swojej kariery edukacyjnej nie uzyskali promocji i musieli powtarzać naukę w danej klasie. Zjawisko to w sposób oczywisty nie jest losowe i wiąże się ściśle z osiągnięciami szkolnymi. Prowadzi ono zarówno do zwiększenia zróżnicowania wieku uczniów na danym etapie nauki, jak i jego związku z poziomem osiągnięć.

Widzimy zatem, że z problemem znaczenia wieku dla osiągnięć szkolnych wiążą się cztery zjawiska: różnicowanie wieku biologicznego w ramach głównej kohorty wiekowej, przyspieszenie startu szkolnego, opóźnienie rozpoczęcia nauki szkolnej i brak promocji do następnej klasy.

Sytuacja stanie się jeszcze bardziej złożona, kiedy zaczniemy analizować znaczenie wieku dla wyników uczniów z okresu przejściowego reformy obniżenia wieku obowiązku szkolnego. Różnicowanie wieku w obrębie tych kohort będzie jeszcze większe, a fakt że decyzje o posłaniu dziecka do szkoły w wieku 6 lub 7 lat są uzależnione zarówno od czynników związanych z samym uczniem i jego rodziną, jak i szkołą oraz cechami środowiskową, przyniesie niełatwe do rozwiązania problemy metodologiczne.

4.1.1. Dlaczego warto podjąć problem wieku w kontekście badania trafności wskaźników EWD i rozwoju metodologii ich szacowania?

Analizy związku wieku z osiągnięciami szkolnymi pokazują, że zarówno różnicowanie wieku w ramach głównej kohorty, jak i bycie młodszym lub starszym niż główna kohorta, wyjaśnia część wariacji osiągnięć szkolnych. Gdy weźmiemy pod uwagę tylko tych uczniów, którzy stanowią dominującą kohortę wiekową, wiele badań pokazuje, że uczniowie starsi przewyższają młodszych w poziomie osiągnięć szkolnych (Lee i Fish, 2010; Morrison, Griffith i Alberts, 1997), chociaż tempo rozwoju młodszych uczniów w zakresie umiejętności czytania i matematyki nie jest wolniejsze od obserwowanego wśród starszych kolegów z tej samej klasy (Morrison i in., 1997). Niektóre badania pokazują dodatkowo, że związek osiągnięć z wiekiem jest silniejszy dla dziewczynek niż dla chłopców, a także dla uczniów pochodzących z mniej zamożnych środowisk (Smith, 2009). Co ciekawe, interakcje te obserwuje się dopiero w starszych klasach. Wyniki badań nie zawsze są spójne, jeśli chodzi o wielkość i istotność efektu wieku dla osiągnięć szkolnych (por. Martin, 2009), jednak w wielu krajach powtarza się prawidłowość, że w kolejnych latach nauki szkolnej siła korelacji osiągnięć z wiekiem dla głównej kohorty wiekowej spada (Lee i Fish, 2010; Smith, 2009). Analizy wykorzystujące dane z polskiego systemu egzaminacyjnego pokazały, że zależność osiągnięć od wieku uczniów obserwujemy także w polskim systemie oświaty (Dolata i Pokropek, 2012). Widzimy więc, że różnicowanie wieku w głównej kohorcie ma znaczenie dla wyjaśniania wyników egzaminacyjnych. Należy więc spróbować odpowiedzieć na pytanie, czy i jakie znaczenie może mieć różnicowanie wieku uczniów dla wskaźników EWD, które opierają się na wynikach egzaminów zewnętrznych.

Analizy prowadzone przez badaczy na całym świecie pokazują, że o ile efekt wieku dla uczniów z głównej kohorty jest raczej niewielki, to różnicowanie wiekowe wynikające z wcześniejszego lub późniejszego posłania dziecka do szkoły oraz braku promocji do następnej klasy jest wyraźnie skorelowane z późniejszymi osiągnięciami. Zagadnienia przyspieszonego i opóźnionego startu szkolnego są często rozpatrywane w kontekście problemu gotowości szkolnej (Martin, 2009; May i Kundert,

1995). Zwolennicy odroczenia obowiązku szkolnego widzą w nim szansę na nabycie przez ucznia odpowiedniego poziomu rozwoju poznawczego i społecznego, który pomoże dziecku odnosić szkolne sukcesy. Wyniki badań pokazują jednak, że uczniowie, którzy zostali posłani do szkoły później niż ich rocznikowi rówieśnicy, mają niższe (lub rzadziej: porównywalne) osiągnięcia szkolne niż koledzy z tej samej klasy (Martin, 2009; May i Kundert, 1995; Morrison i in., 1997). Negatywny efekt odroczenia obowiązku szkolnego wynika przede wszystkim z faktu, że decyzje o opóźnieniu startu szkolnego są podejmowane na podstawie cech ucznia i środowiska, w którym żyje, a które są bezpośrednio powiązane z jego przyszłymi osiągnięciami. Przeprowadzone przez badaczy analizy pokazują jednak, że odroczenie obowiązku szkolnego nie jest w stanie w wystarczającym stopniu zrekompensować przyczyn trudności tak, by uczniowie ci dogonili uczniów z tej samej klasy, nie mówiąc już o dorównaniu swoim rówieśnikom, którzy uczą się w klasach starszych. Na drugim krańcu tego zjawiska są uczniowie o przyspieszonym starcie szkolnym. Ci zazwyczaj mają istotnie wyższe wyniki niż ich *de facto* starsi koledzy z tej samej klasy (Mayer i Knutson, 1999). Wynika to z faktu, że w tej nielicznej grupie wcześniej rozpoczynających naukę szkolną są uczniowie wyselekcjonowani ze względu na cechy, które są powiązane z późniejszymi sukcesami szkolnymi.

Badania poświęcone problemowi braku promocji do kolejnej klasy pokazują natomiast, że uczniowie, którzy powtarzali naukę na którymś etapie, mają niższe osiągnięcia szkolne niż ich koledzy z tej samej klasy (Martin, 2009). Efekt ten jest negatywny niezależnie od płci, klasy i wieku ucznia. Tak więc przeciętny statystycznie uczeń pozostawiony na kolejny rok w tej samej klasie nie jest w stanie nadrobić zaległości tak, by dorównać młodszej kohorcie, z którą uczy się po braku promocji.

Widzimy zatem, że zarówno zróżnicowanie wieku wewnątrz dominującej kohorty wiekowej, jak i bycie od niej młodszym lub starszym, jest w wielu krajach istotnie związane z osiągnięciami szkolnymi. Przyjęty obecnie sposób szacowania wskaźników EWD nie zakłada natomiast kontrolowania wieku uczniów z wyjątkiem uwzględniania w modelowaniu EWD zależności dla uczniów, którzy powtarzali klasę na danym etapie kształcenia (w gimnazjum dla wskaźników gimnazjalnych i w szkole ponadgimnazjalnej dla wskaźników maturalnych). Nie bierze się jednak pod uwagę faktu powtarzania klasy, np. w szkole podstawowej w przypadku wskaźników dla gimnazjów, opóźnionego i przyspieszonego startu szkolnego ani zróżnicowania wieku w głównej kohorcie. Gdyby związek wieku z osiągnięciami okazał się silny, a uczniowie nie byłiby losowo przydzielani do szkół ze względu na wiek, uwzględnienie go w modelach EWD mogłoby poprawić precyzję wskaźników. Dane dotyczące daty urodzenia uczniów są przechowywane w bazach egzaminacyjnych, mogłyby więc być do tego celu wykorzystane.

Analizy zaprezentowane w tym rozdziale mają na celu odpowiedź na pytanie, czy uwzględnienie wieku w modelach edukacyjnej wartości dodanej mogłoby poprawić jakość wskaźników. Spróbujemy także naświetlić problemy, które możemy napotkać przy próbie włączenia wieku do modelowania EWD.

4.2. Inteligencja

Kolejnym wyznacznikiem osiągnięć szkolnych, którym zajmiemy się w niniejszym rozdziale, jest inteligencja. Rozpatrywać będziemy ją z dwóch perspektyw związanych z badaniem trafności wskaźników EWD. Po pierwsze, przyjrzymy się zróżnicowaniu międzyszkolnemu inteligencji rekrutowanych uczniów i ewentualnemu wpływowi tych różnic na EWD. Po drugie, spojrzymy na inteligencję jako cechę ucznia, na którą w pewnym stopniu może wpływać szkoła.

W najbardziej ogólnym ujęciu przyjmuje się, że inteligencja to „zdolność do przystosowania się do okoliczności dzięki spostrzeganiu abstrakcyjnych relacji, korzystaniu z uprzednich doświadczeń i skutecznej kontroli nad własnymi procesami poznawczymi” (Nęcka, 2003, s. 26).

Ujęcia tradycyjne przedstawiają inteligencję jako strukturę poznawczych zdolności, czyli czynników pozostających względem siebie w relacjach o różnym charakterze. W zależności od teorii przyjmuje się, że zdolności mają strukturę hierarchiczną lub są od siebie niezależne (Nęcka, 2003). We współczesnych ujęciach inteligencji odchodzi się od jej statycznego (strukturalnego) rozumienia na rzecz ujęć dynamicznych. Kładą one nacisk na procesy poznawcze, poprzez które realizuje się inteligencja, unikając jej opisu za pomocą takich kategorii jak „zdolność” czy „cecha” (np. Nęcka, 2000). Wielość orientacji teoretycznych nie pozwala na ich omówienie w ramach niniejszego rozdziału, dlatego ograniczymy się do omówienia tych ujęć inteligencji, z których czerpaliliśmy.

Pierwsze z nich, mieszczące się w nurcie tradycyjnym i podkreślające hierarchiczną strukturę inteligencji, to ujęcie Raymonda Cattella. Zgodnie z nim inteligencja ogólna obejmuje dwa komponenty – inteligencję płynną oraz skryształizowaną (Cattell, 1971 za: Nęcka, 2003). Inteligencja płynna to zdolność do dostrzegania relacji między symbolami i wykonywania operacji na nich niezależnie od ich znaczenia i poprzednich doświadczeń. Jest uwarunkowana zarówno genetycznie, jak i środowiskowo, ale druga grupa czynników dotyczy stanu zdrowia, odżywienia i innych biologicznych aspektów funkcjonowania organizmu i wyklucza efekty związane z trenin- giem i wychowaniem. Stanowi ona podłoże do rozwoju inteligencji skryształizowanej. Ta z kolei obejmuje zgromadzone w toku życia: wiedzę, umiejętności i doświadczenie, istotne w danym kontekście kulturowym, a także zdolność do korzystania z nich. Mimo że nie wszystkie pre- dykcje wynikające z tej teorii znalazły potwierdzenie empiryczne (np. Ferrer i McArdle, 2004), wzorce wspomnianych zmian zostały dość dobrze udokumentowane, choć wyniki różnią się w zależności od stosowanych miar inteligencji. Wiadomo, że inteligencja płynna i skryształizo- wana mają różne wzorce zmian, chociaż obie początkowo dynamicznie rosną, by następnie rozpocząć systematyczny spadek. Pierwsza z nich wzrasta w dzieciństwie i adolescencji, osią- gając swój najwyższy poziom we wczesnej dorosłości. Następujący później spadek charak- teryzuje się niższym tempem niż tempo wcześniejszego wzrostu i jednocześnie przebiega

szybciej niż w przypadku inteligencji skryształizowanej. Ta z kolei w dzieciństwie i młodości rośnie w szybszym tempie niż inteligencja płynna, osiągając najwyższy poziom w średniej dorosłości, po czym zaczyna stopniowo, choć bardzo powoli, opadać. Tempo tych zmian jest na tyle wolne, że przez dłuższy czas mogą pozostawać niezauważone (Ferrer i McArdle, 2004; McArdle, Ferrer-Caja, Hamagami i Woodcock, 2002). Niektórzy badawcy wskazują jednak, że w okresie dorosłości wiele funkcji poznawczych znajduje się w fazie stabilizacji, a ich wyraźny spadek następuje znacznie później, niż powszechnie się sądzi (np. Schaie, 2005).

Ujęcie, na którym opieramy się w znacznej mierze w badaniach nad trafnością EWD, to ujęcie procesualne. Zgodnie z definicją Edwarda Nęcka (2009, s. 26) przyjmujemy, że inteligencja jest „procesem mobilizacji zasobów poznawczych umysłu ludzkiego”. Zasoby te to formalne i treściowe cechy umysłu (systemu poznawczego), które można wykorzystać podczas rozwiązywania problemów, a w szczególności: formalne cechy tego systemu (tempo przetwarzania informacji, mobilność procesów umysłowych) i jego centralnych modułów nadzorujących i kontrolujących (wydolność uwagi i pamięci roboczej) oraz formalne i treściowe charakterystyki struktur poznawczych (sposób zorganizowania i treść całości wiedzy jednostki). Od efektywności mobilizacji tych zasobów zależy sukces bądź porażka w danym zadaniu poznawczym. Co istotne, w celu wykonania tego samego zadania, różne osoby mogą mobilizować różne zasoby w różny sposób, np. jeżeli jednostka dysponuje niewystarczającą pojemnością pamięci roboczej, może wykorzystać swoją szybkość przetwarzania i dzięki temu przeprowadzić niezbędne operacje na dostępnych fragmentach wiedzy, nim zostaną utracone. Możliwa jest sytuacja przeciwna – osoba o wolnym tempie przetwarzania może dysponować dużą pojemnością pamięci roboczej i dzięki temu utrzymywać w niej dane niezbędne do przeprowadzenia operacji umysłowych aż do momentu ich ukończenia. W ten sposób kompensowane mogą być inne zasoby (Nęcka, 2009).

Ujęcie to podkreśla występowanie zarówno między-, jak i wewnątrzosobniczych różnic w zakresie inteligencji. Jeżeli jednostka, pomimo dostępności niezbędnych zasobów, nie jest w stanie ich w danym momencie zmobilizować, poniesie porażkę. Podobna sytuacja może mieć miejsce, gdy pojawiają się czynniki motywacyjne lub emocjonalne, drenujące zasoby poznawcze i uniemożliwiające ich wykorzystanie podczas wykonywania zadania. Mimo to jednostka, która posiada duże zasoby, częściej i z lepszym skutkiem będzie w stanie je zmobilizować w wystarczającym stopniu, by poradzić sobie z zadaniem, w porównaniu do osoby, która dysponuje niewielkimi zasobami (Nęcka, 2009).

Ujęcie procesualne inteligencji, choć bezpośrednio nie odnosi się do jej zmian w toku życia, podkreśla rolę procesów poznawczych. Stąd zmiany w jej poziomie muszą być powiązane m.in. z ich zmianami rozwojowymi. Wiadomo, że uwaga, pamięć robocza, funkcje wykonawcze rozwijają się intensywnie w dzieciństwie i adolescencji, natomiast starzenie się związane jest z coraz słabszym funkcjonowaniem w tym zakresie (np. Borella, Carretti i De Beni, 2008;

Cepeda, Kramer i Gonzalez de Sather, 2001; Craik i Bialystok, 2006; Glisky, 2007; Zelazo, Craik i Booth, 2004). Wskazuje się jednak, że znaczna część badań nad poznawczym starzeniem ma charakter przekrojowy, co daje fałszywy i zbyt pesymistyczny obraz następujących w toku życia zmian (np. Schaie, 1965) oraz ignoruje sfery, których sprawność pozostaje na wysokim poziomie nawet do 60. roku życia lub dłużej (np. zasób słownictwa, Schaie, 2005; pamięć semantyczna, np. Glisky, 2007).

Oprócz ilości zasobów, istotna jest także ich mobilizacja. To, jakie zasoby są mobilizowane, zmieniać się może wraz z wiekiem, nabywaniem doświadczenia czy wiedzy eksperckiej z dziedziny, której rozwiązywany problem dotyczy (por. Nęcka, 2009). Nie bez znaczenia pozostaje bieżący stan organizmu (zasoby energetyczne) oraz czynniki wpływające na przebieg procesów poznawczych (np. chwilową pojemność pamięci roboczej) rodzaj pobudzenia (mobilizujące vs. lękotwórcze), wykonywane zadania poboczne i inne (por. Nęcka, 2000; 2009). Sugeruje to, że skuteczność mobilizacji zasobów poznawczych w procesie inteligencji może wiązać się ze znajomością zasobów, którymi jednostka dysponuje, umiejętnością rozpoznania, jakich zasobów wymaga zadanie czy skutecznością strategii kontroli i modyfikacji czynników wpływających na przebieg procesów poznawczych (np. redukcji pobudzenia lękotwórczego, kontroli poziomu napięcia mobilizującego, Nęcka, 2000). Wiadomo, że zarówno umiejętności metapoznawcze (Dignath, Buettner i Langfeldt, 2008; Haller, Child i Walberg, 1988), jak i wspomniane strategie (Thayer, 2008), mogą być rozwijane. Szkoła jako miejsce, w którym uczniowie mają zdobywać wiedzę i umiejętności, wydaje się zobligowana do wspierania uczniów w tym zakresie.

4.2.1. Inteligencja a osiągnięcia szkolne

Inteligencja jest jednym z najważniejszych i najbardziej stabilnych predyktorów sukcesu szkolnego dziecka – pozwala wyjaśniać od kilkunastu do kilkudziesięciu procent zmienności wyników w nauce (Deary, Strand, Smith i Fernandes, 2007; Duckworth, Quinn i Tsukayama, 2012; Heaven i Ciarrochi, 2012; Hofer, Kuhnle, Kilian i Fries, 2012; Sternberg, Grigorenko i Bundy, 2001; Teo, Carlson, Mathieu, Egeland i Sroufe, 1996). Co istotne, relacja ta ujawnia się systematycznie pomimo różnorodności stosowanych zarówno ujęć i miar inteligencji, jak i dziedzin i miar osiągnięć szkolnych. Zwykle bywa silniejsza dla przedmiotów ścisłych (np. matematyka, fizyka) w porównaniu do humanistycznych (np. historia, język ojczysty; Ferrer i McArdle, 2004; Heaven i Ciarrochi, 2012; Teo i in., 1996) i zmienia się na różnych etapach edukacyjnych. Wyższe wartości notuje się w wieku szkolnym, niższe – na etapie studiów wyższych, podczas których na znaczeniu zyskują czynniki natury motywacyjnej, zainteresowania i cechy osobowości (Furnham, Chamorro-Premuzic i McDougall, 2003; Lin i Humphreys, 1977).

Nie ma jednak zgody co do natury zależności inteligencja – osiągnięcia szkolne. Postulowano m.in. ich ekwiwalentność, wskazując, że testy je mierzące zawierają niejednokrotnie bardzo

podobne zadania (np. z zakresu arytmetyki, zasobu słownictwa). Wskazywano także wzajemny ich wpływ, argumentując, że inteligencja wyznacza osiągnięcia, ale zdobyta wiedza sprawia, że możemy czerpać z większych zasobów doświadczenia, a więc podwyższa inteligencję. Trzecie ujęcie omawianej relacji postuluje, że inteligencja stanowi źródło osiągnięć szkolnych (por. Watkins, Lei i Canivez, 2007).

W tym kontekście widoczne jest, że wykorzystywanie w badaniach edukacyjnych miar inteligencji uwzględniających te jej aspekty, które obejmują zgromadzoną przez jednostkę wiedzę i doświadczenie (a więc także wiedzę szkolną) jest problematyczne ze względu na nakładanie się zakresów konstruktów. Z drugiej jednak strony, miary uwzględniające różne aspekty funkcjonowania poznawczego (werbalne, niewerbalne, przestrzenne) dają najpełniejszy obraz możliwości poznawczych jednostki.

Różnice w zakresie operacjonalizacji zarówno osiągnięć jak i inteligencji, w połączeniu z odmiennymi schematami badawczymi (przekrojowe vs. podłużne/z odroczonym pomiarem osiągnięć) oraz różnicami w wieku osób badanych, sprawiają, że trudno jest precyzyjnie określić siłę zależności między nimi. Szacunkowo przyjąć można, że w badaniach podłużnych wykorzystujących miary inteligencji ogólnej oraz miary inteligencji, zawierające zarówno komponenty werbalne, jak i niewerbalne, raportowana w literaturze światowej siła relacji między inteligencją a osiągnięciami szkolnymi kształtuje się na poziomie 0,6–0,8 (np. Deary i in., 2007; Duckworth i in., 2012; Heaven i Ciarrochi, 2012; Hofer i in., 2012; Sternberg i in., 2001) i zależec może w pewnym stopniu od kontrolowanych zmiennych. Słabsze zależności z osiągnięciami notuje się w przypadku inteligencji płynnej – w zależności od wieku osób badanych kształtują się one na poziomie około 0,2 do około 0,4 (Bossaert, Doumen, Buyse i Verschueren, 2011; Hofer i in., 2012).

Wśród polskich publikacji w zasadzie nie spotyka się raportów z badań podłużnych uwzględniających relację inteligencja – osiągnięcia szkolne. Przegląd wyników badań przekrojowych przedstawi Roman Dolata (2008), wskazując, że na etapie szkoły podstawowej (od około drugiej klasy) do końca gimnazjum testy inteligencji płynnej korelują z wynikami egzaminów zewnętrznych oraz ocenami szkolnymi na poziomie odpowiednio 0,4–0,5 oraz 0,3–0,4.

Widoczne jest więc, że inteligencja jest ważną zmienną pozwalającą przewidywać osiągnięcia szkolne. By mówić jednak o potrzebie jej kontrolowania w modelach EWD, konieczne jest odpowiedzenie na pytanie, czy gimnazja różnią się między sobą poziomem zdolności poznawczych uczniów, którzy do nich trafiają. Istnienie procesów selekcyjnych na progu gimnazjum, zarówno jawnych (dozwolone ustawowo), jak i ukrytych (Dolata, Jasińska i Modzelewski, 2012), czy zróżnicowanie poziomu inteligencji osób mieszkających na wsi i w mieście (częstkowych danych dostarczają np. Jaworowska i Szustrowa, 1991a; 1991b), sugerują, że rzeczywiście może ono istnieć.

4.3. Pytania badawcze i hipotezy

W ramach badania trafności wskaźników EWD postawiliśmy szereg pytań badawczych i hipotez powiązanych z omawianymi cechami uczniów. W zakresie wieku oparliśmy się w głównej mierze na wynikach badań dotyczących jego znaczenia dla osiągnięć szkolnych. Po pierwsze, ze względu na raportowany w literaturze efekt wieku dla wyników w nauce wewnątrz głównej kohorty wiekowej, spodziewaliśmy się wykryć słabą, choć istotną tego rodzaju zależność. Oczekiwaliśmy także wyższych wyników w nauce wśród uczniów młodszych niż główna kohorta wiekowa oraz niższych wyników wśród uczniów starszych. Wynika to z powiązania osiągnięć szkolnych odpowiednio z przyspieszeniem oraz opóźnieniem startu szkolnego i/lub nieuzyskiwaniem promocji do następnej klasy.

Ponadto postawiliśmy kluczowe dla nas pytanie o konsekwencje nieuwzględniania w modelach EWD danych dotyczących wieku uczniów. Chcieliśmy się dowiedzieć, czy powoduje to obciążenie wskaźników oraz jaki charakter mają ewentualne zniekształcenia. Są to pytania szczególnie ważne z perspektywy zwiększającego się zróżnicowania wiekowego uczniów na jednym poziomie klasowym spowodowanego przejściowym okresem reformy obniżenia wieku obowiązku szkolnego.

W zakresie inteligencji za podstawę teoretyczną przyjęliśmy jej procesualne ujęcie proponowane przez Nęckę (2009, s. 26), zgodnie z którym inteligencja jest „procesem mobilizacji zasobów poznawczych umysłu ludzkiego”. Zakładamy, że stanowi ona korelat osiągnięć szkolnych, lecz jednocześnie, dzięki rozwojowi wiedzy oraz ekspozycji na zadania wymagające mobilizacji zasobów poznawczych, może podlegać rozwojowi w toku kształcenia.

W ramach tego obszaru tematycznego postawiliśmy trzy hipotezy. Pierwsza z nich dotyczy zróżnicowania poziomu inteligencji uczniów. Jak sygnalizowaliśmy wcześniej, ze względu na różnice w lokalizacji szkół czy rekrutację uczniów spoza rejonu na podstawie wyników sprawdzianu lub innych kryteriów związanych z ich zdolnościami, spodziewać się można między-szkolnego zróżnicowania w poziomie inteligencji wśród uczniów pierwszych klas gimnazjum. Jeżeli hipoteza ta nie zostanie odrzucona, będzie to świadczyć o potrzebie kontrolowania inteligencji w modelach EWD ze względu na systematyczne zróżnicowanie jej poziomu wśród uczniów trafiających do różnych szkół. Postawiliśmy także dodatkową hipotezę o istnieniu istotnego zróżnicowania poziomu inteligencji uczniów między oddziałami wewnątrz szkół. Zróżnicowanie to może być zarówno konsekwencją tworzenia wprost oddziałów homogenicznych ze względu na uprzednie zdolności, jak i tworzenia klas profilowanych.

Ponadto spodziewaliśmy się, że uwzględnienie w modelach EWD poziomu osiągnięć szkolnych uczniów, z którym przychodzą oni do gimnazjum, pozwoli pośrednio kontrolować znaczenie inteligencji dla osiągnięć mierzonych w momencie jej opuszczania (w stopniu, w jakim

była ona istotna dla uprzednich osiągnięć). Pozytywny wynik weryfikacji tego oczekiwania będzie świadczyć na korzyść hipotezy o odzwierciedleniu we wskaźnikach EWD faktycznego wkładu szkoły w osiągnięcia uczniów.

Trzecia hipoteza dotyczy różnic w funkcjonowaniu poznawczym uczniów w zależności od efektywności nauczania w szkole. Jak wskazaliśmy wcześniej, wzrost poziomu inteligencji między innymi w adolescencji, jest typowym zjawiskiem rozwojowym. Uważamy, że na przyrost ten wpływ może mieć szkoła, dlatego spodziewamy się wykrycia dwóch efektów – istotnego międzyszkolnego zróżnicowania przeciętnych przyrostów inteligencji uczniów podczas nauki w gimnazjum oraz pozytywnej korelacji przeciętnych przyrostów inteligencji uczniów ze wskaźnikami EWD.

Jaki może być mechanizm przewidywanego „efektu przyrostu”? Co sprawia, że szkoły mogą różnić się stopniem, w jakim wspierają rozwój inteligencji swoich uczniów? Przypuszczamy, że stawianie uczniów przed zadaniami nie przekraczającymi ich możliwości, ale wymagającymi mobilizacji zasobów poznawczych, połączone z udzielaniem odpowiedniego wsparcia w ich rozwiązywaniu (np. trening rozpoznawania wymagań zadania, własnych zasobów, trening strategii radzenia sobie z różnymi zadaniami, kontroli czynników zakłócających ich wykonywanie itp.), może sprzyjać nauce skutecznej mobilizacji tych zasobów przez ucznia. Umiejętność ta ulegać może ekspresji podczas zetknięcia się ze złożonymi zadaniami poznawczymi w innych sytuacjach, a więc także podczas rozwiązywania testu inteligencji.

4.4. Metoda

4.4.1. Pomiar wieku

W analizach wykorzystano trzy wskaźniki opisujące wiek ucznia: wiek w miesiącach, bycie młodszym niż główna kohorta oraz bycie starszym niż główna kohorta. Wartość wskaźników ustalono na podstawie dat urodzenia zaczerpniętych z baz danych egzaminacyjnych. Na tej podstawie utworzono wskaźnik wieku ucznia w miesiącach wycelowany w taki sposób, że przyjmuje on wartość 0 dla uczniów urodzonych w styczniu 1996 roku (główna kohorta wiekowa). Zmienna reprezentująca uczniów młodszych niż główna kohorta przyjmuje wartość 1 dla uczniów urodzonych później niż w roku dominującym (czyli dla 40 uczniów urodzonych w 1997 roku, co stanowi 0,76% wszystkich uczniów objętych badaniem) oraz 0 dla pozostałych. Zmienna reprezentująca uczniów starszych niż główna kohorta wiekowa, przyjmuje wartość 1 dla uczniów urodzonych wcześniej niż uczniowie z głównej kohorty (czyli dla 67 uczniów urodzonych w latach 1993–1995, co stanowi 1,28% wszystkich uczniów objętych badaniem) i 0 dla pozostałych. Ponieważ uczniów starszych niż główna kohorta było niewielu w badanej próbie, nie rozdrabniano tej kategorii w zależności od liczby lat, które dzieliły tych uczniów od

głównej kohorty ze względu na rok urodzenia (rozkład zmiennej „rok urodzenia” znajduje się w aneksie).

Należy zwrócić uwagę na fakt, że wśród uczniów starszych niż główna kohorta mogli się znaleźć zarówno uczniowie, którym odroczono obowiązek szkolny, jak również i tacy, którzy w trakcie nauki w szkole podstawowej powtarzali naukę w tej samej klasie. Na podstawie dostępnych danych nie jesteśmy w stanie odróżnić od siebie tych dwóch sytuacji. Analizowanie ich łącznie jest jednak uzasadnione z tego względu, że wyniki innych badań pokazują (patrz: przegląd literatury na początku niniejszego rozdziału), że możemy spodziewać się podobnych efektów dla obu grup.

Zaznaczyć trzeba, że próba badawcza nie obejmuje uczniów, którzy powtarzali którąś z klas podczas nauki w gimnazjum. Wynika to z faktu, że uczniowie ci nie przystąpili do egzaminu gimnazjalnego w tym samym roku co główna kohorta wiekowa, więc nie posiadamy dla nich wyników egzaminu gimnazjalnego. Wykorzystanie natomiast danych z innej sesji egzaminacyjnej dla sprawdzianu (w przypadku uczniów, którzy powtarzali pierwszą klasę gimnazjum a więc przystąpili do sprawdzianu szóstoklasisty rok wcześniej) nie jest możliwe z uwagi na nieporównywalność wyników egzaminacyjnych między latami i zbyt małą liczbę uczniów w próbie, będących w takiej sytuacji (szczegółowe informacje na temat definicji próby badawczej można znaleźć w rozdziale 1).

4.4.2. Pomiar inteligencji

W badaniu dokonano dwukrotnego pomiaru inteligencji. Pierwszy z nich miał miejsce w drugim semestrze pierwszej klasy gimnazjum (luty – maj 2010 roku, 88,5% uczniów zostało przebadanych w marcu). Zastosowaliśmy test matryc Ravena, wersja standard, forma klasyczna (TMS-K, Jaworowska i Szustrowa, 1991a), tradycyjnie stosowany jako test inteligencji płynnej. Składa się on z 60 zadań zgrupowanych w 5 równolicznych serii (A–E). Każde zadanie ma postać matrycy z brakującym fragmentem, którą należy uzupełnić jednym spośród ośmiu prezentowanych wycinków. Przyjęliśmy, że inteligencja w rozumieniu Nęcki (2009) ujawni się w poziomie wykonania tego testu, gdyż zadania wchodzące w jego skład wymagają mobilizacji zasobów poznawczych, którymi dysponuje jednostka. Poprawne rozwiązanie danego zadania świadczyć będzie o tym, że zmobilizowała ona zasoby wystarczające do poradzenia sobie z nim, porażka – że zmobilizowane zasoby nie były wystarczające. Jednocześnie pomiar z wykorzystaniem testu inteligencji płynnej pozwolił uniknąć problemu nakładania się zakresów konstruktów inteligencji oraz osiągnięć szkolnych.

Drugi pomiar odbył się pod koniec trzeciej klasy gimnazjum (maj 2012 roku). Ponieważ analizy pokazały, że test w wersji standard był za łatwy dla uczniów już na I etapie badania (szczegóły opisano w aneksie), i tym samym nie pozwalał wystarczająco dobrze różnicować uczniów o wyższym poziomie inteligencji, zaistniała potrzeba wykorzystania na III etapie badania

testu trudniejszego. Priorytetem była jednak możliwość wyrażenia wyników z obu pomiarów na tej samej skali, co stanowi warunek przeprowadzenia oszacowań przyrostów inteligencji między dwoma pomiarami. Cele te zostały osiągnięte przez wykorzystanie w drugim pomiarze trzech serii testu matryc Ravena w wersji standard (serie C, D i E) oraz pierwszych 24 pozycji z testu matryc Ravena w wersji dla zaawansowanych (Jaworowska i Szustrowa, 1991b). Dzięki temu testy wykorzystane na obu etapach badania zawierały kilkadziesiąt zadań kotwiczących, które umożliwiły przedstawienie wyników z obu pomiarów na tej samej skali. Zastosowanie natomiast serii zadań trudniejszych na III etapie badania umożliwiło złagodzenie tzw. efektu sufitowego testu, co zwiększyło precyzję oszacowań wyników dla uczniów o wyższym poziomie inteligencji. Oba pomiary miały charakter badania audytoryjnego i były prowadzone przez psychologów przeszkolonych w zakresie procedur realizacji badania.

W celu wyznaczenia wyników testów z obu pomiarów, wyspecyfikowany został strukturalny model czynnikowy³. Analizy wykonano w programie Mplus 6.1, wykorzystując estymator WLSMV i uwzględniając zgrupowanie uczniów w klasach. W modelu tym zmienne mierzalne miały charakter porządkowy, a macierz korelacji pomiędzy nimi wyznaczona została za pomocą korelacji polichorycznych. Model taki jest równoważny dwuparametrycznemu modelowi IRT (patrz np.: Bartholomew, Knott i Moustaki, 2011). Model ten przyjmował, że poziom inteligencji z I etapu badania wpływa na poziom inteligencji z III etapu badania. Ponadto na parametry par pozycji kotwiczących nałożono ograniczenia w postaci ich równości w obu pomiarach⁴. Umożliwiło to wyrażenie obu pomiarów na tej samej skali, a także wykorzystanie w modelowaniu informacji o powtórzoną pomiarze. Model uzyskał dobre dopasowanie do danych (RMSEA = 0,011, CFI = 0,954, TLI = 0,954). Wyniki uczniów, będące wartościami czynnikowymi z tak wyspecyfikowanego modelu, przeliczono na skalę o średniej 0 i odchyleniu standardowym 1 w badanej próbie dla pierwszego pomiaru. Wyniki z III etapu badania wyrażone są na skali wyników z pierwszego pomiaru. Szczegóły dotyczące właściwości psychometrycznych narzędzi oraz modelu, na podstawie którego wyznaczono wartości czynnikowe, opisano w aneksie.

4.4.3. Schemat analizy danych

W celu weryfikacji hipotez policzono szereg hierarchicznych modeli liniowych, które odzwierciedliły złożony schemat doboru próby (Domański i Pokropek, 2011). Zastosowano przede wszystkim modele trzypoziomowe z losową stałą uwzględniającą poziom ucznia, klasy i szkoły.

³ Autorki składają podziękowania Tomaszowi Żółtakowi (Instytut Badań Edukacyjnych, Pracownia EWD) za pomoc w specyfikacji modelu strukturalnego.

⁴ W ten sposób zakotwiczono tylko te zadania, które w wystarczająco dobry sposób spełniały założenie o posiadaniu tych samych parametrów w obu pomiarach. Do tego celu wybrano 30 zadań. Szczegóły znajdują się w aneksie.

Hipotezy weryfikowano, wykorzystując dwie zmienne zależne: łączny wynik egzaminu gimnazjalnego z matematyki i przedmiotów przyrodniczych oraz łączny wynik egzaminu gimnazjalnego z języka polskiego i przedmiotów humanistycznych. Są one wyrażone na skali o średniej 100 i odchyleniu standardowym 15 w populacji uczniów, którzy przystąpili do egzaminu gimnazjalnego w 2012 roku. Zmienne te zostały szczegółowo scharakteryzowane w rozdziale 1, stąd nie będą tu opisywane. Modelowanie, mające na celu weryfikację znaczenia płci, wieku oraz inteligencji dla osiągnięć szkolnych i EWD, przeprowadzono w dwóch etapach. W pierwszym pokazano związek interesującej zmiennej z wynikami egzaminu, budując trzypoziomowy model regresji, w którym stanowiła ona predyktor na poziomie ucznia. W drugim zbudowano trzypoziomowy model EWD uwzględniający tę zmienną, pokazując jej znaczenie dla wyników egzaminów, jeśli kontroli podlegają zmienne uwzględniane w modelach EWD. Strategia ta pozwoliła sprawdzić, w jakim stopniu w modelach EWD kontrolujemy czynniki niezależne od szkoły, omawiane w niniejszym rozdziale. Było to możliwe dzięki zaobserwowaniu, jak zmienia się związek danej zmiennej z wynikami egzaminów, jeśli kontrolujemy w modelu te zmienne, które są uwzględniane w modelach EWD. W prezentacji wyników pokazano także modele EWD bez uwzględnienia dodatkowego interesującego nas predyktora, które stanowiły punkt odniesienia w zakresie weryfikacji hipotez o poprawie dopasowania modeli.

Analizy wykonano za pomocą oprogramowania HLM 6.06, wykorzystując metodę estymacji *full maximum likelihood*. Do oszacowania istotności efektów stałych wykorzystano odporne (*robust*) błędy standardowe⁵. Wartości istotne na poziomie $p < 0,05$ zaznaczono w tabelach pogrubionym drukiem. Analizy wykonano z uwzględnieniem wag warunkowych (patrz: rozdział 1) na wszystkich trzech poziomach, odpowiednio przeskalowanych za pomocą programu HLM (Scientific Software International, 2012).

Weryfikację hipotezy dotyczącej związku przyrostów inteligencji płynnej z EWD przeprowadzono na podstawie analizy korelacji średnich przyrostów inteligencji ze wskaźnikami EWD. Wskaźniki przyrostu inteligencji dla szkół wyznaczono jako Bayesowskie predykcje a posteriori efektów losowych (*best linear unbiased predictions*) z dwupoziomowego modelu mieszanych efektów, w którym zmienną zależną jest poziom inteligencji uczniów w trzeciej klasie, a zmienną niezależną poziom inteligencji uczniów w pierwszej klasie. Wskaźniki EWD również wyznaczono z dwupoziomowych modeli mieszanych efektów jako Bayesowskie predykcje a posteriori efektów losowych, czyli z zachowaniem pełnej analogii do metodologii ich wyliczania dla danych populacyjnych. Do ich wyznaczenia wykorzystano oprogramowanie Stata 12. Analizy wykonano metodą największej wiarygodności.

⁵ Są to oszacowania błędów standardowych odporne na naruszenie założeń dotyczących normalności rozkładu reszt (patrz np. Konarski, 2009, s. 318–324).

4.5. Wyniki

4.5.1. Wiek a wyniki egzaminu matematyczno-przyrodniczego i EWD

W pierwszej kolejności weryfikowano hipotezy dotyczące wieku. W poniższej tabeli zaprezentowano wyniki trzypoziomowych analiz regresji pokazujących znaczenie wieku dla osiągnięć szkolnych z przedmiotów matematyczno-przyrodniczych. Zmienną zależną w tych modelach jest łączny wynik egzaminu gimnazjalnego z matematyki i przedmiotów przyrodniczych wyrażony na skali o średniej 100 i odchyleniu standardowym 15 w populacji uczniów w Polsce.

Tabela 1. Znaczenie wieku dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej i EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej						
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)
oszacowanie efektów stałych						
poziom ucznia						
sprawdzian			11,94	(0,254)	11,94	(0,255)
sprawdzian ²			1,24	(0,113)	1,28	(0,113)
sprawdzian ³			-0,413	(0,065)	-0,425	(0,066)
pleć ^a			-0,92	(0,263)	-1,00	(0,259)
dysleksja na egz. gimnazjalnym ^b			1,79	(0,783)	1,75	(0,778)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym			-3,93	(0,796)	-3,92	(0,794)
wiek ucznia	0,023	(0,061)			-0,152	(0,043)
starszy niż główna kohorta	-12,65	(1,46)			-1,75	(1,01)
młodszy niż główna kohorta	10,09	(2,39)			2,52	(1,82)
stała	99,95	(0,597)	99,90	(0,339)	99,12	(0,388)
oszacowanie efektów losowych						
wariancja efektów szkół	19,90		4,32		4,31	
wariancja efektów oddziałów	20,73		5,40		5,24	
wariancja na poziomie ucznia	156,51		66,53		66,09	
podsumowanie						
deviance	41902,97		37293,23		37255,54	
liczba szacowanych parametrów	7		10		13	

liczba uczniów: 5249, liczba klas: 291, liczba szkół: 150

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji

Model(1) testował znaczenie wieku dla wyników egzaminacyjnych bez kontroli innych zmiennych. Analiza ta pokazuje, że osiągnięcia szkolne z przedmiotów matematyczno-przyrodniczych nie zależą istotnie od wieku uczniów wyrażonego w miesiącach. Wyniki te nie odbiegają od wyników analiz przeprowadzonych na danych dla całej populacji, które pokazały, że efekt wieku w głównej kohorcie wiekowej dla części matematyczno-przyrodniczej egzaminu gimnazjalnego jest bardzo mały i wynosi 0,07–0,11 punktu na skali o odchyleniu standardowym 15 punktów (Dolata i Pokropek, 2012). W badaniu realizowanym na próbie 5249 uczniów ze 150 szkół tak mały efekt okazał się nieistotnie statystycznie różny od zera.

O ile jednak wiek biologiczny w głównej kohorcie nie ma dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej większego znaczenia, to uczniowie starsi lub młodsi niż główna grupa wiekowa istotnie różnią się wynikami od pozostałych uczniów. Uczniowie starsi uzyskują wyniki niemal o jedno odchylenie standardowe niższe niż pozostali. Uczniowie młodsi niż główna kohorta mają natomiast wyniki wyższe o dwie trzecie odchylenia standardowego od pozostałych uczniów. Różnice te są istotne statystycznie.

Dodanie do modelu uprzednich osiągnięć (mierzonych wynikiem sprawdzianu w klasie szóstej szkoły podstawowej) oraz pozostałych zmiennych uwzględnianych w modelach EWD (patrz model (3)) powoduje natomiast, że efekty bycia młodszym lub starszym niż główna kohorta radykalnie maleją i przestają być istotne statystycznie (choć znaczenie dla braku istotności efektów ma na pewno także mała liczebność osób przynależących do tych kategorii). Widoczne jest więc, że uwzględnienie w modelu wyniku sprawdzianu w dużej mierze pozwala kontrolować różnice w przewidywanych osiągnięciach uczniów starszych i młodszymi niż główna kohorta wiekowa. Jest to możliwe ze względu na silny związek uprzednich osiągnięć z przynależnością do tych grup.

Ciekawy i z pozoru niezgodny z intuicją wynik obserwujemy natomiast dla zależności osiągnięć od wieku biologicznego wyrażonego w miesiącach. Dodanie do modelu wyniku sprawdzianu i pozostałych zmiennych kontrolowanych w modelowaniu EWD sprawia, że efekt dla wieku jest ujemny. Innymi słowy, jeśli kontrolujemy uprzednie osiągnięcia, uczniowie starsi uzyskują średnio niższe wyniki niż uczniowie młodsi. Efekt ten nie jest silny (dla 12 miesięcy wynosi niecałe 2 punkty, czyli 0,12 odchylenia standardowego), ale jest istotnie różny od zera. Jak to wytłumaczyć? Uzyskany wynik można zobrazować następująco: jeśli porównamy ze sobą dwóch uczniów, którzy uzyskali na sprawdzianie ten sam wynik (dodatkowo są tej samej płci, mają taki sam status w zakresie specyficznych trudności w uczeniu się), ale różnią się wiekiem biologicznym, to uczeń młodszemu uzyskał wyższy wynik na egzaminie gimnazjalnym. Jednak fakt, że uczeń młodszemu uzyskał na sprawdzianie taki sam wynik jak uczeń starszy oznacza de facto, że miał wyższe umiejętności niż uczniowie w tym samym wieku (wyrażonym w miesiącach wewnątrz głównej kohorty wiekowej). Podstawą tego rozumowania jest

fakt, że w klasie szóstej szkoły podstawowej także obserwuje się zależność osiągnięć szkolnych od wieku, która jest nawet silniejsza niż w trzeciej klasie gimnazjum i pokazuje, że uczniom starszym łatwiej uzyskać wyższy wynik (Dolata i Pokropek, 2012). Ponieważ w oszacowaniach poziomu umiejętności uczniów na podstawie wyniku sprawdzianu nie uwzględnia się poprawki na wiek, otrzymujemy taki efekt. Pamiętać jednak należy, że prawidłowość ta nie dotyczy dowolnej pary gimnazjalistów, ale dwóch „statystycznych uczniów”.

Dodatkowo należy zauważyć, że dodanie do modelu (2) – czyli modelu EWD – zmiennych opisujących wiek (model (3)) istotnie poprawia jego dopasowanie do danych, $\Delta deviance(3) = 37,69$; $p = 0,05$.

W analizach podjęto także próby zweryfikowania założenia o istnieniu takiej samej zależności osiągnięć od wieku w grupach uczniów młodszych i starszych niż główna kohorta, jak i w głównej kohorcie wiekowej, jednak z uwagi na zbyt małą liczebność tych grup nie udało się wiarygodnie oszacować efektów interakcyjnych. Analiza taka byłaby możliwa i warta przeprowadzenia na danych populacyjnych.

4.5.2. Wiek a wyniki egzaminu z przedmiotów humanistycznych i EWD

Kolejna tabela pokazuje wyniki analogicznych analiz dla osiągnięć szkolnych z przedmiotów humanistycznych. Zmienną zależną w tych modelach jest łączny wynik egzaminu gimnazjalnego z języka polskiego i przedmiotów humanistycznych wyrażony na skali o średniej 100 i odchyleniu standardowym 15 w populacji wszystkich uczniów.

Model (1) pokazuje, że wyniki egzaminu gimnazjalnego w części humanistycznej nie zależą istotnie od wieku uczniów wyrażonego w miesiącach. Analizy na pełnych danych egzaminacyjnych pokazały efekt rzędu 0,10–0,14 punktu na analogicznej skali (Dolata i Pokropek, 2012), jednak biorąc pod uwagę rzetelność oszacowania na danych z próby, wyniki te są zbieżne. Możemy powiedzieć, że znaczenie wieku dla osiągnięć gimnazjalistów w zakresie przedmiotów humanistycznych jest tak małe, że można je pominąć.

Uczniowie starsi i młodszy niż główna kohorta różnią się natomiast istotnie od pozostałych uczniów wynikami egzaminu gimnazjalnego w części humanistycznej. Różnice są podobne jak w przypadku przedmiotów matematyczno-przyrodniczych. Uczniowie starsi niż główna kohorta osiągają wyniki średnio o ponad jedno odchylenie standardowe niższe niż pozostali, a uczniowie młodszy niż główna kohorta uzyskują wyniki wyższe o pół odchylenia standardowego. Różnice te są istotne statystycznie.

Tabela 2. Znaczenie wieku dla wyników egzaminu gimnazjalnego w części polonistyczno-humanistycznej i EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części polonistyczno-humanistycznej						
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)
oszacowanie efektów stałych						
poziom ucznia						
sprawdzian			11,91	(0,233)	11,92	(0,232)
sprawdzian ²			-0,405	(0,131)	-0,354	(0,129)
sprawdzian ³			-0,369	(0,077)	-0,392	(0,079)
pleć ^a			1,71	(0,323)	1,63	(0,319)
dysleksja na egz. gimnazjalnym ^b			1,80	(0,688)	1,74	(0,672)
dysleksja na sprawdzianie ^b			-3,09	(0,783)	-3,05	(0,765)
wiek ucznia	0,027	(0,060)			-0,121	(0,040)
starszy niż główna kohorta	-17,59	(1,63)			-3,57	(1,19)
młodszy niż główna kohorta	7,52	(1,96)			-0,354	(1,84)
stała	100,81	(0,637)	100,90	(0,364)	100,29	(0,456)
oszacowanie efektów losowych						
wariancja efektów szkół	17,27		5,07		5,06	
wariancja efektów oddziałów	18,76		3,39		3,31	
wariancja na poziomie ucznia	170,36		77,19		76,77	
podsumowanie						
deviance	42 295,29		37 999,99		37 969,85	
liczba szacowanych parametrów	7		10		13	

liczba uczniów: 5248, liczba klas: 291, liczba szkół: 150

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe;

a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji

Dodanie do modelu (1) zmiennych kontrolowanych w modelach EWD daje wyniki zbliżone do uzyskanych dla przedmiotów matematyczno-przyrodniczych (model (3)). Efekt dla uczniów młodszych niż główna kohorta staje się nieistotnie statystycznie różny od zera. Efekt dla uczniów starszych niż główna kohorta radykalnie maleje, jednak, w odróżnieniu do analiz dla przedmiotów matematyczno-przyrodniczych, jest istotny. Uczniowie ci osiągają wyniki niższe o prawie jedną czwartą odchylenia standardowego od pozostałych uczniów, jeśli

kontrolujemy uprzednie osiągnięcia. Efekt dla wieku biologicznego jest podobny jak w przypadku przedmiotów matematyczno-przyrodniczych. Także dodanie do modelu EWD (modelu (2)) zmiennych opisujących wiek (model (3)) istotnie poprawia jego dopasowanie do danych, $\Delta deviance(3) = 30,13, p = 0,05$.

4.5.3. Podsumowanie znaczenia wieku dla wyników egzaminacyjnych i EWD

Przeprowadzone analizy pokazały, że osiągnięcia szkolne gimnazjalistów nie zależą istotnie od ich wieku biologicznego. Jest to spójne z wynikami innych badań pokazujących słabnące znaczenie wieku dla osiągnięć w kolejnych latach nauki szkolnej (Dolata i Pokropek, 2012; Lee i Fish, 2010; Smith, 2009). O ile jednak wiek biologiczny nie ma znaczenia dla wyników egzaminów gimnazjalnych, to uczniowie starsi lub młodsi niż główna kohorta wiekowa istotnie różnią się osiągnięciami od pozostałych uczniów. Uczniowie starsi niż główna kohorta uzyskują wyniki o około jedno odchylenie standardowe niższe niż pozostali. Uczniowie młodsi niż główna kohorta osiągają natomiast wyniki wyższe o około pół odchylenia standardowego od pozostałych uczniów.

Dodanie jednak do modeli uprzednich osiągnięć oraz pozostałych zmiennych uwzględnianych w modelach EWD pozwala w dużym stopniu kontrolować różnice w przewidywanych wynikach egzaminacyjnych, mianowicie powoduje znaczący spadek efektów dla uczniów starszych i młodszych niż główna kohorta. Niemniej jednak niektóre z tych efektów, mimo że znacznie mniejsze, pozostają istotne statystycznie. Sugeruje to, że dodanie do modeli EWD informacji o fakcie bycia młodszym lub starszym niż główna kohorta, mogłoby pomóc w lepszym wyznaczaniu wskaźników efektywności pracy szkoły.

Analizy wskazują także, że wraz z dodaniem do modeli wyniku sprawdzianu i pozostałych zmiennych kontrolowanych w modelach EWD, efekt wieku biologicznego staje się ujemny i istotny, choć niewielki. Pokazuje to, że jeśli szkoły znacząco różniłyby się ze względu na wiek uczniów z głównej kohorty wiekowej, to placówki, w których uczy się więcej młodszych uczniów, zyskiwałyby, a te z większą liczbą starszych uczniów traciłyby trochę na wartości EWD. Uwzględnienie w modelowaniu EWD wieku biologicznego jako zmiennej kontrolowanej przy przewidywaniu wyników egzaminacyjnych, albo uwzględnienie wieku przy wyznaczaniu poziomu uprzednich osiągnięć, mogłoby złagodzić to zagrożenie.

Obecnie jednak gimnazja nie różnią się znacząco wiekiem uczniów. Liczba uczniów młodszych i starszych niż główna kohorta wiekowa także jest na razie niewielka. To pozwala przyjąć, że wprowadzenie do modeli EWD wieku nie zmieniłoby istotnie wyliczonych wskaźników. Problemu tego nie będzie można jednak pominąć w momencie, gdy kohorty znacznie bardziej zróżnicowane wiekowo (wskutek sposobu obniżenia wieku obowiązku szkolnego) zaczną przystępować do egzaminów zewnętrznych. Ponieważ problem ten nie jest banalny, w szczególności że trzeba sobie poradzić z nielosowym przydziałem uczniów do poszczególnych

grup wiekowych, uzasadnione wydaje się podjęcie już teraz w pracach nad modelowaniem EWD problemu wieku i poszukanie optymalnych rozwiązań.

W modelach wyliczanych na potrzeby tej publikacji, tam gdzie jest to teoretycznie uzasadnione i praktycznie potrzebne, wiek będzie uwzględniany jako zmienna kontrolująca cechy uczniów, na które szkoła nie ma wpływu. Mimo że efekty dla uczniów młodszych lub starszych niż główna kohorta okazały się prawie zawsze nieistotne statystycznie, sugerujemy, że powinny one pozostać w modelu, jako że ich obecność jest uzasadniona teoretycznie. Pozwala także w sposób bardziej poprawny oszacować zależność osiągnięć od wieku biologicznego.

4.5.4. Międzyszkolna i międzyoddziałowa wariancja inteligencji uczniów

W kolejnym kroku weryfikacji poddane zostaną hipotezy dotyczące międzyszkolnego różnicowania inteligencji na początku nauki w gimnazjum.

Przypomnijmy, że gdyby uczniowie byli losowo przydzielani do szkół i oddziałów, średni poziom ich inteligencji w każdej placówce i oddziale byłby bardzo do siebie zbliżony, a co za tym idzie, każdy nauczyciel pracowałby z uczniami o podobnych zasobach. Nie byłoby wówczas potrzeby kontrolowania w modelach efektywności nauczania poziomu inteligencji uczniów. Jednak tak nie jest. Sama segregacja przestrzenna związana z rejonem, w którym mieści się placówka, przyczynia się do różnicowania cech uczniów przyjmowanych do szkoły, nie mówiąc już o jawnych czy ukrytych procesach selekcji do szkół i oddziałów, które w oczywisty sposób przekładają się na zwiększenie różnicowania międzyszkolnego i międzyoddziałowego potencjału intelektualnego uczniów na wejściu.

Zanim zatem zweryfikujemy hipotezę dotyczącą tego, w jakim stopniu modele EWD pozwalają kontrolować zasoby poznawcze uczniów, zobaczymy, jak bardzo szkoły i oddziały wewnątrz szkół różnią się między sobą poziomem inteligencji uczniów. Wskaźniki, którymi się posłużymy, są analogiczne do tych, wykorzystywanych jako miary różnicowania międzyszkolnego wyników nauczania. Ich wyznaczenie opiera się na dekompozycji wariancji, czyli określeniu, jaka część różnicowania indywidualnych wyników wszystkich uczniów może zostać przypisana różnicowaniu średnich grupowych (czyli w naszym przypadku wyników szkół i oddziałów wewnątrz szkół), a jaka różnicowaniu wyników poszczególnych uczniów w szkołach. Następnie poszczególne składowe wariancji (komponent międzyszkolny i międzyoddziałowy) odnosimy do całkowitej wariancji wyników uczniów, a uzyskany wynik informuje nas, jaki odsetek całkowitego różnicowania osiągnięć szkolnych uczniów możemy przypisać podziałowi uczniów na szkoły oraz na oddziały wewnątrz szkół. Gdyby jednostki analizy z wyższego poziomu (szkoły i oddziały wewnątrz szkół) nie różniły się między sobą średnim poziomem inteligencji uczniów, wskaźnik różnicowania przyjmowałby wartość 0%. W przypadku odwrotnym, gdyby całe różnicowanie wyników indywidualnych

uczniów sprowadzałoby się do różnic między szkołami lub oddziałami wewnątrz szkół, wskaźnik osiągałby wartość 100% dla danego poziomu grupowania.

Dekompozycja wariancji na potrzeby opisanego powyżej problemu została przeprowadzona za pomocą trzypoziomowego modelu pustego (uwzględnia tylko pogrupowanie uczniów na szkoły i oddziały) z losową stałą, w którym zmienną zależną jest wynik testu matrycy Ravena z I etapu badania, wyrażony na skali o średniej 0 i odchyleniu standardowym 1. W poniższej tabeli zaprezentowano wyniki analizy zróżnicowania międzyszkolnego i międzyoddziałowego poziomu inteligencji uczniów. Umieszczono w niej oszacowania efektów losowych oraz wskaźniki zróżnicowania wskazujące, jaki procent wariancji wyników testu inteligencji płynnej można przypisać podziałowi uczniów na szkoły, na oddziały wewnątrz szkół oraz łącznie podziałowi na szkoły i oddziały.

Tabela 3. Zróżnicowanie międzyszkolne i międzyoddziałowe inteligencji uczniów – oszacowania efektów losowych z modelu trzypoziomowego z losową stałą (analiza wariacji z efektami losowymi)

zmienna zależna: wynik testu matrycy Ravena na I etapie			
oszacowanie efektów losowych		wskaźniki zróżnicowania	
wariancja efektów szkół	0,047	poziom szkół	5,5%
wariancja efektów oddziałów klasowych	0,076	poziom oddziałów	8,9%
		poziom szkół i oddziałów łącznie	14,4%
wariancja na poziomie ucznia	0,730		

liczba uczniów: 5223; liczba oddziałów: 291; liczba szkół: 150

Podział uczniów na szkoły i oddziały wyjaśnia istotną część zróżnicowania wyników testu inteligencji przeprowadzonego podczas pierwszego roku nauki w gimnazjum. Łącznie podział na szkoły i klasy wyjaśnia 14% wariancji wyników indywidualnych. Co ciekawe, znacznie bardziej są zróżnicowane oddziały wewnątrz szkół (prawie 9% wariancji wyjaśnianej), niż same szkoły (5,5% wariancji daje się wyjaśnić przez podział na szkoły). Dane te potwierdzają to, że mamy do czynienia nie tylko ze zjawiskiem nierównego rozłożenia się potencjału poznawczego uczniów między szkołami, ale także z nielosowym przydziałem uczniów do oddziałów wewnątrz szkół. Gdyby bowiem, niezależnie od zróżnicowania międzyszkolnego, uczniowie byli przydzielani do klas losowo, wskaźnik zróżnicowania międzyoddziałowego przyjmowałby wartość 0%.

Warto zwrócić uwagę na fakt, że zróżnicowanie międzyszkolne poziomu inteligencji uczniów może być trochę zaniżone w stosunku do faktycznego, obserwowanego w populacji, gdyż w zrealizowanej próbie szkół obserwujemy także niedoszacowanie zróżnicowania międzyszkolnego wyników egzaminacyjnych w stosunku do danych z całej populacji (patrz: część rozdziału 1

poświęconego badaniu reprezentatywności próby). Jednakże pomiar inteligencji przeprowadzony został dość późno, bo w drugim semestrze pierwszej klasy, de facto nie dysponujemy więc danymi, które jednoznacznie pozwoliłyby na oszacowanie międzyszkolnego i międzyoddziałowego zróżnicowania na progu tego etapu edukacyjnego. Do tego momentu bowiem pewne zmiany w zakresie poziomu inteligencji uczniów (zależne od szkoły) mogły się już pojawić, co z kolei przełożyć się mogło na wzrost zróżnicowania międzyszkolnego w tym zakresie. Z jednej więc strony mamy czynnik, który mógł spowodować niedoszacowanie tego zróżnicowania, z drugiej – jego przeszacowanie. Nie jest jednak możliwe rozstrzygnięcie, czy doprowadziły do wzajemnego zniesienia się, czy też jeden z nich przeważał, doprowadzając do zniekształceń. Stąd przedstawiony wynik należy traktować jako przybliżony.

Podsumowując, hipoteza o międzyszkolnym i międzyoddziałowym zróżnicowaniu poziomu inteligencji uczniów nie uzyskała jednoznacznego rozstrzygnięcia. Choć zaobserwowaliśmy istotne zróżnicowanie międzyszkolne, nie wiemy, czy i w jakim stopniu jego poziom został zniekształcony w wyniku mniejszego niż w populacji zróżnicowania szkół oraz ewentualnych zależnych od szkoły zmian w poziomie inteligencji uczniów, które zaszły przed momentem pomiaru. Nie znamy też dynamiki wzrostu tego zróżnicowania w wyniku oddziaływań szkoły, a dane te byłyby pomocne w określeniu, na ile pomiar w drugim półroczu pierwszej klasy odzwierciedla zróżnicowanie „na wejściu” do szkoły. Widoczne jest jednak, że zarówno przyrosty inteligencji uczniów w toku nauki w gimnazjum, jak i wyższy poziom zróżnicowania pod koniec nauki w gimnazjum (opisane w kolejnych częściach rozdziału), choć znaczące, nie są bardzo duże. Stąd przypuszczalnie w okresie między rozpoczęciem nauki w gimnazjum a pomiarem inteligencji duże zmiany w zakresie zróżnicowania międzyszkolnego i międzyoddziałowego nie zaszły. Pozwala to, w połączeniu z przeciwnym działaniem wskazanych wcześniej potencjalnych zniekształceń, wnioskować o istnieniu istotnego zróżnicowania w populacji szkół pod względem poziomu inteligencji uczniów, których rekrutują. Dlatego uważamy za zasadną potrzebę kontrolowania poziomu ogólnych zdolności poznawczych w modelach efektywności nauczania.

4.5.5. Inteligencja a wyniki egzaminu matematyczno-przyrodniczego i EWD

W tej i kolejnej części rozdziału omówimy wyniki analiz mających na celu weryfikację hipotezy o pośredniej kontroli w modelach EWD znaczenia inteligencji dla wyników nauczania mierzonych na zakończenie szkoły. Tabela 4 prezentuje wyniki analiz, w których zmienną zależną jest łączny wynik egzaminu gimnazjalnego z matematyki i przedmiotów przyrodniczych, wyrażony na skali o średniej 100 i odchyleniu standardowym 15 w populacji uczniów.

Tabela 4. Znaczenie inteligencji dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej i EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej						
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)
oszacowanie efektów stałych						
poziom ucznia						
sprawdzian			11,96	(0,255)	9,81	(0,269)
sprawdzian ²			1,26	(0,115)	1,24	(0,113)
sprawdzian ³			-0,419	(0,065)	-0,393	(0,068)
pleć ^a			-0,92	(0,266)	-1,47	(0,244)
dysleksja na egz. gimnazjalnym ^b			1,85	(0,766)	1,29	(0,740)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym			-3,99	(0,782)	-3,08	(0,777)
test matryc Ravena (I pomiar)	8,97	(0,244)			3,90	(0,190)
wiek ucznia						
starszy niż główna kohorta						
młodszy niż główna kohorta						
stała	99,78	(0,379)	99,89	(0,338)	100,02	(0,321)
oszacowanie efektów losowych						
wariancja efektów szkół	10,66		4,26		4,10	
wariancja efektów oddziałów	8,03		5,37		4,05	
wariancja na poziomie ucznia	103,24		66,52		59,20	
podsumowanie						
deviance	39 436,29		37 106,49		36 482,94	
liczba szacowanych parametrów	5		10		11	

liczba uczniów: 5222, liczba klas: 291, liczba szkół: 150

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe;

a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji

Model (1) pokazuje siłę zależności wyników egzaminacyjnych od inteligencji bez kontroli innych zmiennych. Przypomnijmy, że wynik testu matryc Ravena jest wyrażony na skali

o odchyleniu standardowym równym 1. Zatem uczniowie różniący się wynikiem testu inteligencji o jedno odchylenie standardowe, różnią się wynikami egzaminu w części matematyczno-przyrodniczej o prawie 9 punktów, czyli o ponad pół odchylenia standardowego. Dodanie do modelu uprzednich osiągnięć oraz innych zmiennych uwzględnianych w modelach EWD sprawia, że efekt ten znacząco maleje, choć pozostaje nadal istotny. Jeśli kontrolujemy zatem uprzednie osiągnięcia, płeć i posiadanie opinii o specyficznych trudnościach w uczeniu się, różnica w zakresie poziomu inteligencji o jedno odchylenie standardowe powoduje różnicę w wynikach egzaminu o około jedną czwartą odchylenia standardowego.

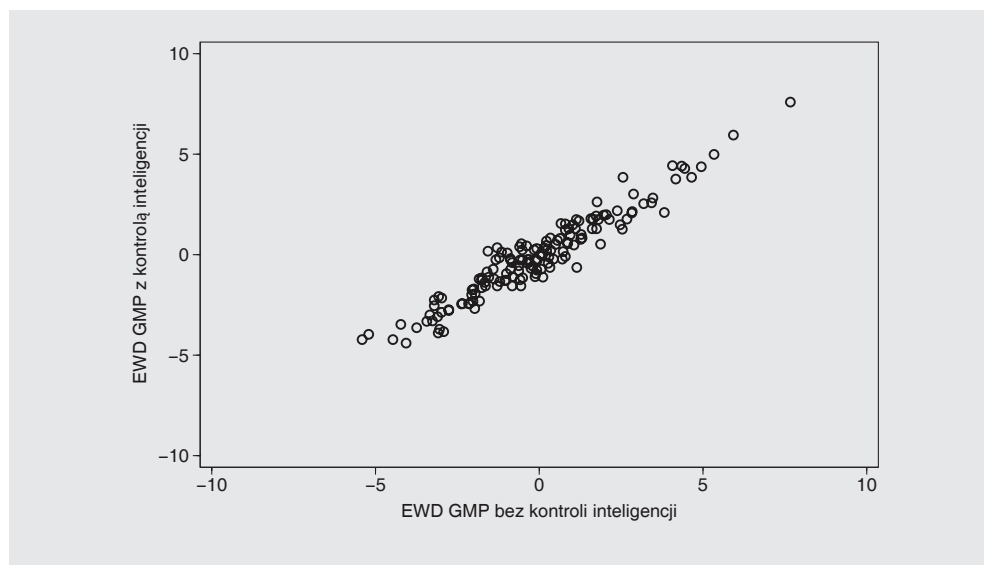
Dodanie do modelu EWD (2) inteligencji znacząco poprawia dopasowanie modelu do danych, $\Delta deviance(1) = 623,56, p = 0,05$, jednak trzeba zwrócić uwagę na to, że w ten sposób przede wszystkim redukujemy niewyjaśnioną wariancję na poziomie uczniów (por. oszacowania efektów losowych dla modeli (2) i (3)). Redukcja wariancji na poziomie szkół jest pomijalna. Porównanie tego ze spadkiem wariancji efektów szkół między modelem EWD (2) a modelem pustym (1), pokazuje, że kontrolując w modelach EWD uprzednie osiągnięcia szkolne, wytrącamy znaczną część zróżnicowania międzyszkolnego potencjału na wejściu. Oznacza to, że dodając do modelu EWD poziom inteligencji ucznia zwiększylibyśmy przede wszystkim precyzję oszacowania EWD, natomiast oszacowania średnich EWD niewiele by się zmieniły. Podkreślić należy, że podobny efekt moglibyśmy uzyskać, zwiększając rzetelność pomiaru osiągnięć „na wejściu” do szkoły.

Na potwierdzenie sformułowanej powyżej tezy wykonano dodatkowe analizy. Obliczono wskaźniki EWD dla szkół z modelu, w którym kontrolujemy tylko uprzednie osiągnięcia, płeć i posiadanie opinii o specyficznych trudnościach w uczeniu się oraz z modelu, w którym dodatkowo uwzględniono wynik testu inteligencji. Wskaźniki te zostały policzone jako Bayesowskie predykcje a posteriori efektów losowych szkół z dwupoziomowego modelu regresji⁶. Predykcje te wyznaczono za pomocą programu Stata 12. Porównanie ich może przynieść odpowiedź na pytanie, jak bardzo różniłyby się od siebie wskaźniki EWD, gdyby przy ich wyliczaniu dodatkowo uwzględnić poziom inteligencji uczniów. Wyniki tego porównania przedstawiono na rysunku 1.

Na osi poziomej znajdują się wartości średniej EWD dla szkół biorących udział w badaniu wyliczone z modelu bez kontroli inteligencji. Na osi pionowej przedstawiono średnie wartości wskaźników EWD wyliczone z modelu, w którym dodatkowo uwzględniono wynik testu inteligencji. Widać, że wskaźniki wyliczone z obu modeli są bardzo do siebie zbliżone. Gdyby uwzględnić błąd ich oszacowania, wskaźniki EWD otrzymane z modelu bez kontroli inteligencji nie różniłyby się istotnie od tych otrzymanych z modelu z kontrolą inteligencji. Współczynnik korelacji między nimi wynosi 0,962.

⁶ Oba modele zostały policzone na próbie szkół i uczniów biorących udział w badaniu i mających dane na wszystkich zmiennych uwzględnianych w obu modelach: wynik sprawdzianu, płeć, zaświadczenie o dysleksji, wynik testu matrycy Ravena z I pomiaru. Tym samym w analizach uwzględniono 150 szkół i 5222 uczniów.

Rysunek 1. Porównanie wskaźników EWD dla części matematyczno-przyrodniczej liczonych bez i z kontrolą inteligencji



Potwierdza to sformułowany wcześniej wniosek, że dodatkowa kontrola inteligencji w modelach EWD nie zmieniałyby znacząco średnich wartości wskaźników dla przedmiotów matematyczno-przyrodniczych. Pozwoliłaby jednak na zmniejszenie błędów ich oszacowań.

4.5.6. Inteligencja a wyniki egzaminu z przedmiotów humanistycznych i EWD

Analogiczne analizy przeprowadzone zostały dla łącznego wyniku egzaminu gimnazjalnego z języka polskiego i przedmiotów humanistycznych. W tabeli 5 zaprezentowano ich wyniki.

Siła zależności wyników egzaminacyjnych z przedmiotów humanistycznych od inteligencji jest zbliżona do obserwowanej w przypadku przedmiotów matematyczno-przyrodniczych (patrz model (1)). Uczniowie różniący się wynikiem testu inteligencji o jedno odchylenie standardowe, różnią się wynikami egzaminu o trochę ponad pół odchylenia standardowego. Podobnie, dodanie do modelu zmiennych uwzględnianych w modelach EWD powoduje radykalny spadek siły zależności wyników testu od inteligencji (patrz model (2)). Efekt nadal pozostaje istotny statystycznie, jednak jest jeszcze słabszy niż dla wyników egzaminu w części matematyczno-przyrodniczej.

Dodanie do modelu EWD (2) inteligencji istotnie poprawiło dopasowanie modelu do danych, $\Delta deviance(1) = 191,97, p = 0,05$.

Tabela 5. Znaczenie inteligencji dla wyników egzaminu gimnazjalnego w części humanistycznej i EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części humanistycznej						
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)
oszacowanie efektów stałych						
poziom ucznia						
sprawdzian			11,91	(0,231)	10,60	(0,256)
sprawdzian ²			-0,409	(0,130)	-0,419	(0,129)
sprawdzian ³			-0,369	(0,076)	-0,353	(0,073)
pleć ^a			1,71	(0,328)	1,38	(0,311)
dysleksja na egz. gimnazjalnym ^b			1,88	(0,686)	1,61	(0,666)
dysleksja na sprawdzianie ^b			-3,07	(0,793)	-2,62	(0,769)
test matryc Ravena (i pomiar)	8,28	(0,249)			2,37	(0,192)
stała	100,51	(0,378)	100,90	(0,365)	100,98	(0,354)
oszacowanie efektów losowych						
wariancja efektów szkół	8,94		4,91		4,67	
wariancja efektów oddziałów	8,82		3,48		2,99	
wariancja na poziomie ucznia	127,03		77,25		74,63	
podsumowanie						
deviance	40 466,17		37 815,89		37 623,93	
liczba szacowanych parametrów	5		10		11	

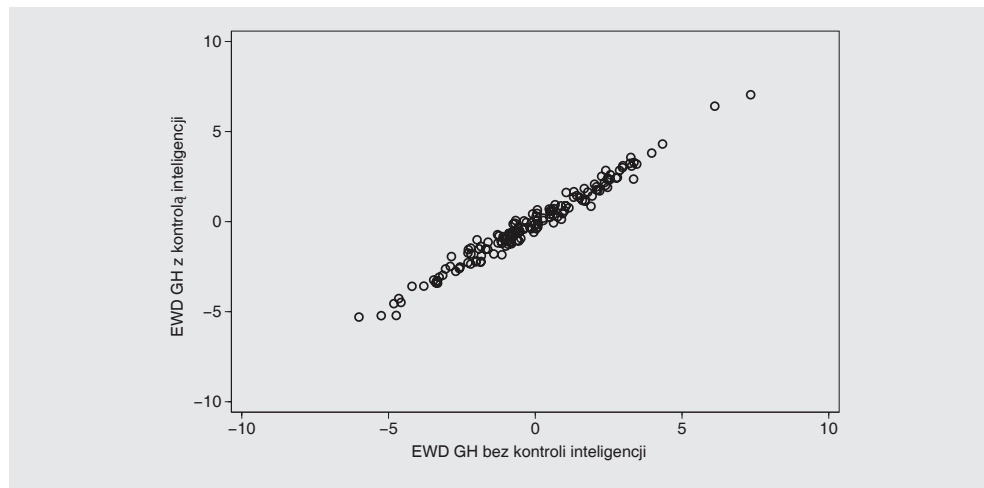
liczba uczniów: 5222, liczba klas: 291, liczba szkół: 150

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe;

a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji

Aby zobaczyć, jak bardzo różniłyby się od siebie wskaźniki EWD dla części humanistycznej, gdyby przy ich wyliczaniu dodatkowo uwzględnić poziom inteligencji uczniów, obliczono wskaźniki EWD dla przedmiotów humanistycznych z modelu bez kontroli inteligencji i z modelu z jej kontrolą. Analizy były przeprowadzone w sposób analogiczny do tych dla przedmiotów matematyczno-przyrodniczych. Wyniki porównania wskaźników przedstawiono na rysunku 2.

Rysunek 2. Porównanie wskaźników EWD dla części humanistycznej liczonych bez i z kontrolą inteligencji



Widoczne jest, że wskaźniki wyliczone z modelu z kontrolą inteligencji (oś pionowa) są niemalże liniowym przekształceniem wskaźników wyliczonych bez kontroli inteligencji (oś pozioma). Uwzględniając błąd ich oszacowania, należałoby dojść do wniosku, że nie różnią się one między sobą. Potwierdza to też bardzo wysoki współczynnik korelacji między nimi, który jest równy 0,988.

Podsumowując, wyniki analiz wskazują, że dodatkowa kontrola poziomu inteligencji uczniów w modelach EWD nie wpłynęłaby na znaczącą zmianę średnich wartości wskaźników. Pozwoliłaby jednak na zmniejszenie błędów ich oszacowań. Jednak podobny efekt można by uzyskać, zwiększając rzetelność pomiaru osiągnięć szkolnych na sprawdzianie. Przyjąć można więc, że uwzględnienie w modelach EWD uprzednich osiągnięć uczniów pozwala w zadowalający sposób kontrolować wpływ inteligencji na szacowanie wskaźników EWD dla szkół.

4.5.7. Efektywność nauczania a przyrost inteligencji uczniów

Inteligencja podlega rozwojowi w toku życia, także w okresie adolescencji. Zebrane w badaniu dane umożliwiają sprawdzenie, jak bardzo zmienia się średni poziom inteligencji uczniów mierzony testem matryc Ravena podczas nauki w gimnazjum. Spodziewamy się powiązania tych zmian z efektywnością pracy szkoły.

Tabela 6 zawiera statystyki opisowe dla obu pomiarów. Przypomnijmy, że wyniki zostały przedstawione na skali o średniej równej 0 i odchyleniu standardowym równym 1 w próbie uczniów, którzy przystąpili do badania na I etapie (uczniowie w pierwszej klasie gimnazjum).

Wyniki pomiaru z III etapu badania zostały przedstawione na skali z I etapu badania, tak więc ich jednostką jest odchylenie standardowe wyników uczniów klas I gimnazjum. W tabeli 6 zaprezentowano wyniki tylko dla tych uczniów, którzy zostali włączeni do próby badawczej ze względu na posiadanie zarówno ich wyniku sprawdzianu w klasie szóstej, jak i wyniku egzaminu gimnazjalnego (szczegóły dotyczące próby badawczej opisano w rozdziale 1). Z tego powodu przedstawiona w tabeli średnia i odchylenie standardowe dla wyników z I etapu badania nie są dokładnie równe 0 i 1.

Tabela 6. Statystyki opisowe dla wyników testu matryc Ravena dla uczniów z próby badawczej (N = 5223)

	średnia	odch. std.	minimum	maksimum
test matryc Ravena (I etap badania)	0,056	0,943	-3,930	3,350
test matryc Ravena (III etap badania)	0,658	1,054	-3,597	4,037

Najbardziej interesujące jest jednak porównanie średnich i odchyłeń standardowych wyników z obu pomiarów. Widoczne jest, że przyrost inteligencji uczniów podczas nauki w gimnazjum wynosi średnio 0,6 odchylenia standardowego wyników uczniów klas pierwszych. Mówiąc bardziej obrazowo: rozkład wyników uczniów klas trzecich przesunął się w kierunku wyższego poziomu inteligencji tak, że jego średnia wypada w punkcie 0,6 odchylenia standardowego wyników uczniów w klasie pierwszej.

W badaniu postawiliśmy hipotezę, że na przyrost inteligencji może mieć wpływ szkoła za sprawą tego, jak dobrze wspiera rozwój poznawczy uczniów, ucząc ich efektywnego wykorzystywania zasobów poznawczych. Oznacza to, że spodziewamy się, że średni przyrost inteligencji będzie różny w różnych szkołach. Jeśli bowiem szkoły nie różniłyby się istotnie tym, jak duży postęp w zakresie poziomu inteligencji osiągnęli uczniowie, obserwowany przyrost należałoby przypisać tylko zmianom rozwojowym.

Zweryfikowanie tej hipotezy wymaga dekompozycji wariancji przyrostów poziomu inteligencji na część wariancji związanej z podziałem uczniów na szkoły i część wariancji opisującej zróżnicowanie indywidualne (wewnątrzszkolne). W tym celu policzono dwupoziomowy model z losową stałą uwzględniający podział uczniów na szkoły. W modelu tym zmienną zależną jest poziom inteligencji uczniów w trzeciej klasie, a zmienną niezależną poziom inteligencji uczniów w pierwszej klasie. Jest to model wartości dodanej, analogiczny do tych wykorzystywanych do wyliczania wskaźników EWD. Wyniki części losowej modelu dają możliwość wyznaczenia wskaźnika zróżnicowania międzyszkolnego. W modelu dekomponuje się bowiem wariancję zmiennej zależnej pod kontrolą zmiennej

niezależnej (czyli w tym przypadku postępów uczniów w poziomie inteligencji) na część zróżnicowania związanego z podziałem uczniów na szkoły i część wariancji, którą można przypisać zmienności indywidualnej.

Tabela 7. Wskaźnik zróżnicowania międzyszkolnego przyrostów inteligencji – oszacowania efektów losowych z modelu z losową stałą (liczba szkół: 150, liczba uczniów: 5223)

wariancja efektów szkół	0,060
wariancja na poziomie ucznia	0,354
wskaźnik zróżnicowania międzyszkolnego	14,5%

Obie składowe wariancji są istotnie statystycznie większe od zera (na poziomie istotności $p < 0,05$). Oznacza to, że szkoły istotnie różnią się tym, jakie średnie przyrosty inteligencji uzyskali ich uczniowie podczas dwóch lat nauki. Podział uczniów na szkoły wyjaśnia 14,5% wariancji przyrostów inteligencji uczniów, co pokazuje wskaźnik zróżnicowania międzyszkolnego. W takim stopniu więc szanse ucznia na wysoki przyrost inteligencji zależał od tego, do jakiej szkoły uczęszczał.

Jak opisano w części teoretycznej rozdziału, pozytywna korelacja między przyrostami w zakresie inteligencji a wskaźnikami EWD będzie silnym dowodem na trafność tych ostatnich⁷. Oznaczać to bowiem będzie, że wartość EWD nie jest tylko związana z tym, jak uczniowie rozwiązują testy egzaminacyjne, ale odzwierciedla także zmiany w skuteczności mobilizowania zasobów poznawczych, czyli w poziomie inteligencji.

W celu zweryfikowania tej hipotezy policzono średnie przyrosty inteligencji dla szkół oraz EWD badanych szkół z wykorzystaniem danych od tych uczniów, dla których dysponowano pomiarami inteligencji (wykorzystano zatem dane z 99,5% badanej próby). Wskaźniki przyrostu inteligencji dla szkół wyznaczono jako Bayesowskie predykcje a posteriori efektów losowych z dwupoziomowego modelu mieszanych efektów, w którym zmienną zależną jest poziom inteligencji uczniów w trzeciej klasie, a zmienną niezależną poziom inteligencji uczniów w pierwszej klasie. Wyliczone z niego efekty losowe można interpretować jako wartość dodatkową dla przyrostów inteligencji. Mówią one o tym, ile szkoła „dodała” uczniom do ich przyrostu inteligencji w porównaniu do innych szkół, do których chodzili uczniowie o takim samym poziomie inteligencji „na wejściu”.

Wskaźniki EWD wyznaczono również z dwupoziomowych modeli mieszanych efektów jako Bayesowskie predykcje a posteriori efektów losowych. Predykcje te wyznaczono za pomocą programu Stata 12. Specyfikacja i opis modeli znajdują się w aneksie.

⁷ Choć brak związku nie będzie wystarczającym powodem do podważenia trafności wskaźników EWD.

Analogiczne wskaźniki policzono także dla oddziałów klasowych⁸, ponieważ założono, że miary te powinny lepiej odzwierciedlać to, co się faktycznie dzieje z uczniem w szkole i to, jakim procesom edukacyjnym, które mogą wspierać rozwój ogólnych i specyficznych zdolności i umiejętności, jest poddawany. Dlatego też spodziewano się wyższej korelacji wskaźników na poziomie oddziałów niż szkół. Tabela 8 zawiera wyniki analiz.

Tabela 8. Korelacja wartości dodanej przyrostów inteligencji ze wskaźnikami EWD – porównanie dla szkół i klas

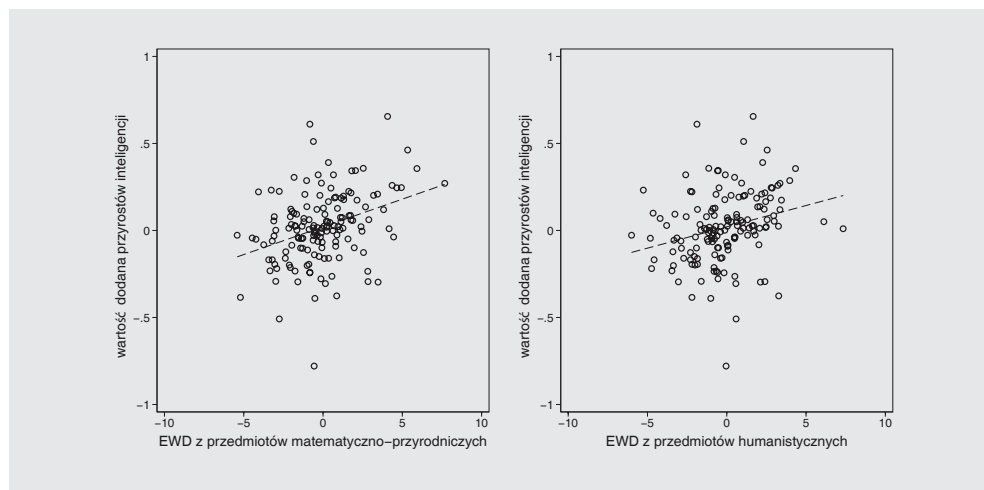
	macierz korelacji dla wskaźników dla szkół ⁹		macierz korelacji dla wskaźników dla oddziałów	
	przyrost IQ	EWD MP	przyrost IQ	EWD MP
EWD MP	0,33		0,32	
EWD PH	0,24	0,59	0,29	0,55

wszystkie współczynniki korelacji są istotne na poziomie istotności $p < 0,05$

Związki, zgodnie z hipotezą, okazały się istotne statystycznie i pozytywne. Oznacza to, że w toku nauki w gimnazjach o wyższej edukacyjnej wartości dodanej uczniowie osiągają średnio większe przyrosty inteligencji. Wartości współczynników korelacji nie są jednak wysokie, co wskazuje na niezbyt dużą siłę zależności. Dla wskaźników na poziomie szkół zależność ta jest troszkę silniejsza dla EWD z przedmiotów matematyczno-przyrodniczych. Nie obserwujemy jednak wzrostu siły zależności dla wskaźników wyznaczonych dla oddziałów. Z czego to może wynikać? Związek pomiędzy dwoma zmiennymi jest tym lepiej oszacowany, im precyzyjniejszy jest pomiar. Mniejsza rzetelność będzie zaniżać korelację między zmiennymi, gdyż wyniki pomiaru są bardziej narażone na losowe odchylenia od wyników prawdziwych. Wskaźniki dla klas są wyliczone dla mniejszej liczby obserwacji, tak więc są bardziej narażone na wpływ losowych czynników lub obciążenie wpływem odstających wyników. Ponadto wskaźniki EWD dla klas są bardziej ściągane do średniej z uwagi na mniejszą liczebność grup i zastosowaną metodę wyliczania efektów losowych (np. Rabe-Hesketh, 2008). To wszystko może powodować brak wzrostu korelacji dla wskaźników na poziomie oddziałów.

⁸ Wskaźniki dla klas (wartość dodaną dla inteligencji i EWD) wyznaczono z dwupoziomowych modeli mieszanych efektów uwzględniających poziom oddziału i ucznia.

⁹ Analogiczne analizy przeprowadzono także na zbiorze danych bez szkół odstających ze względu na wartość dodaną dla przyrostów inteligencji. Z analizy wykluczono dwie szkoły o najniższych wartościach tego wskaźnika, czyli takich które świadczyłyby o braku przyrostu inteligencji lub regresie w tym obszarze (szkoły o wartościach odstających można zauważyć na wykresie prezentującym związek przyrostów inteligencji ze wskaźnikami EWD – rysunek 3). Wyniki analiz nie zmieniły się znacząco po wykluczeniu obserwacji odstających. Wartości współczynników korelacji między przyrostem inteligencji a wskaźnikami EWD wyniosły: 0,33 (dla korelacji z EWD MP) i 0,29 (dla EWD GH). Wyniki te potwierdziły fakt, że na obserwowany związek nie mają wpływu szkoły o odstających wartościach w zakresie wskaźników przyrostów inteligencji.

Rysunek 3. Związek przyrostów inteligencji ze wskaźnikami EWD dla szkół

Dla lepszego zobrazowania omawianych zależności, związek między średnimi przyrostami inteligencji i EWD dla wskaźników policzonych dla szkół zaprezentowano na wykresie (rysunek 3). Na osi pionowej przedstawiono wartość dodaną przyrostów inteligencji, a na osiach poziomych – wskaźniki EWD z przedmiotów matematyczno-przyrodniczych (z lewej) oraz przedmiotów humanistycznych (z prawej). Kropkami zaznaczono pozycję każdej z badanych szkół na tych dwóch wymiarach. Linia przerywana pokazuje liniowe dopasowanie do danych.

Szkoły, które na wymiarze osi pionowej znajdują się w okolicy zera, to gimnazja, w których średni przyrost inteligencji był zbliżony do średniego przyrostu w populacji. W szkołach, które znajdują się powyżej punktu 0, uczniowie zrobili większy niż przeciętny postęp w zakresie przyrostu inteligencji. I analogicznie, w szkołach znajdujących się poniżej 0 – uczniowie osiągnęli średnio mniejszy wzrost inteligencji niż przeciętny w populacji wszystkich uczniów. Na wykresie można zauważyć, że szkoły o powyżej przeciętnej EWD są częściej placówkami o ponadprzeciętnym przyroście inteligencji. Względnie słaba siła zależności objawia się natomiast dość dużym rozproszeniem punktów.

Przeprowadzone analizy pozwoliły na przyjęcie postawionych hipotez. Poziom inteligencji uczniów wzrósł w toku nauki w gimnazjum o około 0,6 odchylenia standardowego na skali wyników uczniów z pierwszych klas. Jednocześnie szkoły różnią się istotnie pod względem średniego poziomu tego przyrostu – przynależność do szkół wyjaśnia 14,5% wariancji przyrostów inteligencji wśród uczniów. Ponadto średnie przyrosty inteligencji w szkołach słabo, choć istotnie korelują ze wskaźnikami EWD. Wraz ze wzrostem EWD rośnie wielkość przeciętnego przyrostu inteligencji. Stanowi to argument przemawiający za trafnością wskaźników EWD.

4.6. Podsumowanie

W niniejszym rozdziale przeanalizowaliśmy problem znaczenia wieku oraz inteligencji uczniów dla badania trafności wskaźników EWD.

W zakresie wieku pytaliśmy o konsekwencje nieuwzględniania tej zmiennej w modelach EWD. Analizy wykazały brak zależności między wiekiem uczniów z głównej kohorty wiekowej a ich wynikami na egzaminie gimnazjalnym, choć zaobserwowano istotny efekt dla uczniów młodszych i starszych niż główna kohorta. W obu częściach egzaminu gimnazjalnego pierwsi z nich uzyskali wyniki przeciętnie o pół odchylenia standardowego wyższe niż pozostali uczniowie, natomiast drudzy – o około jedno odchylenie standardowe niższe. Jednakże w sytuacji kontroli uprzednich osiągnięć oraz innych zmiennych uwzględnianych w modelach EWD, efekty te traciły istotność statystyczną z wyjątkiem uczniów starszych niż główna kohorta w modelu dla przedmiotów humanistycznych. Pozwala to na sformułowanie wniosku, że modele EWD dobrze radzą sobie z efektami bycia młodszym lub starszym niż główna kohorta wiekowa, choć dodanie tej informacji do modelu pozwoliłoby nieco precyzyjniej szacować EWD.

Analizy wykazały także, że w modelach EWD, efekt wieku biologicznego jest ujemny, choć niewielki. Jeśli szkoły znacząco różniłyby się ze względu na wiek uczniów z głównej kohorty wiekowej, to placówki, w których uczy się więcej młodszych uczniów, zyskiwałyby, a te z większą liczbą starszych uczniów traciłyby na wartościach wskaźników EWD. Obecnie mamy jednak do czynienia z niewielkim zróżnicowaniem międzyszkolnym na poziomie gimnazjum w zakresie wieku uczniów w głównej kohorcie i liczby uczniów młodszych lub starszych niż główna kohorta. Dlatego wprowadzenie do modeli EWD informacji o wieku uczniów nie zmieniłoby wskaźników w sposób znaczący. Problem uwzględnienia wieku w szacowaniu EWD nabierze znaczenia w przyszłości, gdy do gimnazjów trafią uczniowie, którzy rozpoczęli naukę szkolną w okresie przejściowym reformy obniżenia wieku szkolnego. Wtedy zróżnicowanie wiekowe uczniów stanie się znacznie większe, co sprawi, że istotne stanie się uwzględnienie danych na temat w modelowaniu.

W kontekście inteligencji weryfikowaliśmy hipotezę o zróżnicowaniu szkół ze względu na poziom inteligencji uczniów, których rekrutują. Ponadto sprawdzaliśmy, czy obecne modele pozwalają w wystarczającym stopniu kontrolować znaczenie inteligencji dla wyników egzaminacyjnych oraz czy przyrosty poziomu inteligencji powiązane są z EWD szkół. Przeprowadzone analizy wykazały, że podział na szkoły i klasy wyjaśnia łącznie 14,4%, a podział na szkoły – 5,5% wariacji poziomu inteligencji uczniów, choć wynik ten może być zniekształcony. Z jednej strony zaniżać go może przypuszczalnie niższe niż w populacji zróżnicowanie szkół z próby w tym zakresie, zaś zawyżać – potencjalne zmiany w inteligencji

uczniów, które zaszły w wyniku oddziaływań szkoły przed momentem pomiaru. Dlatego nie można jednoznacznie rozstrzygnąć, jaka jest wielkość efektu. Ponieważ jednak dane sugerują, że mamy do czynienia z istotnym zróżnicowaniem międzyszkolnym inteligencji rekrutowanych uczniów, kontrola ogólnych zdolności poznawczych w modelach efektywności kształcenia dla polskich gimnazjów jest konieczna – w przeciwnym razie obliczane na ich podstawie wskaźniki nie odzwierciedlałyby wkładu szkoły w osiągnięcia uczniów. Kontrolę tę w zadowalającym stopniu zapewnia uwzględnienie w modelach wyników sprawdzianu po klasie szóstej. Wprowadzie w modelach EWD z uwzględnioną inteligencją jej efekt pozostaje istotny statystycznie oraz istotnie poprawia dopasowanie, jednak wyjaśnia przede wszystkim wariancję na poziomie ucznia. Wobec tego uwzględnianie inteligencji w modelach zwiększyłoby precyzję szacowania EWD, jednak w średnich wynikach dla szkół nie zaszłyby znaczące zmiany. Nadmienić należy, że bardziej rzetelny pomiar na sprawdzianie zapewniłby lepszą kontrolę inteligencji w modelach EWD. To, że jej efekt pozostaje wciąż istotny, może wynikać właśnie z niskiej rzetelności testu na wejściu.

Zaobserwowaliśmy także, że szkoły różnią się średnią wielkością przyrostów inteligencji swoich uczniów (podział na szkoły wyjaśnia 14,5% tego zróżnicowania). Ponadto znaleźliśmy pozytywną korelację pomiędzy wskaźnikami EWD szkół a średnimi przyrostami inteligencji ich uczniów w toku nauki w gimnazjum. Oznacza to, że wraz ze wzrostem EWD szkoły rosną przyrosty inteligencji uczniów.

Podsumowując, uzyskane wyniki stanowią argument przemawiający za trafnością wskaźników EWD – odzwierciedleniu w nich faktycznego wkładu szkoły w osiągnięcia uczniów. Co więcej, ich powiązanie z przyrostami inteligencji sugeruje, że nie wskazują jedynie na to, na ile dobrze szkoła radzi sobie z przygotowaniem uczniów do poradzenia sobie na egzaminie zewnętrznym, lecz są wyrazem zachodzących w szkole procesów, które mają znaczenie dla rozwoju poznawczego uczniów. Przeprowadzone analizy potwierdziły także, że wyniki egzaminów zewnętrznych analizowane bez metody EWD nie są dobrą miarą efektywności pracy szkoły (choć często się je tak traktuje) z powodu ich silnego uzależnienia od poziomu inteligencji uczniów oraz – w mniejszym stopniu – ze względu na wiek zdających.

Literatura

- Bartholomew, D. J., Knott, M. i Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley i Sons.
- Borella, E., Carretti, B. i De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica*, 128(1), 33 – 44. doi:10.1016/j.actpsy.2007.09.008
- Bossaert, G., Doumen, S., Buyse, E. i Verschueren, K. (2011). Predicting children's academic achievement after the transition to first grade: A two-year longitudinal study. *Journal of Applied Developmental Psychology*, 32(2), 47 – 57. doi:10.1016/j.appdev.2010.12.002
- Cepeda, N. J., Kramer, A. F. i Gonzalez de Sather, J. C. M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, 37(5), 715 – 730. doi:10.1037/0012-1649.37.5.715
- Craik, F. I. M. i Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences*, 10(3), 131 – 138. doi:10.1016/j.tics.2006.01.007
- Deary, I. J., Strand, S., Smith, P. i Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13–21. doi:10.1016/j.intell.2006.02.001
- Dignath, C., Buettner, G. i Langfeldt, H.-P. (2008). How can primary school students learn self-regulated learning strategies most effectively? *Educational Research Review*, 3(2), 101 – 129. doi:10.1016/j.edurev.2008.02.003
- Dolata, R. (2008). *Szkoła, segregacje, nierówności*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Dolata, R., Jasińska, A. i Modzelewski, M. (2012). Wykorzystanie egzaminów krajowych jako instrumentu polityki oświatowej na przykładzie procesu różnicowania się gimnazjów w dużych miastach. *Polityka Społeczna*, 1, 41 – 46.
- Dolata, R. i Pokropek, A. (2012). Czy warto urodzić się w styczniu? Wiek biologiczny a wyniki egzaminacyjne. [w:] Niemierko, B. i Szmigiel, M. K. (red.), *Regionalne i lokalne diagnozy edukacyjne: XVIII Krajowa Konferencja Diagnostyki Edukacyjnej*, Wrocław, 21 – 23 września 2012 r. Grupa Tomami.
- Domański, H. i Pokropek, A. (2011). *Podziały terytorialne, globalizacja a nierówności społeczne: wprowadzenie do modeli wielopoziomowych*. Warszawa: Wydawnictwo IFiS PAN.
- Duckworth, A. L., Quinn, P. D. i Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104(2), 439 – 451. doi:10.1037/a0026280

- Ferrer, E. i McArdle, J. J. (2004). An Experimental Analysis of Dynamic Hypotheses About Cognitive Abilities and Achievement From Childhood to Early Adulthood. *Developmental Psychology*, 40(6), 935 – 952. doi:10.1037/0012-1649.40.6.935
- Furnham, A., Chamorro-Premuzic, T. i McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences*, 14(1), 49 – 66. doi:10.1016/j.lindif.2003.08.002
- Glisky, E. L. (2007). Changes in cognitive function in human aging. [w:] *Brain aging: Models, methods, and mechanisms* (s. 3 – 20). Boca Raton, FL, US: CRC Press.
- Haller, E. P., Child, D. A. i Walberg, H. J. (1988). Can Comprehension Be Taught?: A Quantitative Synthesis of „Metacognitive” Studies. *Educational Researcher*, 17(9), 5 – 8. doi:10.3102/0013189X017009005
- Heaven, P. C. L. i Ciarrochi, J. (2012). When IQ is not everything: Intelligence, personality and academic performance at school. *Personality and Individual Differences*, 53(4), 518 – 522. doi:10.1016/j.paid.2012.04.024
- Hofer, M., Kuhnle, C., Kilian, B. i Fries, S. (2012). Cognitive ability and personality variables as predictors of school grades and test scores in adolescents. *Learning and Instruction*, 22(5), 368 – 375. doi:10.1016/j.learninstruc.2012.02.003
- Jaworowska, A. i Szustrowa, T. (1991a). *Podręcznik do Testu Matryc Ravena. Wersja Standard (1956). Polska standaryzacja 1989* (5;11 – 15;11). Warszawa: Pracownia Testów PTP.
- Jaworowska, A. i Szustrowa, T. (1991b). *Podręcznik do Testu Matryc Ravena. Wersja dla Zaawansowanych. Polska standaryzacja*. Warszawa: Pracownia Testów PTP.
- Konarski, R. (2009). *Modele równań strukturalnych: teoria i praktyka*. Wydawnictwo Naukowe PWN.
- Lee, J. i Fish, R. M. (2010). International and Interstate Gaps in Value-Added Math Achievement: Multilevel Instrumental Variable Analysis of Age Effect and Grade Effect. *American Journal of Education*, 117(1), 109 – 137.
- Lin, P. i Humphreys, L. G. (1977). Predictions of Academic Performance in Graduate and Professional School. *Applied Psychological Measurement*, 1(2), 249 – 257. doi:10.1177/014662167700100211
- Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology*, 101(1), 101 – 114. doi:10.1037/a0013100
- May, D. C. i Kundert, D. K. (1995). Does Delayed School Entry Reduce Later Grade Retentions and Use of Special Education Services? *Remedial and Special Education*, 16(5), 288 – 94.

- Mayer, S. E. i Knutson, D. (1999). Does the timing of school affect how much children learn? [w:] Mayer, S. E. i Peterson, P. E. (red.), *Earning and Learning: How Schools Matter* (s. 79 – 102). Washington: Brookings Institution Press.
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F. i Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38(1), 115 – 142. doi:10.1037//0012-1649.38.1.115
- Morrison, F. J. F. J., Griffith, E. M. i Alberts, D. M. (1997). Nature-nurture in the classroom: entrance age, school readiness, and learning in children. *Developmental Psychology*, 33(2), 254 – 262.
- Nęcka, E. (2000). *Pobudzenie intelektu: zarys formalnej teorii inteligencji*. Kraków: Universitas.
- Nęcka, E. (2003). *Inteligencja: geneza, struktura, funkcje*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Nęcka, E. (2009). Inteligencja jest procesem, [w:] Koziński, J. (red.), *Nowe idee w psychologii: psychologia XXI wieku* (s. 21 – 39). Sopot: Gdańskie Wydawnictwo Psychologiczne.
- Peck, B. i Trimmer, K. (1994). The effect of birthday on access to university. Zaprezentowano na: Australian Association for Research in Education, Newcastle, New South Wales: AARE Conference Papers Collection.
- Rabe-Hesketh, S. (2008). *Multilevel and longitudinal modeling using Stata* (2nd ed.). College Station, Tex: Stata Press Publication.
- Routley, V. i de Lemos, M. (1993). Changing trends in school entry age in Victoria. *Journal of Early Childhood*, 18(2), 30 – 37.
- Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, 64, 92 – 107.
- Schaie, K. W. (2005). *Developmental Influences on Adult Intelligence*. Oxford University Press.
- Scientific Software International. (2012). Pobrano 19 marzec 2012, z: <http://www.ssicentral.com/hlm/example6-2.html>
- Smith, J. (2009). Can Regression Discontinuity Help Answer an Age-Old Question in Education? The Effect of Age on Elementary and Secondary School Achievement. *The B.E. Journal of Economic Analysis & Policy*, 9(1). doi:10.2202/1935-1682.2221
- Sternberg, R. J., Grigorenko, E. i Bundy, D. A. (2001). The Predictive Value of IQ. *Merrill-Palmer Quarterly*, 47(1), 1 – 41. doi:10.1353/mpq.2001.0005
- Teo, A., Carlson, E., Mathieu, P., Egeland, B. i Sroufe, L. A. (1996). A Prospective Longitudinal Study of Psychosocial Predictors of Achievement. *Journal of School Psychology*, 34(3), 285 – 306. doi:10.1016/0022-4405(96)00016-7

Thayer, R. E. (2008). *Źródło codziennych nastrojów: kontrola energii, napięcia i stresu*. Gliwice: Wydawnictwo Helion.

Ustawa z dnia 7 września 1991 r. o systemie oświaty.

Watkins, M. W., Lei, P.-W. i Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, 35(1), 59 – 68. doi:10.1016/j.intell.2006.04.005

Zelazo, P. D., Craik, F. I. i Booth, L. (2004). Executive function across the life span. *Acta Psychologica*, 115(2 – 3), 167 – 183. doi:10.1016/j.actpsy.2003.12.005

Rozdział 5

Status społeczny rodziny ucznia

Kolejna grupa czynników pominiętych w modelach edukacyjnej wartości dodanej stosowanych w Polsce, a mogących rzutować na trafność interpretacji przyczynowo-skutkowej wskaźników EWD, to różne aspekty statusu ekonomiczno-społecznego rodziny ucznia. Brak uwzględnienia tej informacji przy szacowaniu EWD jest często podnoszonym argumentem krytycznym. Czy zatem gimnazja znacząco się różnią składem społecznym uczniów? Czy czynniki statusowe determinują postęp edukacyjny w trakcie nauki w gimnazjum? Jest to pytanie szczególnie ważne, bo metody szacowania wartości dodanej używane w innych krajach (np. Anglia, Australia) uwzględniają (lub uwzględniały: w Anglii zrezygnowano z wyliczania tzw. kontekstowego EWD) informację o czynnikach statusowych.

5.1. Kontekst teoretyczny i pytania badawcze

Klasyczne badanie Jamesa Colemana zapoczątkowało serię analiz nad wpływem czynników rodzinnych na osiągnięcia szkolne uczniów. Na podstawie wyników stwierdzono, że dla poziomu osiągnięć uczniów czynniki szkolne znaczą stosunkowo niewiele (Coleman, 1966). To status społeczny rodziny pochodzenia uczniów jest ważny. Osiągnięcia szkolne są znacznie silniej związane ze społecznym pochodzeniem uczniów niż z tym, do której szkoły uczęszcza. Od badań Colemana przeprowadzono tysiące analiz potwierdzających silny statusowy determinizm osiągnięć szkolnych. Aktualnie znacznie lepiej rozumiemy mechanizm transformacji statusu przypisanego ucznia w jego osiągnięcia szkolne. Przegląd badań dotyczących statusowej determinacji osiągnięć szkolnych, również dotyczących polskiej oświaty, można znaleźć w wielu polskich monografiach poświęconych różnym aspektom nierówności edukacyjnych (ostatnio m.in.: Dolata, 2008; Zawistowska, 2012; Dolata, Jakubowski i Pokropek, 2013). W tym miejscu zamieścimy tylko krótki rys ram teoretycznych analiz statusowej determinacji.

Większość XX-wiecznych teorii statusowej determinacji osiągnięć szkolnych próbowała ją wyjaśnić czynnikami środowiskowymi. W obrębie podejścia strukturalno-funkcjonalnego podstawową kategorią wyjaśniającą wpływ statusu społecznego rodziny pochodzenia ucznia na osiągnięcia była kategoria deprywacji środowiskowej (por. Bruner, 2006). Koncepcja kulturowej deprywacji głosiła, że niektóre dzieci mają niskie osiągnięcia szkolne, bo w odpowiednim wieku nie dostarczano im odpowiedniej stymulacji kulturowej. Jeżeli w rodzinie dziecku dużo się czyta, dużo z nim rozmawia, wspólnie się bawi w pewne zabawy, jeżeli na półkach znajduje się dużo książek, to wszystko to staje się wzorcem właściwej stymulacji kulturowej. Zdaniem Jerome'a

Brunera takie społecznie stronnicze rozumowanie staje się punktem wyjścia tzw. programów kompensacyjnych typu *Head Start*, które mają wyrównywać szanse edukacyjne.

Inne wyjaśnienia są formułowane w obrębie paradygmatu konfliktu strukturalnego. W każdy system społeczny wpisany jest strukturalny, czyli trudny, czy wręcz niemożliwy do rozwiązania konflikt międzygrupowy, a trwanie społeczeństwa jest możliwe tylko dzięki przemocy, która antagonistyczne grupy utrzymuje na swych społecznych pozycjach. Marksistowskie wyjaśnienie nierówności edukacyjnych każe widzieć szkołę jako instytucję klasową, czyli realizującą interes grupy dominującej. Demokratyzacja współczesnej szkoły nie pozwala uznać prostego wyjaśnienia klasowego za wystarczające. Jak to się dzieje, że grupom dominującym udaje się uniknąć formującego osobowość niewolnika wpływu powszechnej, jednolitej szkoły? Oczywiście tam, gdzie mamy do czynienia z jawnym dualizmem systemu szkolnego, można twierdzić, że powszechność i jednolitość szkoły jest fikcją. Jednak w systemach faktycznie jednolitych należy dowiedzieć, że pod powierzchnią uniwersalności tak naprawdę mają miejsce bardzo różne procesy socjalizacyjne. Pojęciem, które toruje drogę takiemu demistyfikującemu badaniu, jest ukryty program. Jean Anyon (1983) badając kilka szkół różniących się składem społecznym uczniów, doszła do wniosku, że pod fasadą jednolitych programów nauczania kryją się bardzo różne przekazy socjalizacyjne – ukryte programy. Różnice te dotyczyły nie tyle treści nauczania, co relacji społecznych w szkole i społecznego kontekstu nabywania wiedzy. Z paradygmatu konfliktu wyrastają też koncepcje odwołujące się do teorii kapitałów. Współczesne ujęcia socjologiczne wskazują na wielopostaciowość kapitałów i różnorodne mechanizmy ich konwersji. Kapitał kulturowy rodziny pochodzenia ucznia ma znaczący wpływ na jego funkcjonowanie w szkole. Myślenie o roli kapitału kulturowego w osiąganiu sukcesów szkolnych zapoczątkował Pierre Bourdieu (Bourdieu i Passeron, 1990). Jego zdaniem kapitał kulturowy człowieka jest podstawowym mechanizmem reprodukcji się podziałów społecznych. Przyszłą pozycję w strukturze społecznej wyznacza środowisko wychowawcze dziecka, a następnie uczęszczanie do szkół o określonej renomie. Uczniowie oceniani są na podstawie posiadanego kapitału kulturowego, opartego na elementach kultury wysokiej. Uczniowie z rodzin o niskim statusie społeczno-ekonomicznym uzyskują znacznie niższe wyniki, gdyż dla nich szkoła to środowisko kulturowo obce. W badaniach nad rolą kapitału kulturowego brane są pod uwagę dobra kulturowe, znajdujące się w posiadaniu rodziny, takie jak książki czy dzieła sztuki, wykształcenie i styl życia rodziców, ich podejście do edukacji dziecka lub partycypacja kulturowa samych uczniów.

Inną ważną odmianą kapitału jest kapitał społeczny. Ta forma kapitału w ujęciu Bourdieu (2001) jest pochodną sieci znajomości i wzajemnego uznania. Ważna jest nie tylko liczba i rozległość kontaktów społecznych, które umożliwiają dostęp do różnych zasobów, ale też jakość tych kontaktów, a więc pytanie o kapitał kulturowy i ekonomiczny osób, z którymi się kontaktujemy. Zdaniem Bourdieu za pomocą różnych form kapitału społecznego reprodukowane

są nierówności społeczne. Może on bowiem nie tylko spajać grupy ludzi, ale też powodować ich zamknięcie na przedstawicieli innych grup społecznych, które nie pozostają z nimi w ścisłych relacjach tworzących struktury kapitału społecznego. Zbiorowości o silnym kapitale społecznym i wysokim statusie ekonomicznym mogą dążyć do izolacji od reszty społeczeństwa. Grupy, w których więzy społeczne są bardzo silne, mają duży wpływ na indywidualne działania i wybory jej członków. Inne ujęcie kapitału społecznego znajdujemy w koncepcjach Roberta Putnama i Jamesa Colemana. Według Putnama (2008) kapitał społeczny jest równoznaczny z poziomem zaangażowania społecznego mieszkańców, ich aktywnością i członkostwem w stowarzyszeniach, świadomością obywatelską, intensywnością kontaktów w ramach lokalnej społeczności, czyli z istnieniem silnych więzi społecznych w danej zbiorowości. Zaufanie, sieci i normy wzajemności w rodzinie, szkole, grupie rówieśniczej i całej społeczności, w której wzrasta dziecko, mają wpływ na jego rozwój, możliwości i dokonywane wybory życiowe. Istotne jest poczynione przez Colemana (1988) rozróżnienie pomiędzy kapitałem społecznym, rozumianym jako zestaw norm zachowań, zaufania i sieci społecznych, a wewnątrzrodzinnym kapitałem społecznym. Na osiągnięcia szkolne dzieci mają wpływ relacje między rodzicami a dziećmi, struktura rodziny, zaangażowanie rodziców w edukację, ich aspiracje względem dzieci oraz stopień kontroli, jaką nad nimi sprawują.

Klasyczne dla podejścia teorii konfliktu jest wyjaśnianie determinacji statusowej przez kapitał ekonomiczny. Czynniki ekonomiczne, takie jak sytuacja materialna gospodarstwa domowego, w którym wychowuje się uczeń, mogą mieć wpływ na osiągnięcia przez niego wyniki w nauce. Ważny jest nie tyle dochód przypadający na całą rodzinę, co zasoby gospodarstwa domowego w przeliczeniu na osobę. Wsparciem dla dzieci mogą być płatne zajęcia pozaszkolne lub kursy przygotowujące do egzaminu, na opłacenie których pozwolili sobie mogą zamożniejsi rodzice.

Podejście teorii konfliktu wypromowało bardzo ważne pojęcie wyjaśniające statusową determinację osiągnięć szkolnych – autoselekcja. Decyzję o wyborze szkoły w krajach, gdzie rozluźniono lub zrezygnowano z rejonizacji, podejmują rodzice. Mają oni często różną perspektywę poznawczą, różny stosunek do edukacji, dysponują różnymi zasobami. Rodzice, którzy dostrzegają korzyści związane z edukacją oraz jej wpływ na pozycję dziecka w przyszłości, świadomi są zróżnicowania szkół oraz konsekwencji takiej decyzji i pragną aktywnie dokonać wyboru placówki edukacyjnej, do której uczęszczać będzie ich dziecko. Zwolennicy programów rodzicielskiego wyboru twierdzą, że gdy stworzymy rodzicom zewnętrzne warunki wyboru szkoły dla swoich dzieci, to będą oni wybierali racjonalnie, myśląc o osiąganiu ważnych dla siebie celów i maksymalizując edukacyjną użyteczność. Stephen Ball (1993) twierdzi, że teoretycy rynku zakładają, że kompetencje i środki decyzyjne są równo rozłożone w populacji. Podobnie jak w klasycznej ekonomii przyjmują założenie, że wybory konsumentów na rynku są racjonalne, a tym samym na rynku utrzymują się produkty

o jakości optymalnej ze względu na cenę. Ball twierdzi, że założenie to w odniesieniu do rynku edukacyjnego jest w sposób oczywisty nieprawdziwe. Takie środki jak czas, transport, dostęp do instytucji wspierających opiekę nad dzieckiem nie są równo rozłożone. Można do tej listy barier dodać fakt, że rozpoznawanie jakości usług na rynku edukacyjnym nie jest proste i wymaga znacznych kompetencji kulturowych i dostępu do danych. Zdaniem Balla polityka rynkowa w oświacie publicznej działa jak mechanizm reprodukcji klasowej poprzez dublowanie upośledzenia „źle wybierających” w wyniku połączenia dystrybucji wyborów z dystrybucją środków. Program wyboru szkoły staje się formą liberalnej legitymizacji dla niesprawiedliwej dystrybucji publicznych środków na bezpłatną oświatę. „Trafne wybory” rodziców kumulują się w lepszych szkołach, „nietrafne wybory” (brak wyboru to też wybór) skupiają się w pozostałych. To w oczywisty sposób pogłębia segregację społeczną i dodatkowo sprawia, że w lepszych szkołach jest więcej uczniów, a tym samym szkoły te mają więcej środków i zdobywają swoisty bonus finansowy wynikający z mniejszych kosztów utrzymania szkół działających na granicy możliwości przyjmowania uczniów. W ten sposób beneficjentami bezpłatnej oświaty publicznej stają się grupy społeczne dysponujące kapitałem kulturowym umożliwiającym trafny wybór.

Jeszcze inne kategorie wyjaśniające wprowadzają przedstawiciele podejścia interpretacyjnego. Podejście to nie jest jednolitą teorią, ale zróżnicowanym zbiorem koncepcji wywodzących się głównie z interakcjonizmu symbolicznego George’a Meada. Makrostrukturalne podejście funkcjonalistów czy teoretyków konfliktu wychodziło od pytania, w jaki sposób społeczeństwo kształtuje jednostki i reguluje procesy społeczne. Podejście interpretacyjne zmienia perspektywę badawczą, wychodząc od pytania, jak jednostki tworzą, utrzymują i zmieniają społeczeństwo. Klasycznym przykładem badań o orientacji interpretacyjnej jest studium Paula Willis’a *Learninig: How working class kids get working class jobs* (1981). Odnajdujemy w nim wszystkie charakterystyczne elementy tego podejścia: etnograficzne badanie „obcej” kultury, otwarcie „czarnej skrzynki” procesów szkolnych, by opisać procesy sensotwórcze składające się na edukację, uwzględnienie podmiotowości jednostki i opisywanie nierówności w kategoriach tożsamości tworzonej przez samą jednostkę. Kultura „równych gości” wytworzona przez grupę 12 uczniów nieselekcyjnej szkoły średniej z rodzin robotniczych, choć była negatywistyczną reakcją na wymagania szkoły, stała się równocześnie źródłem tożsamości prowadzących tych młodych ludzi do pozycji społecznych zajmowanych przez ich rodziców. Opozycyjna wobec kultury „frajerów” (mułów), pozwoliła młodym ludziom nadać sens ich szkolnemu losowi. Nie mogło się to stać bez przewartościowania szkolnego sukcesu.

Ważnym uzupełnieniem środowiskowych teorii statusowej determinacji osiągnięć szkolnych jest podejście genetyki zachowania. Przede wszystkim zwolennicy podejścia biologicznego do problematyki nierówności edukacyjnych zwracają uwagę na to, że działanie czynników

społecznych należy odróżnić od efektu przekazu genetycznego. Wiele indywidualnych dyspozycji ucznia wyznaczających poziom osiągnięć szkolnych jest w znaczącym stopniu determinowanych genetycznie, co należy uwzględnić w wyjaśnianiu nierówności społecznych (Jensen, 1969; 1991, Hernstein i Murray, 1994, Plomin i in. 2001). Podsumowując, z perspektywy pedagogicznej wyniki badań nad odziedziczalnością Judith Harris (2000) i Steven Pinker (2005) dochodzą do dwóch kluczowych wniosków. Po pierwsze, wszystkie kluczowe cechy osobowości – w tym inteligencja ogólna – są w znaczącym stopniu determinowane genetycznie. Po drugie, wpływ wychowywania się w tej samej rodzinie na kształtowanie cech osobowości – przy kontroli czynnika przekazu genetycznego – jest słaby lub nie zachodzi. Nawiązując do klasycznych badań Colemana nad czynnikami wyjaśniającymi zróżnicowanie osiągnięć szkolnych, Hans Eysenck i Michael Eysenck (1998, s. 226) piszą: „Ignorując oddziaływanie [czynnika genetycznego] Coleman w rzeczywistości dopuścił do tego, by w jego badaniach dziedziczenie oddziaływało na osiągnięcia szkolne w postaci takich czynników, jak status społeczno-ekonomiczny i środowisko rodzinne ucznia, doprowadziło do sztucznego rozdęcia wpływów środowiska rodzinnego ponad wszelkie rozsądne proporcje. Jeżeli powtórzyć analizę jego wyników z należyтым uwzględnieniem oddziaływania czynników dziedzicznych, treść wyciągniętych wniosków ulega radykalnej zmianie i okazuje się, że jakość szkoły w bardzo poważnym stopniu wpływa na poziom edukacyjnych osiągnięć uczniów”. Najnowsze badania potwierdzają, że przekaz genetyczny jest ważnym kanałem wpływu. Z tego punktu widzenia bardzo ciekawe są najnowsze ustalenia dotyczące wzrostu siły odziedziczalności IQ w trakcie rozwoju osobniczego (Haworth i in., 2010), wykrycie, że czynnik psychologiczny zwany spostrzeganiem własnych zdolności też w dużej mierze jest odziedziczalny i – niezależnie od IQ – bezpośrednio oddziałuje na osiągnięcia szkolne (Grevén i in., 2009), czy pierwsze dowody interakcji środowisko – genotyp (Molenaar i in., 2013). Dla badania trafności metody EWD niezwykle ważne są ustalenia dotyczące odziedziczalności indywidualnych przyrostów osiągnięć szkolnych. Robert Plomin i współpracownicy (Haworth, Asbury, Dale i Plomin, 2011) wykazali na próbie 4000 par dwunastoletnich bliźniąt, że wskaźnik odziedziczalności dla wyników testów osiągnięć wynosi około 50%, zaś wspólne środowisko (wychowywanie się w jednej rodzinie) wyjaśnia około 25% wariancji osiągnięć. Ale co ciekawsze badacze oszacowali również wskaźnik odziedziczalności dla dwuletnich przyrostów osiągnięć (od 10. do 12. roku życia). Okazało się, że również miara postępu w osiągnięciach jest w 50-procentach pod kontrolą genów, a tylko 12-procentach zmienności przyrostów indywidualnych wyjaśniało uczęszczanie do danej szkoły. Choć autorzy deklarują, że wynik jest niespodziewany, to jest on spójny z oszacowaniami znaczenia szkoły dla osiągnięć szkolnych w ramach tradycji badań nad *efektywnością kształcenia*.

Powyższy przegląd teorii statusowej determinacji osiągnięć szkolnych pokazuje, jak różne mogą być jej mechanizmy. Dostarcza też przesłanek do wyboru wskaźników statusu, które

będą mierzone w tym badaniu. Jeżeli chcemy w modelu dobrze opisać procesy selekcji na progu gimnazjum, to potrzebujemy przede wszystkim dobrych miar kapitału kulturowego i ekonomicznego rodziny ucznia i syntetycznych wskaźników pokazujących na pozycję rodziny na różnych wymiarach statusu społeczno-ekonomicznego. Wyniki badań z zakresu genetyki zachowania uczulają nas natomiast na problem interpretacji korelacji miar SES z osiągnięciami. Prawdopodobnie za znaczną część tej korelacji odpowiada mechanizm dziedziczenia biologicznego.

Celem analiz w niniejszym rozdziale nie jest potwierdzenie po raz kolejny statusowej determinacji osiągnięć. W tym rozdziale skupimy się na pytaniach kluczowych z punktu widzenia oceny trafności interpretacji przyczynowej wskaźników EWD. Na ile brak w modelu czynników statusowych obciąża szacunki EWD? Odpowiedź zależy od czterech kwestii:

- Jak silne jest międzyszkolne i międzyoddziałowe zróżnicowanie ze względu na czynniki statusowe?
- Jak silny jest efekt czynników statusowych na wyniki egzaminów gimnazjalnych? Który z aspektów statusu rodziny ucznia jest z tego punktu widzenia najważniejszy?
- Czy efekt ten zanika w modelu EWD, a jeżeli nie, to jak jest silny?
- Czy wskaźniki EWD szacowane za pomocą modeli bez uwzględnienia czynników statusowych i z uwzględnieniem znacząco się różnią?

5.2. Metoda

5.2.1. Pomiar czynników statusowych

W analizach wykorzystano kilka wskaźników statusu rodziny ucznia: indeks statusu ekonomiczno-społecznego (ISEI), indeks prestiżu zawodu wykonywanego przez rodziców (SIOPS), poziom wykształcenia rodziców oraz wskaźnik zasobności (głównie dobra kulturowe) rodziny ucznia. Miary te, choć nie są w pełni niezależne, pozwolą wszechstronnie opisać status społeczny rodziny pochodzenia ucznia. ISEI jest zbudowany na statusie zawodowym i jest najbardziej ogólny, pozostałe miary wskaźnikują węższe konstrukty. Są to miary najczęściej używane w badaniach edukacyjnych o dobrze udokumentowanej mocy predykcyjnej.

5.2.1.1. Międzynarodowa skala statusu społeczno-ekonomicznego ISEI i międzynarodowa skala prestiżu zawodów SIOPS.

W badaniach socjologicznych procesów ruchliwości zwykle stosuje się miary SES, które wskazują na pozycję jednostki w strukturze społecznej na podstawie wykonywanego zawodu. Zawód traktuje się jako swoisty łącznik między nakładami koniecznymi, by objąć daną pozycję społeczną, a nagrodami otrzymywanymi za jej pełnienie. Takie widzenie zawodu

wywodzi się z funkcjonalnej teorii uwarstwienia Kingsleya Davisa i Wilberta Moore'a. Zgodnie z twierdzeniami tej teorii zawód pozwala na wnioskowanie ze znacznym prawdopodobieństwem o innych cechach pozycji społecznej jednostki. Zawody rodziców uczniów zostały zakodowane zgodnie z Międzynarodową Klasyfikacją Zawodów ISCO-08 (International Labour Organization, 2007). Klasyfikacja ta jest powszechnie przyjętą podstawą do prowadzenia porównań międzynarodowych w zakresie rynków pracy. Została ona przyjęta przez Międzynarodową Organizację Pracy (ILO) w grudniu 2007 roku. Klasyfikacja ta wyróżnia 570 zawodów połączonych w 120 mniejszych grup, które z kolei łączą się w 34 grupy średnie, te zaś w 10 grup dużych. Szczegóły klasyfikacji wraz z definicjami można odnaleźć na stronach Międzynarodowej Organizacji Pracy. Dobry przewodnik do kodowania poszczególnych zawodów wraz z wyjaśnieniami znajduje się na stronie Harry'ego Ganzeboom¹.

Na podstawie ISCO-08 oraz odpowiednich kluczy przejścia, każdemu z zawodów przyporządkowane zostały dwie miary społecznej klasyfikacji zawodów. Pierwsza z nich to międzynarodowa skala statusu społeczno-ekonomicznego – *International Socio-Economic Index* – ISEI (Ganzeboom, De Graaf i Treiman, 1992), druga to międzynarodowa skala prestiżu zawodów (*Standard International Occupational Prestige Scale* – SIOPS (Treiman, 1977).

Harry Ganzeboom, Paul De Graaf i Donald Treiman opracowali międzynarodową skalę statusu społeczno-ekonomicznego. W odróżnieniu od skali prestiżu nie jest to skala subiektywna, lecz została skonstruowana jako ważony średni uzyskany poziom edukacji oraz ważone średnie dochody przedstawicieli poszczególnych zawodów, niekiedy korygowane o wpływ wieku (Ganzeboom i in., 1992). Estymacja wartości ISEI zmierza do takiego ich oszacowania, aby maksymalizować efekt pośredni wykształcenia na dochody poprzez wykonywany zawód, przy jednoczesnej minimalizacji efektu bezpośredniego wykształcenia na dochód. W odróżnieniu od ISEI skonstruowanych dla ISCO-68 (Ganzeboom i in., 1992) jak i ISCO-88 (Ganzeboom i Treiman, 1996; Ganzeboom i Treiman, 2003) w najnowszej wersji dla ISCO-08 oszacowania zostały dokonane zarówno na danych dla kobiet i mężczyzn. Informacje pochodziły z 198 500 kwestionariuszy zebranych w ramach Międzynarodowego Programu Badań Społecznych (*International Social Survey Programme* (ISSP)) realizowanego w latach 2002–2007 (Ganzeboom, 2010).

Prowadząc analizy porównawcze 85 skal prestiżu zawodów opracowanych w ponad 60 różnych społeczeństwach i w różnych okresach Treiman zauważył powszechny, w dużej mierze stały w czasie układ prestiżu zawodów, który nazwany został „stałą Treimana”. Na tej podstawie opracował międzynarodową skalę prestiżu zawodów,

¹ (home.fsw.vu.nl/hbg.ganzeboom/isco08/)

zaktualizowaną razem z Ganzeboomem w kolejnych latach (Ganzeboom i Treiman, 1996; Ganzeboom i Treiman, 2003). Jest to skala z natury subiektywna, gdyż opiera się na ocenach prestiżu poszczególnych zawodów dokonywanych w ramach dużych badań przekrojowych przez respondentów lub przez sędziów kompetentnych (Ganzeboom i in., 1992). Po ogłoszeniu ISCO-08 przygotowane zostały również odpowiednie przejścia pomiędzy nową klasyfikacją zawodów, a SIOPS. Algorytmy dostępne są na stronie internetowej Ganzebooma.

Prestiżu mierzonego skalą SIOPS i międzynarodowego statusu społeczno-ekonomicznego ISEI nie należy traktować jako wymiennych miar statusu zawodowego, tak na poziomie koncepcyjnym, jak i empirycznym. Prestiż jest raczej miarą uznania społecznego (akceptacji), zaś ISEI zasobów i bardziej bezpośrednich nagród ekonomicznych (Ganzeboom i Treiman, 2003).

W analizach używamy wartości wskaźników ISEI i SIOPS tego z rodziców, dla którego przypisana jest wyższa wartość. W wypadku braku danych dla jednego z rodziców, przypisywano wartość wskaźnika drugiego. Zmienne te będziemy skrótowo nazywać HISEI i HSIOPS.

5.2.1.2. Wykształcenie rodziców

Wyróżnionym w ankiecie skierowanej do rodziców poziomom wykształcenia przypisano wartości skalowe w latach nauki. Pewne decyzje podjęto arbitralnie. Na przykład przypisanie niepełnemu podstawowemu sześciu lat nauki jest uproszczeniem, choć faktycznie przez dłuższy czas utrzymywała się w polskiej szkole praktyka, że dopiero po ukończeniu szóstej klasy można było kierować mocno „przerośniętych” uczniów do przysposobienia zawodowego. Bardziej arbitralne jest przypisanie do kategorii „stopień doktora, doktora habilitowanego lub profesora” wartości skalowej 20. Sposób wyskalowania wszystkich kategorii wykształcenia w latach nauki pokazuje tabela 1.

Tabela 1. Przeliczenie poziomu wykształcenia rodziców na lata nauki

poziom wykształcenia	liczba lat nauki	procent wśród ojców	procent wśród matek
nieukończone podstawowe	6	0,2	0,2
podstawowe	8	8,5	8,6
zasadnicze zawodowe	10	49,0	34,5
średnie: technikum, liceum zawodowe	13	22,6	23,1
średnie: liceum ogólnokształcące	12	3,5	7,4
policealne lub pomaturalne	13	2,4	6,8
licencjackie lub inżynierskie	15	2,8	4,0
wyższe studia magisterskie lub lekarskie	17	9,3	14,5
stopień doktora, doktora habilitowanego lub profesora	20	0,5	0,4
nie wiem, brak danych	-	4,1	1,9

Korelacja między wyskalowanym poziomem wykształcenia matki i ojca wynosi 0,6. Wskaźnik używany w analizach to poziom wykształcenia dla tego rodzica, dla którego przebiera on wyższą wartość. W wypadku braku danych dla jednego z rodziców, przypisywano wartość wykształcenia drugiego. Zmienną tę będziemy skrótowo nazywać HEDU.

5.2.1.3. Zasobność rodziny ucznia

Do analizy poziomu zamożności gospodarstwa domowego ucznia zaadaptowana została pula pytań dotycząca wyposażenia gospodarstwa domowego wykorzystywana w badaniach PISA. W wersji użytej w badaniu składa się ona z 17 dychotomicznych pytań, w których respondenci pytani byli o posiadanie następujących elementów wyposażenia: biurka do nauki dla dziecka, własnego pokoju dla dziecka, spokojnego miejsca do nauki dla dziecka, komputera, który dziecko może używać do nauki, edukacyjnych programów komputerowych, dostępu do internetu, kalkulatora, literatury pięknej (np. Sienkiewicz); zbiorów poezji, dzieł sztuki (np. obrazy), encyklopedii, słowników, zmywarki do naczyń, DVD albo wideo, anteny satelitarnej lub telewizji kablowej, aparatu cyfrowego, mikroskopu lub lunety. Do konstrukcji skali użyto też niezwykle predykcyjnego pytania o liczbę książek w domu.

Ze względu na problemy z estymacją modelu pomiarowego, tak poprzez klasyczną konfirmacyjną analizę czynnikową (CFA), jak i z wykorzystaniem eksploracyjnych modeli równań strukturalnych (ESEM), z oryginalnej skali usunięta została pozycja „dostęp do komputera, który dziecko może używać do nauki”.

Estymacja modelu odzwierciedlającego podejście wykorzystywane w analizach PISA (OECD, 2009) z wykorzystaniem CFA napotkała istotne problemy. W związku z tym zastosowano inne podejście. Estymacja modelu *bifactor* w ramach podejścia ESEM pozwoliła na uzyskanie modelu dobrze dopasowanego do danych: RMSEA 0,021, CFI – 0.989, TLI – 0.980, WRMR – 1,06. Poza czynnikiem głównym mierzącym zasobność gospodarstwa domowego rodziny ucznia, wyróżnione zostały trzy czynniki lokalne: miejsce do nauki, sprzęt elektroniczny oraz dobra kulturowe. W analizach, jako zmienną opisującą aspekt posiadania statusu rodziny ucznia, użyto miary ogólnej zasobności gospodarstwa domowego. Z analizy czynników lokalnych zrezygnowano, ponieważ ich uwzględnienie nie wносиło nic ważnego do analizy trafności metody EWD.

Poniższa tabela zawiera wartości podstawowych parametrów rozkładu używanych w analizach wskaźników statusu.

Tabela 2. Parametry rozkładu zmiennych statusowych. Dane ważone

zmienne	średnia	odchylenie standardowe	skośność
wykształcenie (HEDU)	12,78	2,65	0,42
status (HISEI)	43,06	20,85	0,56
prestż (HSIOPS)	45,10	10,77	0,64
zasobność	-0,02	1,00	-0,23

5.2.2. Schemat analizy danych

Analogenicznie jak w pozostałych rozdziałach, w celu odpowiedzi na pytania badawcze użyto hierarchicznych modeli regresji. Zastosowano modele trzypoziomowe z losową stałą uwzględniające poziom ucznia, oddziału klasowego i szkoły. Analizy przeprowadzono oddzielnie dla wyniku egzaminu gimnazjalnego w części matematyczno-przyrodniczej i humanistycznej. Przypomnijmy – wyniki egzaminu są przedstawiane na skali 100:15 w populacji uczniów, którzy przystąpili do egzaminu gimnazjalnego w 2012 roku. Analizy mające na celu weryfikację znaczenia czynników statusowych przeprowadzono czteroetapowo. Najpierw sprawdzono, na ile szkoły i oddziały klasowe różnią się ze względu na czynniki statusowe. W drugim kroku pokazano związek poszczególnych czynników statusowych z wynikami egzaminu. W trzecim kroku szacowano siłę tych samych efektów w modelu EWD. Porównanie wyników z drugiego i trzeciego etapu pozwoliło sprawdzić, w jakim stopniu w modelach EWD kontrolujemy niezależne od szkoły czynniki statusowe. Innymi słowy sprawdzono, jak zmienia się związek danej zmiennej SES z wynikami egzaminów, jeśli kontrolujemy w modelu te zmienne, które są uwzględniane w modelach EWD. W końcu sprawdzono dla potencjalnie użytecznego w modelowaniu EWD czynnika statusowego – wykształcenia rodziców – na ile modele EWD pozwalają kontrolować wariancję międzyszkolną i międzyoddziałową i czy dodanie do modelu informacji o tym czynniku znacząco zmieniłoby oszacowania efektywności nauczania dla szkół.

Analizy wykonano za pomocą programu HLM 6.06, wykorzystując metodę estymacji *full maximum likelihood*. Do oszacowania istotności efektów stałych wykorzystano odporne (*robust*) błędy standardowe. Wartości istotne na poziomie istotności $p < 0,05$ zaznaczono w tabelach pogrubionym drukiem. Analizy wykonano z uwzględnieniem wag warunkowych (patrz rozdział metodologiczny) na wszystkich trzech poziomach, odpowiednio przekalowanych za pomocą programu HLM.

5.3. Wyniki

Zgodnie z przyjętą strategią analizy, zaczynamy od odpowiedzi na pytanie, czy szkoły różnią się ze względu na czynniki statusu społecznego rodziny ucznia.

5.3.1. Międzyszkolna i międzyoddziałowa wariancja wskaźników statusowych

Analizę międzyszkolnego i międzyoddziałowego zróżnicowania wyników egzaminacyjnych przeprowadzono, wykorzystując trzypoziomowe modele puste. Oszacowanie efektów losowych na poziomie szkół, oddziałów klasowych i uczniów pozwoliło wyliczyć współczynniki korelacji wewnątrzgrupowej. Ich wartość mówi, jaką część zmienności danego czynnika statusowego można przypisać podziałowi uczniów na szkoły i oddziały.

Tabela 3. Międzyszkolne i międzyoddziałowe zróżnicowanie wskaźników statusowych rodziny ucznia: wykształcenie, HISEI, HSIOPS i zasobność. Wyniki trzypoziomowych analiz regresji, modele puste

zmienna zależna:	wykształcenie HEDU	status HISEI	prestż HSIOPS	zasobność rodziny
oszacowanie efektów losowych				
wariancja efektów szkół	1,08	78,78	9,81	0,113
wariancja efektów oddziałów klasowych	0,29	19,34	4,65	0,027
wariancja na poziomie ucznia	5,12	308,70	89,26	0,866
współczynniki korelacji wewnątrzgrupowej				
poziom szkół	0,166	0,194	0,095	0,131
poziom klas	0,045	0,048	0,045	0,026
poziom klas i szkół łącznie	0,211	0,241	0,140	0,157
liczebności				
liczba uczniów	4690	4782	4782	4901
liczba klas	291	291	291	291
liczba szkół	150	150	150	150

Zgodnie ze stanem naszej wiedzy (Dolata, 2008) przeprowadzona analiza wykazała istotne międzyszkolne i międzyoddziałowe zróżnicowanie ze względu na czynniki statusowe. Podział uczniów na szkoły wyjaśnia około 17% wariancji wskaźnika wykształcenia i około 19% zmienności miary statusu społeczno-ekonomicznego rodziców ucznia. W tym zakresie podział na oddziały w obrębie szkoły wyjaśnia średnio około 5% zmienności. Trochę słabsze jest zróżnicowanie szkół ze względu na miarę prestiżu zawodu rodziców i wskaźnik zasobności rodziny, ale

i tak podział na szkoły wyjaśnia 10–13% wariacji. Zróźnicowanie międzyoddziałowe tych miar jest – analogicznie jak dwóch poprzednich – niższe niż międzyszkolne i wynosi około 3–5%. Ciekawe, że w wypadku inteligencji znacznie bardziej były zróźnicowane oddziały wewnątrz szkół (prawie 9% wariacji wyjaśnianej), niż szkoły (5,5% wariacji daje się wyjaśnić przez podział na szkoły). To ciekawy wątek, ale ponieważ nie wiąże się bezpośrednio z tematem raportu, porzucimy go.

Wyniki analizy potwierdziły, że szkoły bardzo różnią się zasobami statusowymi. To stwarza potencjalne zagrożenie dla modelowania EWD, w którym nie uwzględnia się zmiennych SES. Zasadne jest zatem wykonanie kolejnych kroków zaplanowanej strategii analizy.

Zanim przejdziemy do kolejnych obliczeń warto sformułować dwie uwagi. Po pierwsze, okazuje się, że prosta do zebrania informacja o wykształceniu rodziców, wydaje się dobrze opisywać międzyszkolne zróźnicowanie zasobów statusowych. Gdyby okazało się, że uwzględnienie tego zróźnicowania jest niezbędne, by ograniczyć stroniczość modeli EWD, możliwe jest zbieranie takich danych w systemie. Po drugie, należy zwrócić uwagę, że proste porównywanie siły międzyszkolnego zróźnicowania ze względu na poszczególne aspekty statusu rodziny ucznia jest metodologicznie ryzykowne. Użyte miary mają prawdopodobnie różną rzetelność, w związku z tym zaobserwowane różnice mogą odbijać po części realne różnice w sile międzyszkolnego zróźnicowania, po części niższą lub wyższą rzetelność zastosowanej miary.

5.3.2. Status społeczny a wyniki egzaminu gimnazjalnego w części matematyczno-przyrodniczej

W tym podrozdziale przeprowadzone zostaną dwa kolejne etapy analizy trafności metody EWD dla wyników egzaminu matematyczno-przyrodniczego. Najpierw zobaczymy, jak poszczególne czynniki statusowe są powiązane z wynikami egzaminacyjnymi bez kontroli uprzednich osiągnięć. Wyniki odpowiednich analiz zawiera tabela 4.

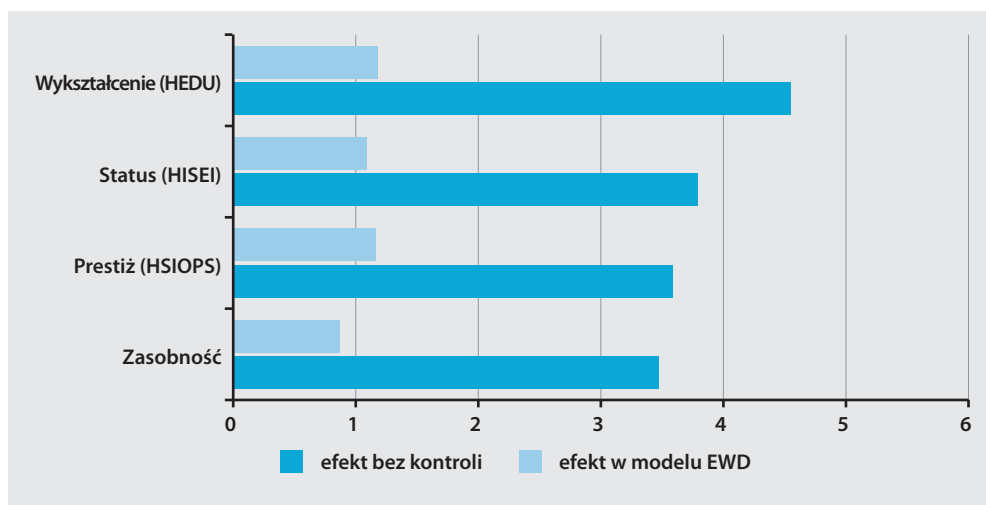
Dla wszystkich czterech aspektów statusu rodziny ucznia uzyskano istotne statystycznie efekty. Współczynniki regresji w tabeli mówią, jak zmienia się wynik egzaminu gimnazjalnego (skala 100;15), gdy dana miara statusowa zmienia się o jedno odchylenie standardowe. I tak zmiana wykształcenia o jedno odchylenie oznacza wzrost wyników o około 4,6 punktu egzaminacyjnego, czyli prawie o jedną trzecią odchylenia standardowego. Pozostałe trzy efekty są trochę słabsze i mieszczą się w przedziale 3,5–3,8 punktu egzaminacyjnego. Choć nie leży to w obszarze naszych analiz, zwróćmy uwagę, że efekty statusowe są słabsze niż ten obserwowany w wypadku IQ. Dla inteligencji dla tej części egzaminu zanotowano efekt 8,97, czyli prawie dwukrotnie silniejszy niż dla wykształcenia rodziców.

Tabela 4. Czynniki statusowe a wyniki z egzaminu w części matematyczno-przyrodniczej, model bez kontroli uprzednich osiągnięć. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych, współczynniki regresji dla standaryzowanych (0;1) zmiennych wyjaśniających, zmienna zależna standaryzowana (100;15)

zamienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej								
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)
oszacowanie efektów stałych								
poziom ucznia								
wykształcenie (HEDU)	4,55	(0,29)	–	–	–	–	–	–
status (HISEI)	–	–	3,79	(0,24)	–	–	–	–
prestż (HSIOPS)	–	–	–	–	3,59	(0,23)	–	–
zasobność	–	–	–	–	–	–	3,47	(0,21)
stała	100,83	(0,46)	100,67	(0,46)	100,44	(0,45)	100,32	(0,47)
oszacowanie efektów losowych								
wariancja efektów szkół	13,34		13,53		13,70		16,06	
wariancja efektów oddziałów	14,80		17,24		17,30		17,13	
wariancja na poziomie ucznia	147,75		151,80		152,54		150,47	
podsumowanie								
deviance	34748,12		35519,09		35541,29		36361,54	
liczba szacowanych parametrów	5		5		5		5	
liczba uczniów	4690		4782		4782		4901	
liczba oddziałów	291		291		291		291	
liczba szkół	150		150		150		150	

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe

Rysunek 1. Siła efektów czynników statusowych w modelu bez kontroli uprzednich osiągnięć i w modelu EWD. Egzamin gimnazjalny w części matematyczno-przyrodniczej



Na ile skutecznie wyeliminuje efekt czynników statusowych wykorzystanie modelu EWD? Czy uwzględnienie uprzednich osiągnięć i pozostałych zmiennych kontrolnych – płci ucznia i dysleksji – pozwoli wyeliminować, bądź przynajmniej silnie ograniczyć, efekty SES? Spójrzmy na wyniki.

W modelach EWD wszystkie czynniki statusowe są w dalszym ciągu istotnymi predyktorami wyników egzaminacyjnych w części matematyczno-przyrodniczej. Oznacza to, że czynniki te wpływają nie tylko na poziom osiągnięć szkolnych na progu gimnazjum, ale również determinują postęp ucznia podczas trzech lat nauki w gimnazjum. Jednak siła efektów jest znacznie słabsza niż w modelu bez kontroli uprzednich osiągnięć. Lepiej pozwala to prześledzić poniższy wykres.

Tabela 5. Czynniki statusowe a wyniki z egzaminu w części matematyczno-przyrodniczej, model EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych, współczynniki regresji dla standaryzowanych (0;1) zmiennych wyjaśniających, zmienna zależna standaryzowana (100;15)

zamienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej								
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)
oszacowanie efektów stałych								
poziom ucznia								
sprawdzian	11,60	(0,27)	11,70	(0,27)	11,69	(0,28)	11,71	(0,27)
sprawdzian ²	1,24	(0,12)	1,25	(0,11)	1,24	(0,11)	1,26	(0,11)
sprawdzian ³	-0,39	(0,06)	-0,40	(0,07)	-0,39	(0,07)	-0,39	(0,06)
płeć ^a	-0,76	(0,28)	-1,02	(0,28)	-1,04	(0,27)	-0,98	(0,27)
dysleksja na egzaminie gim.	1,10	(0,73)	1,07	(0,73)	1,02	(0,72)	1,08	(0,70)
dysleksja na sprawdzianie	-3,06	(0,72)	-3,05	(0,72)	-3,00	(0,71)	-3,00	(0,69)
wykształcenie (HEDU)	1,18	(0,18)	–	–	–	–	–	–
status (HISEI)	–	–	1,09	(0,17)	–	–	–	–
prestż (HSIOPS)	–	–	–	–	1,16	(0,18)	–	–
zasobność	–	–	–	–	–	–	0,87	(0,14)
Stała	10,94	(0,301)	102,51	(0,320)	102,49	(0,315)	102,39	(0,319)
oszacowanie efektów losowych								
wariancja efektów szkół	4,39		4,25		4,04		4,39	
wariancja efektów oddziałów	4,97		5,04		4,80		5,05	
wariancja na poziomie ucznia	65,52		65,90		65,90		65,62	
podsumowanie								
deviance	31124,48		31718,78		31709,29		32478,62	
liczba szacowanych parametrów	11		11		11		11	
liczba uczniów	4690		4782		4782		4901	
liczba oddziałów	291		291		291		291	
liczba szkół	150		150		150		150	

a – efekt dla dziewcząt

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe

Siła efektów w modelach EWD spadła 3–4 krotnie do poziomu między 0,9 a 1,2 punktu egzaminacyjnego. Czy uzyskane wyniki powinny zatem skłonić nas do uznania, że miary EWD są stronnicze i zarekomendować włączenie zmiennych statusowych do modeli? Zanim taki sąd sformułujemy, zobaczymy, jak skutecznie model EWD radzi sobie z problemem na poziomie szkoły. Zajmiemy się tym w kolejnym podrozdziale.

5.3.3. Znaczenie czynników statusowych dla szacowania EWD w obszarze matematyczno-przyrodniczym

Należy pamiętać, że analizowana metoda nie służy do szacowania indywidualnych przyrostów, a efektów losowych dla grup uczniów (przede wszystkim szkół). W związku z tym warto przyjrzeć się, jak skutecznie modele EWD wyjaśniają wariancję na trzech poziomach analizy. Tabela 6 zawiera wyniki dekompozycji wariancji wyników egzaminacyjnych w obszarze matematyczno-przyrodniczym dla modelu pustego, modelu EWD i modelu EWD z kontrolą wykształcenia rodziców. Wybrano wykształcenie, ponieważ jest to silny predyktor wyników na egzaminie gimnazjalnym i potencjalnie „zbieralna” dana w systemie oświatowym.

Tabela 6. Porównanie wariancji międzyszkolnej, międzyoddziałowej i indywidualnej w trzech modelach analizy. Egzamin w części matematyczno-przyrodniczej

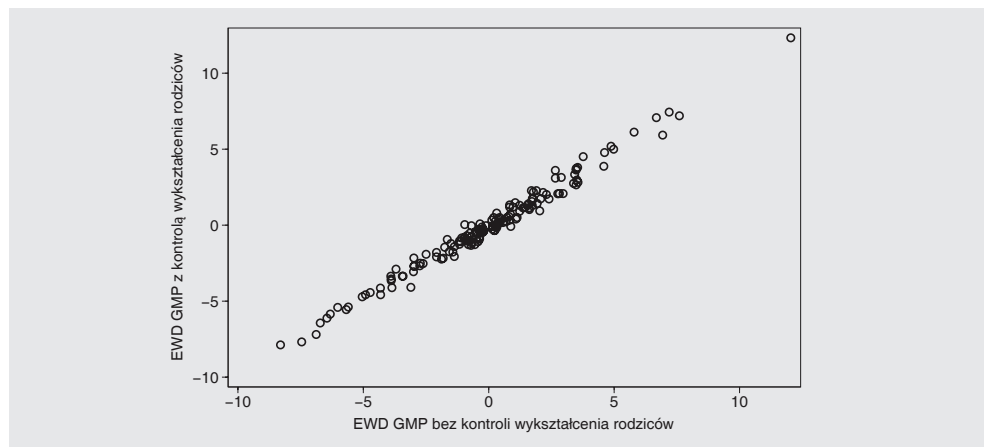
wariancja	model pusty	model EWD	model EWD z wykształceniem rodziców
międzyszkolna	20,21	4,26	4,39
międzyoddziałowa	22,77	5,57	4,97
indywidualna	158,92	66,66	65,52

Punktem wyjścia analizy są wyniki z modelu pustego. Wariancja międzyszkolna wynosi około 20 punktów egzaminacyjnych, międzyoddziałowa wynosi w przybliżeniu 23, na poziomie indywidualnym równa jest około 159. W modelu EWD wariancja międzyszkolna to 21% wariancji z modelu pustego. Dodanie do zestawu zmiennych wyjaśniających wykształcenia rodziców niewiele zmienia – wariancja międzyszkolna w tym przypadku nawet lekko rośnie, do 22% wariancji w modelu pustym. Model EWD dość skutecznie wyjaśnia też wariancję międzyoddziałową – uwzględnienie zmiennej wykształcenia rodziców sprawia, że ta składowa wariancji to 24% zmienności w modelu pustym. Dodanie do modelu zmiennej wykształcenia nieznacznie zmniejsza tę składową wariancji – 22% wariancji z modelu pustego. Ważne dla analizy trafności metody EWD jest to, że kontrola wariancji międzyszkolnej i międzyoddziałowej w modelu EWD jest znacznie skuteczniejsza niż kontrola wariancji indywidualnej. Ta ostatnia ulega redukcji jedynie do poziomu 42% oszacowania z modelu pustego.

Przedstawione wyniki wskazują, że modele EWD lepiej sobie radzą z kontrolą efektów statusowych w wypadku wyższych poziomów analizy – szkoły i oddziału. To oznacza, że dodanie do modelu zmiennej wykształcenia raczej nie wpłynie silnie na oszacowania EWD szkoły lub oddziału. Przetestujmy to empirycznie dla poziomu szkoły w naszej próbie 150 gimnazjów.

Na podstawie wyników egzaminu gimnazjalnego z 2012 roku w części matematyczno-przyrodniczej oszacowano dla każdej szkoły dwa wskaźniki edukacyjnej wartości dodanej: bez kontroli wykształcenia (EWD) i z kontrolą zmiennej wykształcenia rodziców (EWD_{hedu}). Do szacowania użyto trzypoziomowego modelu efektów mieszanych, wskaźniki są Bayesowskimi predykcjami a posteriori. Poniższy wykres rozrzutu pozwala na porównanie tych oszacowań.

Rysunek 2. Porównanie wskaźników EWD dla części matematyczno-przyrodniczej liczonych bez i z kontrolą wykształcenia rodziców



Na osi poziomej znajdują się wartości EWD dla szkół biorących udział w badaniu wyliczone z modelu bez kontroli wykształcenia rodziców. Na osi pionowej przedstawiono wartości wskaźników EWD wyliczone z modelu, w którym dodatkowo uwzględniono tę informację. Widać, że wskaźniki wyliczone z obu modeli są bardzo do siebie zbliżone. Gdyby uwzględnić błąd ich oszacowania, wskaźniki EWD otrzymane z modelu bez kontroli wykształcenia nie różniłyby się istotnie od tych otrzymanych z modelu z kontrolą inteligencji. W próbie 150 gimnazjów uzyskano jednoznaczny obraz – dodanie do modelu zmiennej wykształcenia rodziców w bardzo niewielkim stopniu wpłynęło na oszacowane EWD szkół. Współczynnik korelacji Pearsona EWD z EWD_{hedu} wynosi $r = 0,991$.

Przed ostatecznym wnioskiem wstrzymajmy się do przedstawienia wyników dla egzaminu gimnazjalnego w części humanistycznej.

5.3.4. Status ekonomiczny a wyniki egzaminu gimnazjalnego w części humanistycznej

Analogiczne analizy przeprowadzone zostały dla łącznego wyniku egzaminu gimnazjalnego z języka polskiego i przedmiotów humanistycznych. W tabeli 7 zaprezentowano ich wyniki.

Tabela 7. Czynniki statusowe a wyniki z egzaminu gimnazjalnego w części humanistycznej, model bez kontroli uprzednich osiągnięć. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych, współczynniki regresji dla standaryzowanych (0;1) zmiennych wyjaśniających, zmienna zależna standaryzowana (100;15)

zamienna zależna: wynik egzaminu gimnazjalnego w części humanistycznej							
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4) (s.e.)
oszacowanie efektów stałych							
poziom ucznia							
wykształcenie (HEDU)	5,42	(0,271)	–	–	–	–	–
status (HISEI)	–	–	4,32	(0,220)	–	–	–
prestż (HSIOPS)	–	–	–	–	3,95	(0,225)	–
zasobność	–	–	–	–	–	–	4,45 (0,233)
stała	101,67	(0,432)	101,50	(0,431)	101,22	(0,428)	101,15 (0,446)
oszacowanie efektów losowych							
wariancja efektów szkół	9,69		10,71		11,23		12,02
wariancja efektów oddziałów	12,71		13,23		13,74		13,90
wariancja na poziomie ucznia	156,49		163,53		165,16		158,99
podsumowanie							
deviance	37336,72		38280,51		38334,04		39112,25
liczba szacowanych parametrów	5		5		5		5
liczba uczniów	4689		4781		4781		4900
liczba oddziałów	291		291		291		291
liczba szkół	150		150		150		150

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe

Analogicznie jak w części matematyczno-przyrodniczej egzaminu dla wszystkich czterech aspektów statusu rodziny ucznia uzyskano istotne statystycznie efekty. Przypomnijmy – współczynniki regresji w tabeli mówią, jak zmienia się wynik egzaminu gimnazjalnego (skala 100;15), gdy dana miara statusowa zmienia się o jedno odchylenie standardowe. Zmiana wykształcenia o jedno odchylenie oznacza wzrost wyników o około 5,4 punktu egzaminacyjnego, czyli trochę ponad jedną trzecią odchylenia. To silniejszy efekt niż w wypadu egzaminu matematyczno-przyrodniczego. Różnicy tej nie można przypisać różnicom w rzetelności testów egzaminacyjnych, więc można ją sensownie interpretować jako trochę silniejszą determinację

statusową w tym obszarze kształcenia. Pozostałe trzy efekty są trochę słabsze i wahają się wokół 4 punktów egzaminacyjnych. Tak jak dla części matematyczno-przyrodniczej efekty statusowe są słabsze niż ten obserwowany w wypadku IQ. Dla inteligencji dla tej części egzaminu zanotowano efekt 8,28.

Tabela 8. Czynniki statusowe a wyniki egzaminu gimnazjalnego w części humanistycznej, model EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych, współczynniki regresji dla standaryzowanych (0;1) zmiennych wyjaśniających, zmienna zależna standaryzowana (100;15)

zamienna zależna: wynik egzaminu gimnazjalnego w części humanistycznej								
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)
oszacowanie efektów stałych								
poziom ucznia								
sprawdzian	11,28	(0,249)	11,03	(0,296)	11,01	(0,288)	11,06	(0,301)
sprawdzian ²	-0,43	(0,127)	-0,255	(0,167)	-0,267	(0,167)	-0,271	(0,167)
sprawdzian ³	-0,34	(0,079)	-0,338	(0,097)	-0,331	(0,097)	-0,339	(0,095)
pleć ^a	1,86	(0,328)	-1,47	(0,325)	-1,49	(0,322)	-1,45	(0,309)
dysleksja na egzaminie gim.	1,97	(0,660)	1,51	(0,767)	1,45	(0,763)	1,45	(0,728)
dysleksja int.*	-3,49	(0,915)	-3,02	(0,784)	-2,96	(0,778)	-2,92	(0,755)
wykształcenie (HEDU)	2,00	(0,177)	–	–	–	–	–	–
status (HISEI)	–	–	1,57	(0,169)	–	–	–	–
prestż (HSIOPS)	–	–	–	–	1,50	(0,172)	–	–
zasobność	–	–	–	–	–	–	1,75	(0,169)
stała	101,23	(0,367)	101,29	(0,365)	101,22	(0,358)	101,18	(0,357)
oszacowanie efektów losowych								
wariancja efektów szkół	5,47		5,33		4,96		5,23	
wariancja efektów oddziałów	2,49		2,12		1,97		2,41	
wariancja na poziomie ucznia	74,49		75,59		75,87		75,20	
podsumowanie								
deviance	33785,25		34500,62		34507,55		35339,34	
liczba szacowanych parametrów	11		11		11		11	
liczba uczniów	4689		4781		4781		4900	
liczba oddziałów	291		291		291		291	
liczba szkół	150		150		150		150	

* interakcja dysleksji na sprawdzianie i egzaminie gimnazjalnym

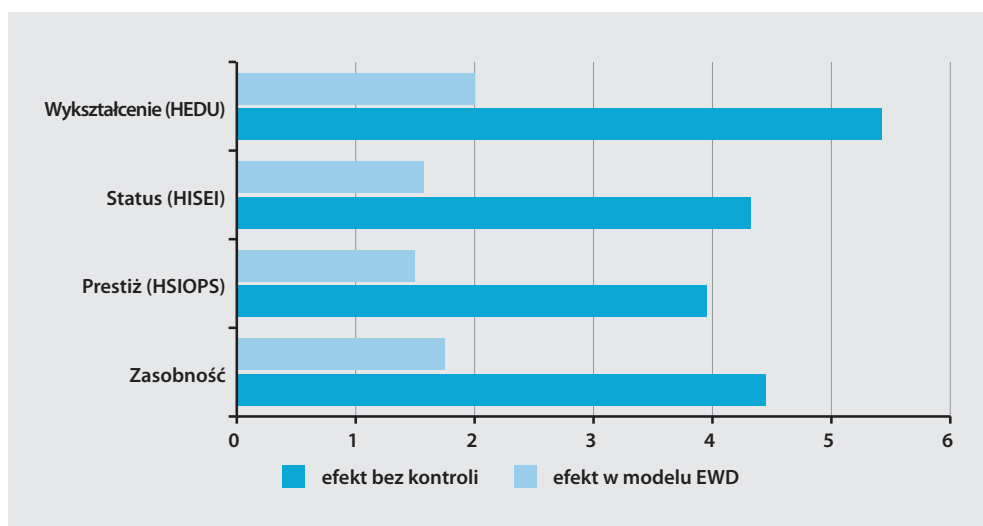
a – efekt dla dziewcząt

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe

Podążając tokiem narracji z poprzedniej części analiz, zadajmy pytanie, na ile skutecznie wyeliminuje efekt czynników statusowych wykorzystanie modelu EWD? Czy uwzględnienie uprzednich osiągnięć i pozostałych zmiennych kontrolnych pozwoli wyeliminować bądź przynajmniej silnie ograniczyć efekty SES? Spójrzmy na wyniki.

W obszarze humanistycznym w modelach EWD wszystkie czynniki statusowe są w dalszym ciągu istotnymi predyktorami wyników egzaminacyjnych. Oznacza to, że czynniki te wpływają nie tylko na poziom osiągnięć szkolnych na progu gimnazjum, ale również determinują postęp ucznia czyniony podczas trzech lat nauki w gimnazjum. Choć siła efektów jest znacznie słabsza niż w modelu bez kontroli uprzednich osiągnięć, to jest jednak znacząco silniejsza niż w wypadku obszaru matematyczno-przyrodniczego. Graficzne przedstawienie wartości współczynników regresji dla modelu bez kontroli uprzednich osiągnięć i modelu EWD zawiera poniższy wykres.

Rysunek 3. Siła efektów czynników statusowych w modelu bez kontroli uprzednich osiągnięć i w modelu EWD. Egzamin gimnazjalny w części humanistycznej



Widzimy, że w części humanistycznej model EWD słabiej kontroluje efekty statusowe. Obserwujemy znaczącą redukcję siły efektów, ale jest ona proporcjonalnie mniejsza i „pozostałe” efekty są silniejsze i sięgają nawet 2 punktów egzaminacyjnych. Czy oznacza to konieczność włączania informacji statusowej do modeli EWD w obszarze humanistycznym? Przyjrzyjmy się problemowi z uwzględnieniem szkolnego i oddziałowego poziomu analizy.

5.3.5. Znaczenie czynników statusowych dla szacowania EWD w obszarze humanistycznym

Zacznijmy od dekompozycji wariancji egzaminu humanistycznego na składowe: międzyszkolną, międzyoddziałową i indywidualną.

Tabela 9. Porównanie wariancji międzyszkolnej, międzyoddziałowej i indywidualnej w trzech modelach analizy. Egzamin w części humanistycznej

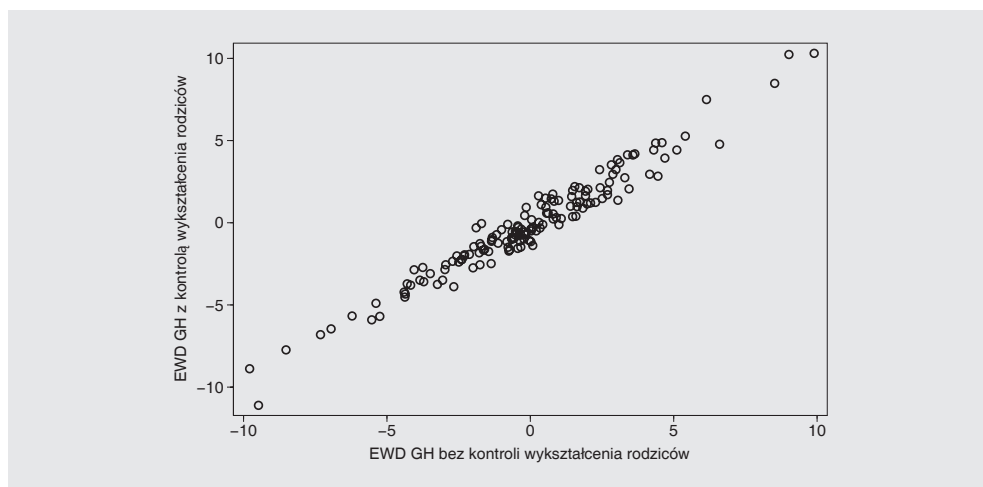
wariancja	model pusty	model EWD	model EWD z wykształceniem rodziców
międzyszkolna	19,83	5,72	5,47
międzyoddziałowa	17,53	2,62	2,49
indywidualna	174,59	77,11	74,49

Tak jak w wypadku egzaminu matematyczno-przyrodniczego, punktem wyjścia jest oszacowanie składowych wariancji dla trzech poziomów analizy. Odpowiednie wartości zawiera kolumna oznaczona „model pusty”. W modelu EWD wariancja międzyszkolna to 29% wariancji z modelu pustego. Dodanie do modelu zmiennej wykształcenie rodziców, podobnie jak poprzednio, niewiele zmienia – wariancja międzyszkolna w tym modelu to 28% wariancji w modelu pustym. Szczególnie efektywna jest w modelu EWD kontrola wariancji międzyoddziałowej – uwzględnienie zmiennych kontrolnych sprawia, że ta składowa wariancji to jedynie 15% zmienności w modelu pustym. W tym wypadku dodanie do modelu zmiennej wykształcenia też niewiele zmienia. Podobnie jak w części matematyczno-przyrodniczej obserwujemy, że kontrola wariancji międzyszkolnej i międzyoddziałowej w modelu EWD jest znacznie skuteczniejsza niż kontrola wariancji indywidualnej. Niewyjaśniona zmienność wyników egzaminacyjnych na poziomie ucznia to 44% oszacowania z modelu pustego.

Czyli w obszarze humanistycznym podobnie jak w matematyczno-przyrodniczym model EWD z kontrolą wykształcenia rodziców tylko nieznacznie lepiej wyjaśnia wariancję na poziomie szkolnym i oddziałowym. Jednak czy to oznacza tożsamość wskaźników wartości dodanej policzonych bez i z kontrolą zmiennej „wykształcenie rodziców”?

Na podstawie wyników egzaminu gimnazjalnego z 2012 roku w części humanistycznej oszacowano dla każdej szkoły dwa wskaźniki edukacyjnej wartości dodanej: bez kontroli wykształcenia (EWD) i z kontrolą zmiennej wykształcenia rodziców (EWD_hedu). Do szacowania użyto trzypoziomowego modelu efektów mieszanych, wskaźniki są Bayesowskimi predykcjami a posteriori. Poniższy wykres rozrzutu pozwala na porównanie tych oszacowań.

Rysunek 4. Porównanie wskaźników EWD dla części humanistycznej liczonych bez i z kontrolą wykształcenia rodziców



Na osi poziomej znajdują się wartości EWD dla szkół biorących udział w badaniu wyliczone z modelu bez kontroli wykształcenia rodziców. Na osi pionowej przedstawiono wartości wskaźników EWD wyliczone z modelu, w którym dodatkowo uwzględniono tę informację. Obraz jest bardzo zbliżony do tego uzyskanego dla obszaru matematyczno-przyrodniczego. W próbie 150 gimnazjów dodanie do modelu zmiennej wykształcenia rodziców w niewielkim stopniu wpłynęło na oszacowane EWD szkół. Współczynnik korelacji Pearsona między EWD i EWD_{hedu} wynosi $r = 0,975$. Warto jednak odnotować, że ten współczynnik jest nieco niższy, niż dla poprzednio analizowanego obszaru nauczania.

Wyniki analiz wskazują zatem, że kontrola w szacowaniu EWD dla szkół wykształcenia rodziców uczniów nie wpłynęłaby na znaczącą zmianę wartości wskaźników.

5.4. Podsumowanie

Dobrym punktem wyjścia do podsumowania analiz trafności metody EWD z punktu widzenia czynników statusowych może być tabela 10. Zestawiono w niej korelacje szkolnych średnich poziomu wykształcenia rodziców ze średnim wynikiem egzaminacyjnym, wskaźnikiem EWD i wskaźnikiem EWD wyliczonym przy kontroli zmiennej wykształcenia rodziców na poziomie indywidualnym.

Tabela 10. Średni poziom wykształcenia rodziców^a w szkole a średni wynik egzaminu gimnazjalnego, EWD i EWD z uwzględnieniem wykształcenia. Współczynniki korelacji Pearsona

egzamin	średnia egzaminacyjna w szkole ^a	EWD szkoły	EWD _{hedu} szkoły
część matematyczno-przyrodnicza	0,680	0,270	0,139
część humanistyczna	0,662	0,181	-0,035

wartości zaznaczone pogrubionym drukiem są istotne statystycznie na poziomie istotności $p < 0,05$, a – oszacowania średnich z pustych modeli trzypoziomowych

Wnioski z tej analizy są bardzo pouczające. Średni wynik egzaminacyjny szkoły ma blisko 50% wariancji wspólnej ze średnią wykształcenia rodziców! Jeżeli używamy średniej egzaminacyjnej szkoły jako miary efektywności nauczania – a jest to w dalszym ciągu dość często spotykane – to w praktyce stawiamy znak równości między dobrą szkołą a szkołą pracującą w korzystnym społecznie środowisku. Analogiczne korelacje dla wskaźników EWD spadają do poziomu 0,2–0,3. To oznacza, że wariancja wspólna średniej poziomu wykształcenia rodziców w szkole i miary efektywności nauczania spada do poziomu 4–9%. Korelacja ta spada prawie do zera, gdy w modelowaniu EWD na poziomie ucznia uwzględnimy zmienną wykształcenia rodziców. Czy to oznacza, że wskaźniki uwzględniające SES są bardziej trafne? Tu już sprawa nie jest tak jednoznaczna. Nie można wykluczyć, że średni poziom wykształcenia rodziców jest skorelowany z efektywnością nauczania. Z pewnością nie jest to korelacja rzędu 0,7, ale korelacji rzędu 0,2–0,3 wykluczyć nie można. Światło na ten problem mogą rzucić wyniki analiz przedstawione w kolejnych rozdziałach.

Spróbujmy podsumować. Największe różnice między gimnazjami stwierdzono dla miary ISEI i poziomu wykształcenia rodziców – około 20% wariancji całkowitej tych zmiennych to wariancja międzyszkolna. Równocześnie miary te bardzo silnie skorelowane są z wynikami egzaminacyjnymi. Czy uwzględnienie w modelach EWD uprzednich osiągnięć skutecznie kontroluje wpływ czynników statusowych? Dla wszystkich miar statusu obserwujemy istotne statystycznie zależności na poziomie indywidualnym. Jednak siła efektu jest dwu-, czterokrotnie słabsza w porównaniu z modelem bez uwzględnienia wyników sprawdzianu. Oszacowane z wykorzystaniem zmiennych statusowych wskaźniki EWD dla szkół korelują ze zwykłymi miarami EWD na poziomie 0,98–0,99. Oznacza to, że brak w modelu EWD zmiennych statusowych w bardzo niewielkim stopniu obciąża oszacowania wskaźników dla szkół, a tym samym ma marginalne znaczenie dla trafności metody.

Literatura

- Anyon, J. (1983). Social class and hidden curriculum of work. [w:] Giroux, H. i Purple, D. (eds) *The hidden curriculum and moral education*. McCutchan, Berkeley.
- Asparouhov, T., i Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling. A Multidisciplinary Journal*, 16(3), 397–438. doi:10.1080/10705510903008204
- Ball, S., J. (1993). Education Markets, Choice and Social Class: The Market as a Class Strategy in the UK and USA. *British Journal of Sociology of Education*, Vol. 14, Issue 1, s. 3–20.
- Bourdieu, P. (2001). *Zaproszenie do Socjologii refleksyjnej*. Warszawa: Oficyna Naukowa.
- Bourdieu, P. i Passeron, J. (1990). *Reprodukcja. Elementy teorii systemu nauczania*. Warszawa: PWN.
- Brown, A. R., Finney, S. J. i France, M. K. (2011). Using the bifactor model to assess the dimensionality of the Hong Psychological reactance scale. *Educ. Psychol. Meas. Educational and Psychological Measurement*, 71, 170–185.
- Bruner, J. (2006). *Kultura edukacji*. Kraków: Universitas.
- Chen, F. F., West, S. G. i Sousa, K. H. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. *Multivariate Behavioral Research*, 41, 189–225.
- Coleman, J. S. (1966). *Equality of educational opportunity*. Washington, D.C. US Office of Education.
- Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *The American Journal of Sociology*, 94, s. 95–120.
- Dolata, R. (2008). *Szkoła – segregacje – nierówności*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Dolata R., Jakubowski M. i Pokropek A. (2013). *Polska oświata w międzynarodowych badaniach umiejętności uczniów PISA OECD. Wyniki, trendy, kontekst i porównywalność*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Eysenck H. i Eysenck M. (1998). *Podpatrywanie umysłu*. Gdańsk: GWP.
- Ganzeboom, H. B. G. (2010). *How was new ISEI-08 constructed*. Pobrano 7 czerwiec 2013, z <http://home.fsw.vu.nl/hbg.ganzeboom/isco08/qa-isei-08.htm>
- Ganzeboom, H. B. G. i Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25(3), 201–239. doi:10.1006/ssre.1996.0010

- Ganzeboom, H. B. G. i Treiman, D. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. [w:] Hoffmeyer-Zlotnik, J. P. i Wolf, C. (red.), *Advances in Cross-National Comparison* (s. 159–193). Springer US. Pobrano z http://dx.doi.org/10.1007/978-1-4419-9186-7_9
- Ganzeboom, H. B. G., De Graaf, P. M. i Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56. doi:10.1016/0049-089X(92)90017-B
- Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T. i Plomin, R. (2009). More Than Just IQ: School Achievement Is Predicted by Self-Perceived Abilities—But for Genetic Rather Than Environmental Reasons. *Psychological Science*, Vol. 20 Issue 6, s. 753–762.
- Harris, J. R. (2000). Geny czy wychowanie? *Co wyrośnie z naszych dzieci i dlaczego*. Warszawa: Jacek Santorski & Co Wydawnictwo.
- Haworth, C. M. A., Asbury, K., Dale, P. S. i Plomin, R. (2011). Added Value Measures in Education Show Genetic as Well as Environmental Influence. *PLoS ONE*, Volume 6, Issue 2, e16006.
- Haworth, C.M.A., Wright, M.J, Luciano, M., Martin, N.G., de Geus, E.J.C., van Beijsterveldt, C.E.M., Bartels, M, Posthuma, D, Boomsma, D.I., Davis, O.S.P., Kovas, Y., Corley, R.P., DeFries, J.C., Hewitt, J.K., Olson, R.K., Rhea, S.A., Wadsworth, S.J., Iacono, W.G., McGue, M., Thompson, L.A., Hart S.A., Petrill, S.A., Lubinski, D. i Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood, *Molecular Psychiatry*, No15, s. 1112–1120.
- Herrnstein, R. i Murray, J. (1994). *The Bell Curve: Intelligence and class structure in American life*. New York. Free Press.
- Holzinger, K. i Swineford, F. (1937). The Bi-factor method. *Psychometrika*, 2, 41–54.
- International Labour Organization (ILO). (2007). *Bureau of Statistics, work unit of the Policy Integration Department*. Pobrano 5 czerwiec 2013, z <http://www.ilo.org/public/english/bureau/stat/isco/intro.htm>
- Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39 (1), s. 1–123.
- Jenses, A. R. (1991). Spearman's g factor and the problem of educational equality. *Oxford Review of Education*. 17 (2), s. 169–187.
- Lai, J.-S., Butt, Z., Wagner, L., Sweet, J. J., Beaumont, J. L., Vardy, J., ... Cella, D. (2009). Evaluating the Dimensionality of Perceived Cognitive Function. *Journal of Pain and Symptom Management*, 37, 982–995.

- Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S. i Nagengast, B. (2011). Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement. *Journal of Psychoeducational Assessment*, 29(4), 322–346. doi:10.1177/0734282911406657
- Marsh, Herbert W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U. i Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471–491. doi:10.1037/a0019227
- Marsh, Herbert W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S. i Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling. A Multidisciplinary Journal*, 16(3), 439–476. doi:10.1080/10705510903008220
- Marsh, Herbert W., Nagengast, B. i Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, 49(6), 1194–1218. doi:10.1037/a0026913
- Marsh, Herbert W., Nagengast, B., Morin, A. J. S., Parada, R. H., Craven, R. G. i Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology*, 103(3), 701–732. doi:10.1037/a0024122
- Mattsson, M. (2012). Investigating the factorial invariance of the 28-item DBQ across genders and age groups: An Exploratory Structural Equation Modeling Study. *Accident Analysis & Prevention*, 48, 379–396. doi:10.1016/j.aap.2012.02.009
- Molenaar, D., Sluis, S., Boomsma, D., Haworth, C., Hewitt, J., Martin, N., Plomin, R., Wright, M. i Dolan, C. (2013). Genotype by Environment Interactions in Cognitive Ability: A Survey of 14 Studies from Four Countries Covering Four Age Groups. *Behavior Genetics*. Vol. 43 Issue 3, s. 208–219.
- Morin, A. J. S., Maïano, C. (2011). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport and Exercise*, 12(5), 540–554. doi:10.1016/j.psychsport.2011.04.003
- Myers, N. D., Chase, M. A., Pierce, S. W., Martin, E. (2011). Coaching efficacy and exploratory structural equation modeling: a substantive-methodological synergy. *Journal of sport & exercise psychology*, 33(6), 779–806.
- OECD (2009). PISA 2006 Technical Report. Washington: OECD.
- Organisation for Economic Co-Operation and Development Staff. (2009). PISA PISA 2006 *Technical Report*. Washington: Organization for Economic Cooperation & Development.

- Pinker, S. (2005). *Tabula rasa. Spory o naturę ludzką*. Gdańsk: GWP.
- Plomin, R., DeFries, J. C., McClearn, G. E. i McGuffin, P. (2001). *Genetyka zachowania*. Warszawa: Wydawnictwa Naukowe PWN.
- Putnam, R. D. (2008). *Samotna gra w kręgle. Upadek i odrodzenie wspólnot lokalnych w Stanach Zjednoczonych*. Warszawa: Wydawnictwa Akademickie i Profesjonalne.
- Sánchez-Carracedo, D., Barrada, J. R., López-Guimerà, G., Fauquet, J., Almenara, C. A. i Trepal, E. (2012). Analysis of the factor structure of the Sociocultural Attitudes Towards Appearance Questionnaire (SATAQ-3) in Spanish secondary-school students through exploratory structural equation modeling. *Body Image*, 9(1), 163–171. doi:10.1016/j.bodyim.2011.10.002
- Treiman, D. J. (1977). *Occupational prestige in comparative perspective*. New York: Academic Press.
- Willis, P. (1981). *Learninig: How working class kids get working class jobs*. New York: Columbia University Press.
- Yu, C. Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes*. University of California, Los Angeles. Pobrano z <http://books.google.pl/books?id=8TdtuAAACAAJ>
- Zawistowska, A. (2012). *Horyzontalne nierówności edukacyjne we współczesnej Polsce*. Warszawa: Wydawnictwo Naukowe Scholar.

Rozdział 6

Korepetycje i rodzinne wsparcie edukacyjne¹

W powszechnej opinii przewaga szkolna uczniów pochodzących z rodzin o wyższym statusie społeczno-ekonomicznym wynika w dużej mierze z większych możliwości uczęszczania na różnorodne dodatkowe zajęcia edukacyjne i ogólnorozwojowe. Pośrednio lub bezpośrednio pomagają im one uzyskać wyższe wyniki w nauce w porównaniu do uczniów, których rodziny ze względów finansowych na tego rodzaju wsparcie nie mogą sobie pozwolić lub nie uważają tego typu wsparcia edukacyjnego za ważne. Z perspektywy nierówności edukacyjnych stanowi to czynnik pogłębiający te nierówności. Rodzi także problemy, jeśli chcemy oceniać pracę szkół. Wyniki egzaminacyjne uczniów traktuje się jako kryterium efektywności nauczania. Jeżeli jednak zależą one, przynajmniej częściowo, od uczestnictwa uczniów w dodatkowych zajęciach pozaszkolnych, to wypracowywane są one zarówno przez szkołę, jak i rodzinę ucznia. Tymczasem odpowiedzialność za nie przypisywana jest wyłącznie szkole. W konsekwencji mogłoby się wydawać, że szkoła pracująca w środowisku uboższym (np. na terenach wiejskich), z uczniami niemającymi możliwości skorzystania z dodatkowych zajęć, stoi na straconej pozycji w porównaniu ze szkołą ze środowiska zamożniejszego, np. wielkomiejskiego. W drugiej z nich bowiem wyniki uczniów mają szansę na większy „zastrzyk” w postaci wkładu rodzinny, za który laury zbiera później szkoła. Analiza tego problemu mieści się w obszarze badania trafności interpretacji przyczynowej wskaźników EWD. Jeżeli wsparcie rodzinne i korepetycje znacząco determinują postęp edukacyjny w gimnazjum, to modele EWD nieuwzględniające tych zmiennych, dawałyby obciążone wskaźniki EWD.

Potoczne przekonanie o pozytywnym wpływie korepetycji opiera się na trzech zdroworozsądkowych założeniach. Po pierwsze przyjmuje, że korepetycje (i inne formy pomocy w nauce, z których korzysta uczeń) są zjawiskiem na tyle częstym, że istotnym z perspektywy wyników egzaminacyjnych szkoły. Po drugie zakłada, że kierunek tej zależności jest pozytywny – uczęszczanie na dodatkowe prywatne lekcje korzystnie wpływa na wyniki w nauce i podobnego efektu nie da się osiągnąć samodzielną pracą. Innymi słowy uczniowie pobierający korepetycje, lepiej opanowują materiał w porównaniu do tych, którzy na korepetycje nie uczęszczają. Po trzecie, ten pozytywny wpływ ujawnia się także na egzaminie zewnętrznym. Prawdziwość tych założeń nie jest jednak weryfikowana a samo zjawisko korepetycji rzadko stanowi przedmiot pogłębionych badań na gruncie polskim. Czy rzeczywiście uczestnictwo w prywatnych lekcjach jest częste? Czy jest wyrazem aktywnego dążenia ucznia do podniesienia wiedzy ponad przeciętny poziom oraz do tego prowadzi? A może jest reaktywne – stanowi dowód trudności dziecka w nauce i ma na celu

¹ Wstępna wersja fragmentu tego rozdziału pojawiła się w pracy Hawrot i Jasińska (2013). *Kto się uczy poza szkołą? Predyktory korzystania z płatnej i bezpłatnej pomocy w nauce*. w: Niemierko i Szmigiel, K. (red.). *Polska edukacja w świetle diagnoz prowadzona z różnych perspektyw badawczych*. Kraków: Grupa Tomami.

nadrobienie zaległości i opanowanie materiału na minimalnym wymaganym w szkole poziomie? Co istotne, potocznie przyjmuje się, że częstotliwość i przyczyny podejmowania dodatkowego kształcenia są jednakowe podczas całej kariery szkolnej dziecka. Czy jednak uczeń w szkole podstawowej z tych samych przyczyn i równie często korzysta z pozaszkolnego wsparcia w nauce co uczeń przygotowujący się do egzaminu gimnazjalnego? A może te same przyczyny korzystania z dodatkowego kształcenia obecne są przez cały okres nauki w szkole, lecz na różnych etapach dotyczą różnych odsetków uczniów? Odpowiedzi na te pytania mają duże znaczenie między innymi dla przewidywania kierunku omawianej zależności. Jeżeli w rzeczywistości uczestnictwo w korepetycjach jest raczej sygnałem trudności, to uczniowie na nie uczęszczający, mogą uzyskiwać na egzaminie wyniki niższe niż uczniowie bez tych kłopotów.

Jak już wspomniano, posyłanie dziecka na korepetycje może być wyrazem aspiracji rodziców, ich troski o dziecko i jego przyszłość, stopnia, w jakim uznają zdobycie wykształcenia za istotne. W tej sytuacji jednak należałoby się spodziewać, że także ubożsi rodzice mogą podejmować podobne działania. Dziecko może korzystać z pomocy rodziców, starszego rodzeństwa lub wolontariuszy, zarówno w ramach codziennej nauki, jak i przygotowania do egzaminu zewnętrznego. Z tej perspektywy istotne wydaje się raczej podejmowanie przez rodzinę działań o wspierającym charakterze, a mniejsze znaczenie ma forma. Z drugiej jednak strony wiadomo, że to właśnie wyższy status społeczny związany jest z większym zaangażowaniem w życie szkolne dziecka (Grolnick, Benjet, Kurowski i Apostoleris, 1997), wobec czego to właśnie rodzice o wyższym statusie częściej mogą decydować się na np. samodzielną pomoc w nauce. Podobna sytuacja może mieć miejsce, gdy wsparcie stanowi reakcję na szkolne problemy dziecka. Niezależnie jednak od rodzaju tych działań (płatnych vs. bezpłatnych), jeżeli mają one znaczenie dla późniejszych osiągnięć dziecka, będą stanowić pozaszkolny wkład w jego wynik egzaminacyjny.

Problem zależności wyników egzaminacyjnych od korzystania przez uczniów z pozaszkolnego wsparcia jest także istotny w kontekście wskaźników edukacyjnej wartości dodanej. Jeżeli rzeczywiście korzystanie ze wsparcia w nauce stanowi element życia szkolnego znaczącej części uczniów i jest silnie powiązane z ich wynikami egzaminacyjnymi, naturalnym staje się pytanie o trafność wskaźników EWD. Wszak w ich modelowaniu wykorzystuje się właśnie wyniki egzaminów zewnętrznych, nie uwzględniając żadnych informacji o uzyskiwanym przez uczniów wsparciu pozaszkolnym. Co istotne, takie dane są niedostępne i z bardzo dużym prawdopodobieństwem nigdy dostępne nie będą. Oczwistym jest więc, że konieczna jest weryfikacja, na ile wskaźniki EWD są odporne na ewentualne zniekształcenia wynikające z nieuwzględniania w modelach tych informacji.

W niniejszym rozdziale zajmiemy się problemem znaczenia różnych form rodzinnego wsparcia w nauce dla wyników egzaminacyjnych oraz EWD. Przyjrzymy się powszechności tego zjawiska oraz poddamy weryfikacji założenie o pozytywnym kierunku zależności między

otrzymywaniem przez ucznia tego wsparcia a jego wynikami na egzaminie gimnazjalnym. Ponadto, próbując odpowiedzieć na pytanie o trafność wskaźników EWD jako miar efektywności pracy szkoły, sprawdzimy, jakie są konsekwencje nieuwzględniania w modelach EWD informacji o uzyskiwanej przez ucznia pomocy w postaci korepetycji oraz bezpłatnego wsparcia.

6.1. Korepetycje i wsparcie edukacyjne

Wsparcie w nauce udzielane dziecku może przyjmować wiele form, np. zajęć ogólnorozwojowych (sportowych, artystycznych i innych), korepetycji, bezpłatnych zajęć organizowanych przez szkołę (wyrównawczych, kół zainteresowań). Z punktu widzenia trafności EWD szczególne znaczenie mają te, które bezpośrednio dotyczą materiału obowiązującego na egzaminach i jednocześnie są niezależne od szkoły. Stąd w centrum naszego zainteresowania leżą dwa rodzaje zajęć: korepetycje oraz bezpłatne wsparcie w nauce przedmiotów, które są obecne na egzaminie gimnazjalnym. Korepetycje rozumiemy jako dodatkowe odpłatne zajęcia z przedmiotów objętych nauczaniem w szkole (skoncentrowane np. na odrabianiu prac domowych, powtarzaniu materiału, nadrabianiu zaległości, przygotowaniu do sprawdzianów), realizowane indywidualnie lub w małych grupach. Za zjawisko odmienne od korepetycji uznajemy bezpłatne wsparcie edukacyjne, czyli bezpłatną pomoc w nauce i odrabianiu lekcji udzielaną przez rodzinę, wolontariuszy, świetlice środowiskowe i instytucje inne niż szkoła.

Ujęcie to jest zbieżne z definicją Marka Braya (2007), który definiuje korepetycję (*private tutoring*) jako dodatkowe zajęcia z przedmiotów objętych nauczaniem w szkole, realizowane przez korepetytorów w celach zarobkowych. Akcentuje ono, że korepetycjami nie są płatne i bezpłatne dodatkowe zajęcia ogólnorozwojowe (np. artystyczne, sportowe, lekcje języka ojczystego dla mniejszości narodowych) ani bezpłatne dodatkowe zajęcia edukacyjne organizowane przez szkołę (np. koła zainteresowań, zajęcia wyrównawcze dla uczniów z problemami w nauce itp.). Jednakże na świecie korepetycje często realizowane są w innych formach niż w Polsce, np. w dużych grupach, przez internet lub telefonicznie, a ich organizacją i prowadzeniem zajmują się niejednokrotnie wyspecjalizowane firmy. Udzielać ich mogą nauczyciele szkół państwowych swoim własnym uczniom lub uczniom z innych szkół, jak również nauczyciele bez stałego zatrudnienia oraz studenci i osoby niewykonujące zawodu nauczyciela (Bray, 2007). Nasza definicja ogranicza w pewnym stopniu różnorodność tych form, uwzględniając w ten sposób ich polską specyfikę. Jednakże pamiętać należy, że przyczyny, częstotliwość, konsekwencje i formy korepetycji wykazują duże zróżnicowanie kulturowe (Bray, 2007), stąd należy zachować ostrożność w uogólnianiu wyników badań zagranicznych, które w niniejszym rozdziale będziemy przytaczać.

6.1.1. Korepetycje i wsparcie edukacyjne w świetle badań

Przed badaczami zainteresowanymi problematyką korepetycji i wsparcia edukacyjnego stoi wiele wyzwań. W związku ze znacznymi kosztami trudno uzyskać reprezentatywne próby uczniów i rodziców, które pozwoliłyby na szacowanie skali obu zjawisk. Problematiczne jest też dotarcie do samych korepetytorów. Zazwyczaj działają oni niezależnie, nie zrzeszają się i nie przyznają do tej działalności ze względu na nieodprowadzanie z jej tytułu należnych podatków (Ireson, 2004). Nie prowadzi się także badań w placówkach, które w ramach swojej działalności mogą bezpłatnie pomagać w odrabianiu lekcji i przygotowaniu się do sprawdzianów. W związku z deklaracyjnym i retrospekcyjnym charakterem danych pojawiają się także trudności w uzyskaniu od uczniów oraz rodziców wiarygodnych i szczegółowych informacji dotyczących częstotliwości i regularności otrzymywanej przez dziecko pomocy w nauce (Shaughnessy, Zechmeister i Zechmeister, 2008). Dlatego niejednokrotnie uzyskiwane dane są fragmentaryczne i nie dają informacji o faktycznej skali zjawiska. Problematiczne jest również zróżnicowanie jakości korepetycji i wsparcia edukacyjnego, które utrudnia badanie ich znaczenia dla osiągnięć szkolnych.

6.1.1.1. Powszechność zjawiska korepetycji

Powszechność korepetycji wśród uczniów na różnych etapach edukacyjnych nie jest w Polsce znana. Mimo że pojawiają się próby jej oszacowania, niereprezentatywność próbek respondentów i fakt obejmowania badaniem osób łatwo dostępnych, np. studentów (np. Putkiewicz, 2006) sprawia, że nie dają one wiarygodnych informacji.

Częstkowych danych dostarczają wyniki sondażu CBOS zrealizowanego w październiku 2012 roku, a dotyczącego wydatków na edukację. Szesnaście procent respondentów (rodziców dzieci uczęszczających do szkoły podstawowej, gimnazjalnej lub ponadgimnazjalnej) zadeklarowało, że ich dzieci uczęszczają lub będą uczęszczać na korepetycje lub kursy przygotowawcze w rozpoczynającym się roku szkolnym. Trzydzieści procent zadeklarowało posłanie dzieci na dodatkowe zajęcia z języków obcych. Niestety, badanie zrealizowano na niewielkiej próbie, a w raporcie brakuje informacji na temat jej reprezentatywności (CBOS, 2012).

Orientacyjnych informacji dostarczają także wyniki badania PISA z 2006 roku (OECD, 2011). Sugerują one, że wśród piętnasto- i szesnastolatków odsetek uczniów uzyskujących pomoc w nauce poza lekcjami szkolnymi jest znaczący. W badaniu pytano o uczęszczanie na dodatkowe lekcje poza zajęciami obowiązkowymi, odbywające się albo indywidualnie, albo grupowo (w małych i dużych grupach), realizowane przez nauczycieli ze szkoły ucznia i spoza niej. Około 42% uczniów z krajów OECD zadeklarowało, że uczestniczy w tego typu zajęciach z języka polskiego, zaś z matematyki – około 48%. Odsetki polskich uczniów wyniosły odpowiednio 44% i 51%. Dane te dotyczą jednak wszelkich płatnych i bezpłatnych lekcji zapewnianych zarówno przez szkołę, jak i rodziców czy instytucje inne niż szkoła. Najbardziej

interesujące z perspektywy badania faktycznego wkładu szkoły w osiągnięcia uczniów są zajęcia, które z relatywnie największym prawdopodobieństwem są organizowane poza nią. Do nich zaliczyć możemy lekcje poza regularnymi godzinami nauki prowadzone przez nauczycieli spoza szkoły. Do uczestnictwa w nich przyznało się 29% polskich uczniów. Pamiętać jednak należy, że dane te dotyczą zarówno zajęć płatnych, jak i bezpłatnych.

Powszechność korepetycji różni się znacznie między krajami, co uniemożliwia wnioskowania w tym zakresie na podstawie danych z państw o zbliżonej kulturze. Wiadomo jednak, że zazwyczaj cieszą się większą popularnością w kulturach podkreślających znaczenie wysiłku dla osiągnięć w porównaniu do kultur, które podkreślają znaczenie zdolności. Wydaje się też, że uczniowie częściej korzystają z takiej formy wsparcia w klasach poprzedzających egzaminy wysokiej stawki (Bray, 2007). Można też przypuszczać, że powszechność obu form wsparcia w nauce zmienia się na różnych etapach edukacyjnych. Pomoc udzielana przez rodziców może być powszechniejsza na początku szkoły podstawowej, natomiast wraz z kolejnymi etapami edukacyjnymi coraz częściej uczniowie mogą decydować się na pomoc płatną.

Jeszcze mniej danych znajdujemy w literaturze w zakresie bezpłatnego wsparcia edukacyjnego otrzymywanego przez uczniów. Nie udało się nam dotrzeć do polskich prac zawierających informacje o skali tego zjawiska. W badaniach holenderskich, które objęły około 12 000 uczniów, 44% dwunastolatków zadeklarowało, że rodzice regularnie pomagają im w odrabianiu prac domowych (Driessen, Smit i Slegers, 2005).

Podsumowując, brakuje rzetelnych danych dotyczących powszechności korzystania przez uczniów z korepetycji i bezpłatnej pomocy w nauce na różnych etapach edukacyjnych, także podczas nauki w gimnazjum. Nie wiadomo więc, czy są to zjawiska na tyle częste, by mogły mieć znaczenie dla wyników egzaminu gimnazjalnego znaczącej liczby uczniów, a tym samym potencjalnie obciążać wskaźniki EWD.

6.1.1.2. Korepetycje i wsparcie edukacyjne a osiągnięcia szkolne

Zakłada się, że uczęszczanie na korepetycje jest korzystne, tj. uczniowie dzięki korzystaniu z nich podwyższają poziom umiejętności szkolnych. Nie są nam znane raporty z polskich badań dotyczących efektywności korepetycji (ich znaczenia dla osiągnięć szkolnych). Wyniki badań prowadzonych w różnych krajach okazują się natomiast niekonkluzywne. W części z nich znaleziono pozytywną zależność, podczas gdy w innych była ona ujemna lub nieistotna statystycznie.

Hai-Anh Dang (2007) raportuje wyniki analiz prowadzące do wniosku, że wzrost wydatków wietnamskich gospodarstw domowych na korepetycje podwyższa prawdopodobieństwo znalezienia się ucznia w szkolnym rankingu opartym na średniej ocen w kategorii uczniów dobrych i bardzo dobrych oraz obniża prawdopodobieństwo znalezienia się w kategorii

uczniów słabych i przeciętnych. Zależność ta była widoczna zarówno na etapie szkoły podstawowej, jak ponadpodstawowej (drugi etap edukacyjny), choć w drugiej z nich okazała się silniejsza. Efekty te były jednak nieduże. W badaniu kontrolowano różnorodne zmienne ważne dla osiągnięć i uczestnictwa w korepetycjach.

O pozytywnym, lecz niewielkim efekcie korepetycji w Stanach Zjednoczonych donosi też Derek Briggs (2001), jednak w analizach uwzględniał on tylko płatne kursy przygotowawcze do egzaminów zewnętrznych, pomijając zajęcia indywidualne z korepetytorem. Uczestnictwo w takich kursach było predyktorem wyższych wyników na dwóch egzaminach zewnętrznych z języka ojczystego i czytania. W przypadku matematyki efekt był jednak zróżnicowany (pozytywny lub negatywny) w zależności od egzaminu. W badaniu kontrolowano zmienne powiązane zarówno z osiągnięciami na egzaminach, jak i uczęszczaniem na korepetycje. Jednakże objęto nim specyficzną próbę uczniów – tych, którzy podeszli do co najmniej dwóch nieobowiązkowych egzaminów wstępnych na studia. Samo w sobie mogło być to źródłem autoselekcji uczniów do egzaminów, wpływającej na szacowanie siły i kierunku badanej zależności.

Poza schemat badań korelacyjnych wyszli Christoph Mischo i Ludwig Haag (2002), realizując badania quasi-eksperymentalne wśród luksemburskich uczniów z klas 8–11, którzy zdecydowali się na uczestnictwo w regularnych korepetycjach (9 miesięcy, 4 razy w tygodniu po 90 minut). Grupę kontrolną stanowili wskazani przez nich i jednocześnie podobni do nich koledzy ze szkoły. Grupy nie różniły się ocenami i poziomem motywacji, lęku i przekonania o swoich zdolnościach. Po dziewięciu miesiącach oceny uczniów korzystających z korepetycji wzrosły w porównaniu do ocen uczniów z grupy kontrolnej, a sam efekt okazał się umiarkowany. W badaniu nie kontrolowano jednak innych zmiennych ważnych dla osiągnięć i decyzji o korzystaniu z korepetycji, jednak problem ten łagodzi w pewnym stopniu strategia wyboru do grupy kontrolnej uczniów ze szkół, do których uczęszczali uczniowie z grupy kryterialnej. Ponadto oceny szkolne traktowano jak zmienną ciągłą. Należy też zwrócić uwagę na znaczną ilość czasu poświęcaną na korepetycje, nietypową dla większości uczniów w Polsce.

Obszerną analizę znaczenia korepetycji dla wyników państwowego egzaminu wstępnego na wyższe uczelnie w Chinach przedstawia Yu Zhang (2013). Wykorzystano w niej różne sposoby pomiaru uczestnictwa w korepetycjach – pobieranie korepetycji z jakiegokolwiek przedmiotu, wysokość ponoszonych z tego tytułu wydatków i poświęcaną ilość czasu. Uzyskane wyniki są bardzo zróżnicowane, dlatego nie będą w całości prezentowane. Generalnie, nie wykryto istotnego efektu w całej próbie, jednak zależność pojawiła się w niektórych grupach uczniów w zakresie wybranych przedmiotów w szkołach o różnej efektywności. Autorka konkluduje, że może być ona pozytywna w zakresie niektórych przedmiotów wśród uczniów z terenów miejskich i o niższych osiągnięciach lub ze słabszych szkół. W przypadku wyniku ogólnego z egzaminu w całej

próbie efekt okazał się ujemny, jednak w dużej mierze odpowiadała za niego negatywna zależność wśród uczniów z terenów wiejskich. Wśród uczniów z miast żaden efekt nie wystąpił.

Literatura dotycząca znaczenia wsparcia edukacyjnego otrzymywanego przez dziecko dla jego osiągnięć szkolnych wydaje się jeszcze uboższa w porównaniu do literatury z zakresu korepetycji. W badaniach problem ten najczęściej pojawia się w kontekście zaangażowania rodziców w edukację dziecka i ograniczony jest do udzielania przez rodziców pomocy w odrabianiu prac domowych lub sprawdzania tych prac. Rozpatrywany jest wtedy, obok zagadnień takich jak: jakość i częstotliwość kontaktów ze szkołą i nauczycielami czy monitorowanie postępów w nauce, jako jeden z przejawów zaangażowania rodziców w życie szkolne dziecka (Hattie, 2009; Rosenzweig, 2001). Wyniki badań nie rozstrzygają jednoznacznie kierunku zależności między otrzymywaniem przez dziecko tego rodzaju wsparcia a jego osiągnięciami. Z przeglądu literatury Kathleen Hoover-Dempsey i współpracowników (2001) wynika, że część badaczy raportuje pozytywne zależności, podczas gdy inni – negatywne. Autorzy zauważają, że różnice te wynikać mogą z różnych przyczyn udzielania dziecku takiego wsparcia. Z jednej strony może być to reakcja na wyniki w nauce poniżej oczekiwań lub wymagań, z drugiej – wyraz chęci obserwowania postępów dziecka i wspierania go. W pierwszym przypadku udzielona pomoc może dawać pozytywne rezultaty, lecz niekoniecznie wystarczające, by dziecko zaczęło otrzymywać zadowalające oceny. Wyniki szczególnie dla nas interesujące, bo zaobserwowane w grupie adolescentów, raportują Sehee Hong i Hsiu-Zu Ho (2005). Wskazują oni na negatywny pośredni wpływ kontroli rodzicielskiej, w tym kontroli prac domowych, na osiągnięcia szkolne nastolatków, lecz wykryte efekty okazały się relatywnie słabe. Podobne wyniki w swojej metaanalizie uzyskała Charlotte Rosenzweig (2001).

Pomoc w odrabianiu prac domowych przez rodziców jest jednak tylko jedną z wielu form pozaszkolnego wsparcia edukacyjnego. Przypomnijmy, że obejmować ono może także pomoc udzielaną przez instytucje inne niż szkoła (na przykład świetlice środowiskowe), lokalną społeczność lub wolontariuszy. Nie udało się nam dotrzeć do badań dotyczących pozostałych form tego wsparcia.

Podsumowując, potoczne przekonania o pozytywnym wpływie uczęszczania na korepetycje oraz otrzymywania pomocy w nauce nie znajdują jednoznacznego potwierdzenia w wynikach badań, choć nie są przez nie obalane. Sam problem okazuje się znacznie bardziej złożony badawczo, niż powszechnie się sądzi, gdyż wymaga kontrolowania zmiennych powiązanych z decyzją o udzieleniu wsparcia, w tym (w perspektywie podłużnej) samych osiągnięć w momencie podjęcia tej decyzji. Tymczasem rozpatrywanie znaczenia korepetycji i bezpłatnego wsparcia edukacyjnego dla EWD ma sens jedynie wtedy, gdy rzeczywiście mają one wpływ na wyniki egzaminów zewnętrznych. Jeżeli są niepowiązane, nie istnieje ryzyko obciążenia wskaźników ze względu na korzystanie przez uczniów z pozaszkolnych form kształcenia.

6.1.1.3. Korelaty korzystania z płatnego i bezpłatnego wsparcia w nauce

Zauważyć należy, że zarówno pobieranie korepetycji, jak i otrzymywanie bezpłatnego wsparcia w nauce, nie jest w populacji losowe, lecz odbywa się poprzez autoselekcję – uczniowie i ich rodziny na podstawie pewnych kryteriów podejmują decyzję o korzystaniu z tego rodzaju wsparcia. Co istotne, kryteria te mogą być jednocześnie powiązane z osiągnięciami szkolnymi, np. korepetycje mogą wybierać uczniowie z kłopotami w nauce po to, by nadrobić materiał lub poprawić oceny. Stąd badanie zależności między uczęszczaniem na korepetycje i otrzymywaniem bezpłatnej pomocy oraz wynikami egzaminów zewnętrznych wymaga ich kontroli. Przyjrzyjmy się więc, jakie czynniki wskazywane są w literaturze jako powiązane zarówno z osiągnięciami szkolnymi, jak i podejmowaniem decyzji o korzystaniu z pozaszkolnych form kształcenia.

Pozytywnej korelacji z osiągnięciami (Hattie, 2009; OECD, 2010; Sirin, 2005) oraz z częstotliwością korzystania przez dziecko z korepetycji (Dang, 2007; Tansel i Bircan, 2006) spodziewać się można w przypadku statusu społecznego. Druga z tych zależności wynikać może z między innymi większych środków finansowych, które przeznaczyć może rodzina o wyższym statusie na dodatkowe, płatne kształcenie dzieci – dochód rodziny pozwala przewidywać wydatki w tym zakresie (Holloway, Yamamoto, Suzuki i Mindnich, 2008). W badaniach ujawnia się także pozytywna zależność między statusem społecznym a aspiracjami i oczekiwaniami rodziców wobec dziecka (Singh i in., 1995; Trusty, 1998), jak również między aspiracjami, oczekiwaniami a wynikami w nauce (Hattie, 2009; Hong i Ho, 2005; Rosenzweig, 2001; Singh i in., 1995). Istotne jest też miejsce zamieszkania. Uczniowie ze środowisk wielkomiejskich osiągają lepsze wyniki w nauce w porównaniu do uczniów ze środowisk wiejskich (Dolata, 2008; IBE, 2012), przy czym zamieszkanie na wsi wiąże się też z niższym statusem społeczno-ekonomicznym (np. wskaźniki zagrożenia ubóstwem są ponad dwukrotnie wyższe na wsiach niż w miastach (GUS, 2012a, s. 299). Przypuszczać też można, że w Polsce rynek korepetycji jest większy w miastach niż na wsiach ze względu na np. możliwości finansowe mieszkańców, większą dostępność potencjalnych korepetytorów², gęstą sieć komunikacyjną. Dobrze udokumentowany w literaturze jest też pozytywny związek inteligencji z osiągnięciami szkolnymi (por. np. Hawrot i Pisarek, 2013), jednakże kierunek zależności inteligencja – pobieranie korepetycji i inteligencja – korzystanie z bezpłatnego wsparcia nie jest już tak oczywisty. Może być pozytywny, jeżeli rodzice zapewniają więcej zajęć dodatkowych uczniom bardziej uzdolnionym, chcąc zapewnić im optymalny rozwój. Zależność może mieć jednak odwrotny kierunek, jeżeli dodatkowe wspieranie ma charakter reaktywny, gdyż bardziej inteligentne dzieci rzadziej mają kłopoty w nauce.

² Np. gęstość zaludnienia w miastach jest wielokrotnie wyższa niż na wsiach (odpowiednio 1084 osoby/km² i 52 osoby/km² (GUS, 2012b, s. 31); odsetek osób z wyższym wykształceniem jest ponad dwukrotnie wyższy w miastach niż na wsiach, (GUS, 2012a, s. 193); uczelnie wyższe, których studenci niejednokrotnie udzielają korepetycji, skupione są w miastach.

Znaczenie mogą mieć też same aspiracje ucznia oraz jego motywacja. Z tej perspektywy uczestnictwo w dodatkowych płatnych zajęciach może być wyrazem przykładania większej wagi do nauki i sukcesu szkolnego w dłuższej perspektywie czasowej czy być wyrazem chęci rozwijania umiejętności. Taki rodzaj motywacji wydaje się być powiązany z wyższymi osiągnięciami szkolnymi (Black i Deci, 2000; Vansteenkiste, Sierens, Soenens, Luyckx i Lens, 2009). Z drugiej strony uczniowie o niskiej motywacji lub o motywacji zewnętrznej mogą osiągać niskie wyniki (Black i Deci, 2000; Vansteenkiste i in., 2009), w konsekwencji czego są posyłani na korepetycje w celu poprawy ocen. Przypuszczać można, że o ile ci pierwsi mogą zyskiwać na uczestnictwie w korepetycjach i rzeczywiście dzięki nim uzyskiwać wyższe wyniki w nauce, to ci drudzy – niekoniecznie.

Zbliżonych zależności można się spodziewać w przypadku bezpłatnej pomocy w nauce i odrabianiu lekcji. Może być ona reakcją rodziców na wyniki poniżej oczekiwań lub wymagań, z drugiej – wyraz chęci obserwowania postępów dziecka i wspierania go (Hoover-Dempsey i in., 2001).

Widoczne jest więc, że badanie zależności między wynikami w nauce a uczęszczaniem na korepetycje i otrzymywaniem bezpłatnego wsparcia jest złożone. Jednym z kluczowych aspektów jest kontrolowanie zmiennych powiązanych jednocześnie z decyzją o dodatkowej nauce pozaszkolnej i osiąganymi wynikami. Nieuwzględnianie ich w analizach doprowadzić może do wniosków nieznajdujących odzwierciedlenia w rzeczywistości.

6.2. Metoda

6.2.1. Pytania badawcze

W celu sprawdzenia, na ile wskaźniki EWD stanowią trafną miarę efektywności pracy szkoły, sformułowaliśmy szereg pytań związanych ze znaczeniem dla EWD pobierania przez uczniów korepetycji i otrzymywania bezpłatnej pomocy w nauce i odrabianiu lekcji. Jednak u podłoża rozpatrywania wrażliwości (czy też odporności) wskaźników EWD na efekty korepetycji i wsparcia leżą dwa założenia. Po pierwsze – oba zjawiska dotyczą znaczącej liczby uczniów. Po drugie – fakt otrzymywania przez dziecko tych form pomocy przynosi efekty w obszarze osiągnięć dziecka i przez to stanowi pozaszkolny wkład w jego umiejętności. Jak wyżej wskazano, brakuje danych dotyczących powszechności obu zjawisk. Ponadto wyniki badań w zakresie znaczenia korepetycji i nieodpłatnej pomocy w nauce dla osiągnięć szkolnych są niekonkluzywne, a w przypadku tego ostatniego – także fragmentaryczne. Tymczasem dla badania obciążenia wskaźników EWD ze względu na korzystanie przez uczniów z dodatkowego kształcenia jest właśnie powszechność tych zjawisk oraz siła ich związku z osiągnięciami szkolnymi uczniów. Dlatego zapytaliśmy, jaki odsetek uczniów korzysta z korepetycji oraz bezpłatnego wsparcia w nauce. Byliśmy też zainteresowani tym, czy istnieje zależność między otrzymywaniem przez dziecko obu rodzajów wsparcia a osiągnięciami szkolnymi.

Jednak ze względu na konieczność kontrolowania wielu zmiennych podczas analizowania tych zależności, wcześniej postawiliśmy szereg hipotez dotyczących związków różnych cech ucznia i rodziny z korzystaniem przez niego z korepetycji i bezpłatnego wsparcia w nauce (status społeczny, aspiracje edukacyjne rodziców wobec dzieci, inteligencja, motywacja, bezradność intelektualna). Hipotezy te nie będą przytaczane, ponieważ ich weryfikacja była potrzebna, by określić jakie zmienne należy kontrolować w modelach dotyczących tych zależności, a więc nie należą one do głównego wątku analiz.

W kolejnym kroku, w kontekście badania trafności wskaźników EWD jako miar efektywności pracy szkoły, zapytaliśmy o konsekwencje nieuwzględniania w modelach EWD informacji o pobieraniu przez uczniów korepetycji i korzystaniu z bezpłatnego pozaszkolnego wsparcia w nauce. Pytania te są istotne o tyle, że czynniki te stanowią rodzinny wkład w wyniki dziecka w nauce i jednocześnie nie można ich pozyskać na potrzeby modelowania wskaźników. Ich odporność na efekt korepetycji i wsparcia edukacyjnego jest więc kluczowa – w przypadku jej braku nie można bowiem ich interpretować jako miar efektywności pracy szkoły.

6.2.2. Schemat pomiarów

W analizach wykorzystano dane zebrane w różnych punktach czasowych. W zakresie osiągnięć szkolnych u progu gimnazjum oraz na jego zakończenie pozyskaliśmy (z baz egzaminacyjnych) informacje o wynikach na sprawdzianie po klasie szóstej oraz na egzaminie gimnazjalnym. Egzaminy odbyły się odpowiednio w maju 2009 oraz maju 2012 roku. Ponadto podczas pierwszej rundy badania (drugi semestr pierwszej klasy gimnazjum; luty – kwiecień 2010 r.) zebrane zostały następujące dane: płeć i poziom inteligencji dziecka, status społeczno-ekonomiczny rodziny dziecka, aspiracje edukacyjne rodziców względem dziecka oraz wielkość miejscowości, w której znajduje się szkoła. Pod koniec trzeciej klasy gimnazjum (maj 2012 r.) pozyskaliśmy dane dotyczące poziomu bezradności intelektualnej uczniów w zakresie języka polskiego i matematyki, ich motywacji do nauki, otrzymywania bezpłatnej pomocy w nauce w chwili obecnej oraz retrospekcyjne dane na temat korzystania z korepetycji w nauce z przedmiotów objętych programem nauczania podczas trzech lat nauki w gimnazjum.

6.2.2.1. Narzędzia badawcze i wskaźniki

Informacje o uczęszczaniu na korepetycje oraz korzystaniu z bezpłatnego wsparcia w nauce zebrano za pomocą ankiety wypełnianej przez ucznia. Zawierała ona szereg pozycji filtrujących, które w zależności od udzielonej odpowiedzi kierowały respondentą do różnych pytań.

W przypadku korepetycji zapytano, czy uczniowie kiedykolwiek podczas nauki w gimnazjum uczęszczali na korepetycje z jakiegokolwiek przedmiotu. Pytania zostały poprzedzone informacją o tym, o jaki rodzaj płatnej pomocy w nauce chodzi. Na potrzeby analiz przygotowano wskaźniki

uczęszczania na korepetycje z różnych przedmiotów: (1) matematyczno-przyrodniczych (matematyki, biologii, chemii, fizyki, geografii), (2) humanistycznych (języka polskiego, wiedzy o społeczeństwie, historii), (3) matematyki, (4) języka polskiego. Należy zwrócić uwagę, że wskaźniki te nie zawsze są rozłączne. Przyjęły one wartość 1, jeżeli uczeń zadeklarował, że kiedykolwiek podczas nauki w gimnazjum korzystał z korepetycji z danego przedmiotu/grupy przedmiotów, oraz wartość 0, gdy tego nie zadeklarował. Ponadto obliczono dla każdego oddziału wskaźniki powszechności korepetycji z danej grupy przedmiotów, tj. odsetki uczniów w oddziale korzystających z tego rodzaju wsparcia – był to drugi rodzaj wskaźnika.

W zakresie bezpłatnego wsparcia uczniowie zostali zapytani o to, czy ktoś obecnie pomaga im bezpłatnie w nauce lub odrabianiu lekcji z języka polskiego, matematyki oraz grup przedmiotów: historii i wiedzy o społeczeństwie; biologii, chemii, fizyki, geografii; języków obcych objętych programem nauczania w szkole. Odpowiedzi należało udzielić na pięciostopniowej skali (nie/tak, do 0,5 godz. tygodniowo/tak, od 0,5 do 1 godziny tygodniowo/tak, od 1 do 3 godzin tygodniowo/tak, więcej niż 3 godziny tygodniowo). W instrukcji sprecyzowano, o jaki rodzaj wsparcia w nauce chodzi. Na potrzeby analiz przygotowano cztery rodzaje wskaźników. Pierwszy z nich to ilość udzielanej dziecku bezpłatnej pomocy w nauce i odrabianiu lekcji z czterech grup przedmiotów: (a) języka polskiego, (b) matematyki, (c) historii i wiedzy o społeczeństwie, (d) przedmiotów przyrodniczych (biologii, chemii, fizyki, geografii). Wskaźnik przyjął wartości od 1 (nie otrzymuje takiej pomocy) do 5 (tak, więcej niż 3 godziny tygodniowo).

Drugi rodzaj wskaźników powstał w wyniku przekształcenia wskaźników opisanych powyżej. Kolejnym wartościom przypisano liczbę godzin³ otrzymywanej tygodniowo pomocy. Następnie obliczono sumy godzin uzyskiwanego wsparcia dla grup przedmiotów, tworząc w ten sposób wskaźniki nasilenia korzystania z korepetycji z przedmiotów matematyczno-przyrodniczych oraz humanistycznych. Tak utworzone wskaźniki miały rozkłady zbliżone do logarytmiczno-normalnego. Aby móc modelować związek ilości otrzymywanego bezpłatnego wsparcia z wynikami egzaminacyjnymi jako zależność liniową, zlogarytmizowano je. Trzeci typ wskaźnika, informujący o powszechności bezpłatnego wsparcia w nauce w oddziałach klasowych, utworzono poprzez obliczenie oddziałowej średniej arytmetycznej z opisanych powyżej wskaźników indywidualnych,

Czwarty typ wskaźników utworzono dla dwóch grup przedmiotów: (a) humanistycznych (język polski, historia, wiedza o społeczeństwie) oraz (b) matematyczno-przyrodniczych (matematyki, biologii, chemii, geografii). Przyjęły one wartość 1, jeżeli uczeń zadeklarował, że otrzymywał bezpłatną pomoc w nauce z danej grupy przedmiotów lub 0, jeżeli tego nie zadeklarował. Pomiar

³ wartości 1 (nie) – 0 godzin; wartości 2 (do 0,5 godziny) – 0,5 godziny; wartości 3 (od 0,5 do 1 godziny) – 0,75 godziny; wartości 4 (od 1 do 3 godzin) – 2 godziny; wartości 5 (powyżej 3 godzin) – 5 godzin, brak danych zastąpiono średnią danej zmiennej w próbie.

poziomu bezradności intelektualnej w zakresie matematyki i języka polskiego przeprowadzono za pomocą skali bezradności intelektualnej⁴ Grzegorza Sędką w wersjach dla języka polskiego i matematyki. Podstawowe informacje o bezradności intelektualnej zamieszczono w aneksie. Obie wersje skali składają się z 20 stwierdzeń dotyczących doświadczeń na lekcjach oraz w zetknięciu z wiedzą z danej dziedziny, typowych dla stanu bezradności intelektualnej lub jego braku. Uczeń miał za zadanie ustosunkować się do nich, zaznaczając odpowiedź na pięciopunktowej skali (nigdy/rzadko/czasem/często/zawsze). W ramach weryfikacji właściwości narzędzia przeprowadzono konfirmacyjną analizę czynnikową dla zmiennych porządkowych (Konarski, 2009). Na jej podstawie wygenerowano wartości czynnikowe dla każdego ucznia, które następnie wystandaryzowano (wyrażono w jednostkach odchylenia standardowego). Tak przygotowany wskaźnik wykorzystany został w analizach. W aneksie znajdują się szczegóły dotyczące specyfikacji modeli, ich dopasowania, ładunków czynnikowych pozycji oraz statystyk opisowych dla otrzymanych oszacowań poziomu bezradności intelektualnej uczniów. W analizach potraktowaliśmy wyniki tej skali jako dodatkową informację o doświadczaniu przez dziecko trudności w nauce.

Pomiar motywacji do nauki został przeprowadzony za pomocą skali motywacji ASRS. Narzędzie jest wstępną adaptacją holenderskiego testu motywacji autorstwa Martina Vansteenkiste i współpracowników (2009). W wersji oryginalnej składa się z 16 pozycji podzielonych na cztery równoliczne podskale mierzące cztery typy motywacji: motywację zewnętrzną, introjekcyjną, identyfikacyjną i wewnętrzną w rozumieniu teorii autodeterminacji (Ryan i Deci, 2000). W zastosowanej wersji usunięto pozycję pierwszą (podskala motywacji zewnętrznej) ze względu na wątpliwości dotyczącej jej trafności. Każda pozycja ma formę stwierdzenia dotyczącego przyczyny uczenia się szkolnego. Wypełniając test, respondent na pięciostopniowej skali (zupełnie nieważne/raczej nieważne/ani ważne, ani nieważne/raczej ważne/bardzo ważne) odpowiada, na ile każdy z tych powodów jest dla niego ważny, tj. odzwierciedla to, dlaczego on sam uczy się. W ramach badania właściwości narzędzia przeprowadzono konfirmacyjną analizę czynnikową dla zmiennych porządkowych (Konarski, 2009). Na jej podstawie wygenerowano wartości czynnikowe dla każdego ucznia, które następnie wystandaryzowano (wyrażono w jednostkach odchylenia standardowego). Tak przygotowane wskaźniki motywacji wykorzystane zostały podczas analizy skupień mającej na celu wyłonienie grup uczniów o różnych profilach motywacji. W jej toku każdy uczeń został zaklasyfikowany do jednego z pięciu profili motywacji: motywacji niskiej, motywacji o przeciętnym nasileniu, motywacji wysokiej, motywacji wewnętrznej lub motywacji zewnętrznej. Tak przygotowany wskaźnik został wykorzystany w analizach regresji. W aneksie znajdują się szczegóły dotyczące prac adaptacyjnych, specyfikacji i dopasowania modeli konfirmacyjnej analizy czynnikowej oraz dane dotyczące analizy skupień i wyłoniionych profili motywacji.

⁴ Narzędzie (za zgodą autora) dostarczyło Interdyscyplinarne Centrum Badań i Rozwoju Organizacji przy Zakładzie Psychologii Społecznej Instytutu Psychologii UJ.

Narzędzia do pomiaru pozostałych zmiennych oraz utworzone wskaźniki zostały opisane w innych rozdziałach niniejszej publikacji, w związku z tym nie będą ponownie omawiane (osiągnięcia szkolne – rozdział 2, inteligencja – rozdział 4, wskaźniki statusu społeczno-ekonomicznego oraz aspiracji edukacyjnych rodziców wobec dzieci – rozdział 5, wielkość miejscowości, w której znajduje się szkoła – rozdział 7).

6.2.3. Schemat analizy danych

W celu przedstawienia kontekstu dla rozważanego problemu w pierwszej części analiz dokonano próby scharakteryzowania zjawiska uczęszczania na korepetycje i otrzymywania bezpłatnego wsparcia w nauce, przedstawiając różnicowanie międzyszkolne i międzyoddziałowe oraz korelaty dla wspomnianych form otrzymywanego wsparcia. Mimo że analizy te nie służą bezpośrednio znalezieniu odpowiedzi na kluczowe dla podjętego problemu pytania badawcze, to są ważne z dwóch powodów. Po pierwsze, ukazują znaczenie różnicowania międzyszkolnego w zakresie otrzymywanej przez uczniów pomocy w nauce oraz dają przybliżone informacje o cechach szkół, w których korzystanie przez uczniów z korepetycji i bezpłatnego wsparcia jest bardziej nasilone. Dostarczy to informacji, dla jakich szkół wskaźniki EWD mogą być bardziej obciążone ze względu na korzystanie przez uczniów z dodatkowej, niezależnej od szkoły pomocy w nauce. Po drugie, wyniki tych analiz pokażą, które ze zmiennych kontrolowanych są związane nie tylko z osiągnięciami szkolnymi, ale także z prawdopodobieństwem otrzymywania dodatkowego wsparcia w nauce. Pozwolą więc lepiej wyspecyfikować modele weryfikujące znaczenie korzystania z dodatkowej pomocy w nauce dla wyników egzaminacyjnych oraz dokonać ich trafniejszej interpretacji. Pozwolą także lepiej zrozumieć różnice w zakresie znaczenia uczęszczania na korepetycje i otrzymywania bezpłatnej pomocy w nauce dla osiągnięć szkolnych.

Modele analizy dobrano tak, by pokazać zależności między otrzymywaniem dodatkowego wsparcia w nauce a poszczególnymi zmiennymi kontrolowanymi przy wytrąceniu znaczenia pozostałych zmiennych. Odzwierciedlono w nich także złożony schemat doboru próby (Domański i Pokropek, 2011). Aby ułatwić porównania wyników modeli pokazujących korelaty uczęszczania na korepetycje i otrzymywanie bezpłatnej pomocy w nauce, wskaźniki korzystania z płatnego i bezpłatnego dodatkowego wsparcia w nauce policzono w analogiczny do siebie sposób, tworząc zmienne dychotomiczne przyjmujące wartość 1 w sytuacji, w której badany korzystał z danego typu wsparcia oraz 0, jeśli z takiej pomocy nie korzystał. Do analiz wykorzystano wielopoziomowe modele regresji. Zastrzegamy jednak, że ze względu na schemat zbierania danych nie można interpretować wszystkich uzyskanych związków w kategoriach przyczynowo-skutkowych. Pomiaru części zmiennych dokonano pod koniec trzeciej klasy gimnazjum (poziom motywacji, bezradności intelektualnej), a wskaźnik korzystania z korepetycji dotyczy okresu trzech lat nauki w gimnazjum. Modele regresji wybrano w celu pokazania

związków korelacyjnych (po wytrąceniu znaczenia pozostałych zmiennych w modelu). Ze względu na dychotomiczny charakter zmiennych zależnych do analiz wykorzystano modele wielopoziomowych regresji logistycznych z efektami mieszanymi⁵ (Snijders i Bosker, 2012). Analizy przeprowadzono osobno dla przedmiotów matematyczno-przyrodniczych i osobno dla przedmiotów humanistycznych.

Analizy wykonano za pomocą oprogramowania HLM 6.06 (Scientific Software International, 2012), wykorzystując metodę estymacji *full PQL* (*penalized lub predictive quasi-likelihood*⁶). Do oszacowania istotności efektów stałych wykorzystano odporne (*robust*) błędy standardowe⁷. W tabelach z wynikami zaprezentowano oszacowania efektów stałych. Wartości istotne na poziomie istotności $p < 0,05$ zaznaczono pogrubionym drukiem. Dodatkowo, w celu ułatwienia interpretacji współczynników regresji, podano wyliczone na ich podstawie wartości ilorazu szans. Wyraża on zmianę szansy na zaobserwowanie danego zdarzenia (wartość 1 dla zmiennej zależnej), w przypadku gdy zmienna niezależna rośnie o jedną jednostkę, a pozostałe zmienne niezależne w modelu pozostają na tym samym poziomie. Analizy wykonano z uwzględnieniem wag warunkowych (patrz rozdział 1) na wszystkich trzech poziomach, odpowiednio przeskalowanych za pomocą programu HLM. Z analiz dla danej zmiennej zależnej wykluczono obserwacje, dla których zanotowano choć jeden brak danych na dowolnej zmiennej niezależnej uwzględnionej w porównywanych modelach.

Druga część analiz koncentruje się na znaczeniu korzystania z pomocy w nauce dla wyników egzaminacyjnych. W modelach tych wykorzystano dwie zmienne zależne: łączny wynik egzaminu gimnazjalnego z matematyki i przedmiotów przyrodniczych oraz łączny wynik egzaminu gimnazjalnego z języka polskiego i przedmiotów humanistycznych. Są one wyrażone na skali o średniej 100 i odchyleniu standardowym 15 w populacji uczniów, którzy przystąpili do egzaminu gimnazjalnego w 2012 roku. Analizy przeprowadzono, budując hierarchiczne modele liniowe. W pierwszym kroku pokazano siłę związku danej formy wsparcia w nauce z wynikami egzaminacyjnymi, uwzględniając ją jako jedyną zmienną niezależną w modelu, a w kolejnych krokach badano zmianę siły związku, jeśli w modelu kontrolowane są inne predyktory osiągnięć, m.in. zmienne uwzględniane w modelowaniu EWD.

⁵ Czytelnika niezaznajomionego z tymi modelami zachęcamy do lektury poświęconego tej tematyce rozdziału podręcznika Toma Snijdersa i Roela Boskera (2012), w którym w przystępny sposób przedstawiono podstawowe założenia tych modeli.

⁶ Metoda estymacji, która daje dość podobne oszacowania efektów stałych do innych powszechnie wykorzystywanych metod (jak np. *laplace approximation*), jednak może dać obciążone oszacowania efektów losowych (Snijders i Bosker, 2012 s. 300–301). Podstawowym przedmiotem zainteresowania są jednak w naszym przypadku oszacowania efektów stałych, stąd zastosowano domyślną dla HLM 6 metodę estymacji.

⁷ Są to oszacowania błędów standardowych odporne na naruszenie założeń dotyczących normalności rozkładu reszt (patrz np. Konarski, 2009, s. 318–324).

Analizy wykonano za pomocą oprogramowania HLM 6.06, wykorzystując metodę estymacji *full maximum likelihood*. Do oszacowania istotności efektów stałych wykorzystano odporne (*robust*) błędy standardowe. Wartości istotne na poziomie istotności $p < 0,05$ zaznaczono w tabelach pogrubionym drukiem. Analizy wykonano z uwzględnieniem wag warunkowych (patrz rozdział 1) na wszystkich trzech poziomach, odpowiednio przeskalowanych za pomocą programu HLM.

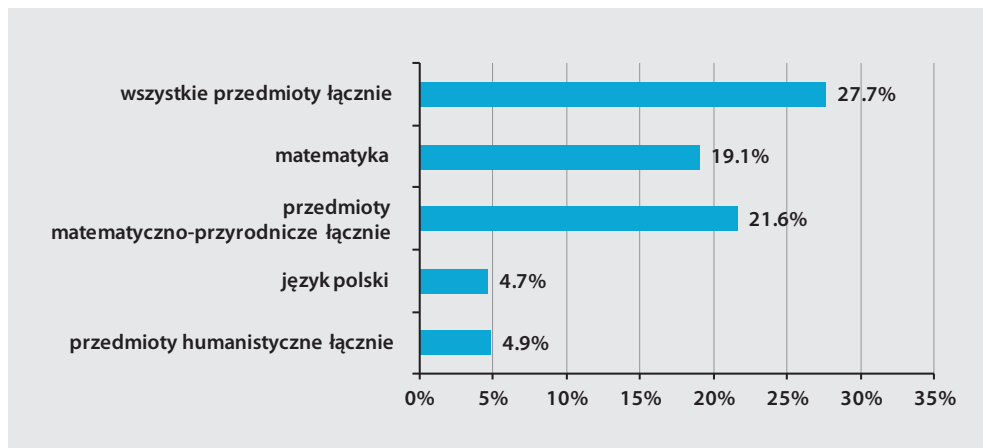
6.3. Wyniki

6.3.1. Skala zjawiska korepetycji oraz nieodpłatnej pomocy w nauce i odrabianiu lekcji

Problemem uczęszczania na korepetycje lub otrzymywania bezpłatnej pomocy w nauce warto się zająć w kontekście badania trafności wskaźników EWD tylko wtedy, jeśli znaczący odsetek uczniów otrzymuje taką pomoc. Gdyby zjawisko to było marginalne, wyniki egzaminacyjne większości uczniów w kraju nie byłyby uzależnione od pozaszkolnego wsparcia edukacyjnego. Przyjrzyjmy się zatem danym pokazującym skalę interesującego nas zjawiska wśród uczniów szkół gimnazjalnych.

Poniższy wykres pokazuje odsetki uczniów, którzy zadeklarowali skorzystanie z korepetycji z różnych grup przedmiotów choć raz podczas trzech lat nauki w gimnazjum. Podczas interpretacji danych, należy pamiętać, że uczniowie mogli wymienić więcej niż jeden przedmiot, tak więc kategorie nie są rozłączne.

Rysunek 1. Odsetek uczniów deklarujących uczęszczanie na korepetycje z wymienionych przedmiotów podczas nauki w gimnazjum

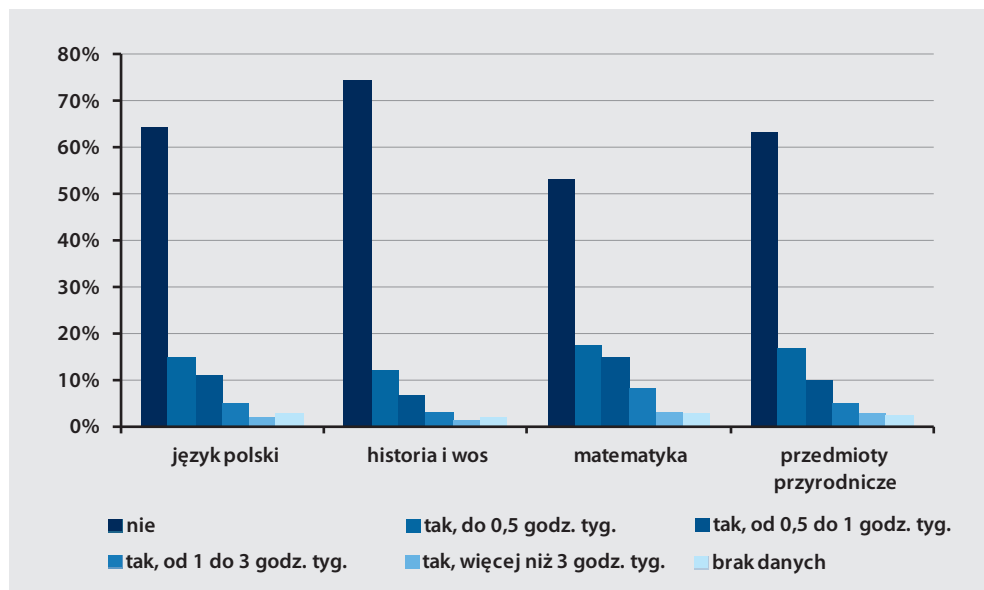


wszystkie przedmioty łącznie: matematyka, biologia, chemia, fizyka, geografia, język polski, wiedza o społeczeństwie, historia
przedmioty matematyczno-przyrodnicze: matematyka, biologia, chemia, fizyka, geografia
przedmioty humanistyczne: język polski, wiedza o społeczeństwie, historia

Prawie 28% uczniów zadeklarowało, że kiedykolwiek podczas nauki w gimnazjum uczęszczało na korepetycje (mogły to być korepetycje pobierane systematycznie lub od czasu do czasu). Tak więc znaczna część populacji uczniów pobierała naukę nie tylko w szkole. Co jednak ważne, o wiele częściej uczniowie otrzymywali dodatkowe wsparcie w postaci korepetycji z przedmiotów matematyczno-przyrodniczych (jedna piąta uczniów) niż z przedmiotów humanistycznych (jedna dwudziesta uczniów). Pokazuje to, że problem możliwego obciążenia wyników egzaminów przez fakt uczęszczania na korepetycje dotyczy w większym stopniu matematyki i przedmiotów przyrodniczych.

W przypadku nieodpłatnego wsparcia pytano uczniów o to, czy ktoś z rodziny, znajomych (lub innych osób) pomaga im bezpłatnie w nauce lub odrabianiu prac domowych z wyszczególnionych przedmiotów lub grup przedmiotów (przedmioty przyrodnicze zostały wyszczególnione w ankiecie jako: biologia, chemia, fizyka i geografia). Dodatkowo proszono ich o oszacowanie liczby godzin otrzymywanego wsparcia tygodniowo. Poniższy wykres prezentuje odsetek uczniów, którzy zaznaczyli poszczególne odpowiedzi dla kolejnych przedmiotów lub grup przedmiotów.

Rysunek 2. Odsetek uczniów deklarujących, ile średnio godzin tygodniowo otrzymuje bezpłatnej pomocy w nauce i odrabianiu lekcji od członków rodziny lub znajomych (lub innych osób)⁸



⁸ Dane zaprezentowane na wykresie pozwalają także zobaczyć, że kształt rozkładów tych zmiennych jest zbliżony do rozkładu logarytmiczno-normalnego. Aby wykorzystać je w dalszych analizach, mających na celu szacowanie związków liniowych, niezbędne będzie ich przekształcenie (zlogarytmizowanie).

Najczęściej uczniowie deklarowali, że otrzymują bezpłatne wsparcie z matematyki (łącznie 44% uczniów udzieliło odpowiedzi twierdzącej). 34,5% uczniów odpowiedziało, że otrzymuje wsparcie z przedmiotów przyrodniczych. Z języka polskiego – podobnie – 32,8% zadeklarowało, że otrzymuje bezpłatną pomoc z tego przedmiotu. Z historii i wiedzy o społeczeństwie uczniowie otrzymują najmniej wsparcia, prawdopodobnie dlatego, że najmniej go potrzebują (23,6% zadeklarowało, że otrzymuje co najmniej pół godziny tygodniowo takiej pomocy).

Przedstawione tu dane pokazują, że mimo iż nazwanie powszechnym zjawiska korzystania z niezależnego od szkoły wsparcia edukacyjnego byłoby przesadą, to jest to problem, którego nie wolno pominąć w rozważaniach poświęconych trafności metody EWD czy szerzej: uwarunkowań osiągnięć szkolnych. Uwidoczniała się też jedna interesująca prawidłowość. Uczniowie znacznie częściej otrzymują wsparcie (czy to bezpłatne czy w postaci płatnych korepetycji) z przedmiotów matematyczno-przyrodniczych (z przewagą matematyki) niż z przedmiotów humanistycznych.

6.3.2. Korelaty korzystania z korepetycji

W celu odpowiedzi na pytanie o to, co charakteryzuje uczniów, którzy uczęszczają na korepetycje lub otrzymują nieodpłatną pomoc w nauce, wyspecyfikowano szereg wielopoziomowych modeli regresji logistycznych. W pierwszej kolejności policzono tzw. modele puste, uwzględniające tylko pogrupowanie uczniów na szkoły i oddziały wewnętrznie szkół. Następnie dodawano krokowo do modeli kolejne predyktory jako efekty stałe, testując ich znaczenie dla prawdopodobieństwa otrzymywania wsparcia w nauce. W celu ułatwienia interpretacji wyników w niniejszym rozdziale zaprezentowano przede wszystkim modele puste, modele uwzględniające wszystkie zmienne, które w literaturze wskazywane są jako powiązane z prawdopodobieństwem otrzymywania pomocy w nauce oraz modele z wybranymi predyktorami, jeśli pokazywały istotne dla interpretacji różnice w uzyskanych efektach.

Jako pierwsze zostaną przedstawione wyniki modeli pokazujących korelaty uczęszczania na korepetycje z przedmiotów matematyczno-przyrodniczych. W drugiej kolejności – pokazujące korelaty otrzymywania bezpłatnej pomocy w nauce od rodziny, znajomych lub instytucji innych niż szkoła. W przypadku zagadnienia korelatów uczęszczania na korepetycje skupiono się przede wszystkim na przedmiotach matematyczno-przyrodniczych, gdyż odsetek uczniów pobierających korepetycje z przedmiotów humanistycznych jest tak mały, że można przyjąć, że ma niewielkie znaczenie dla problemu trafności wskaźników EWD. Ponadto z powodu małego odsetka uczniów uczęszczających na korepetycje z przedmiotów humanistycznych, oszacowanie modeli z tą zmienną wyjaśnianą okazało

się problematyczne. Część modeli miała problem z osiągnięciem zbieżności (model pusty, model z predyktorami z poziomu szkół). Ponadto w takiej sytuacji trudno o zaobserwowanie w modelu istotnych efektów. W aneksie przedstawiono model ze zmiennymi, które mimo wszystko okazały się mieć istotne znaczenie, by zainteresowany czytelnik mógł porównać uzyskane wyniki. Uczulamy jednak, że należy je interpretować w dużą ostrożnością.

W tabeli 1 przedstawiono wyniki analiz, pokazujące związek różnych zmiennych z prawdopodobieństwem uczęszczania na korepetycje z przedmiotów matematyczno-przyrodniczych. Zmienna zależna przyjmuje wartość 1, jeśli uczeń zadeklarował, że podczas nauki w gimnazjum uczęszcza lub uczęszczał na korepetycje z matematyki lub przedmiotów przyrodniczych, wartość 0 – jeśli nie uczęszczał.

Model (1) to model bez predyktorów, uwzględniający tylko pogrupowanie uczniów na szkoły i oddziały. Oszacowania efektów losowych pokazują, że istnieje istotne zróżnicowanie międzyszkolne prawdopodobieństwa uczęszczania na korepetycje. Innymi słowy w różnych szkołach różny odsetek uczniów pobiera dodatkowe płatne lekcje z przedmiotów matematyczno-przyrodniczych. Obserwujemy także istotną wariancję efektów na poziomie oddziałów, co oznacza, że mamy również do czynienia ze zróżnicowaniem w tym zakresie między oddziałami wewnątrz szkół. Wariancja tych efektów jest jednak niższa niż dla poziomu szkół. Wariancja na poziomie uczniów nie jest podana, ponieważ w przypadku modeli logistycznych, inaczej niż w przypadku regresji liniowych, wariancja wewnątrz grup jest z założenia różna (Snijders i Bosker, 2012, s. 291). Wynika to z faktu, że dla zmiennych dychotomicznych wariancja zmiennej wprost zależy od jej średniej⁹. Dlatego sposób wyznaczenia współczynnika korelacji wewnątrzgrupowej nie jest jednoznaczny, tak jak w przypadku modeli liniowych (Snijders i Bosker, 2012, s. 304–307).

W modelu (2) uwzględniono wszystkie zmienne niezależne, które w literaturze przedmiotu wskazywane są jako powiązane z prawdopodobieństwem uczęszczania na korepetycje. Zmienne opisujące ucznia uwzględniono na pierwszym poziomie analizy, natomiast zmienną zdającą sprawę z lokalizacji szkoły, na trzecim poziomie.

⁹ Przyjmując, że p to średnia zmiennej dychotomicznej, wariancja tej zmiennej jest równa $p(1 - p)$.

Tabela 1. Korelaty ucześnieczania na korepetycje z przedmiotów matematyczno-przyrodniczych. Wyniki trzydziomowych analiz regresji logistycznych, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: ucześnieczanie na korepetycje z przedmiotów matematyczno-przyrodniczych									
	(1)	(s.e.)	iloraz szans	(2)	(s.e.)	iloraz szans	(3)	(s.e.)	iloraz szans
oszacowanie efektów stałych									
poziom ucznia									
pleć ^a				0,631	(0,098)	1,888	0,632	(0,098)	1,882
sprawdzian				-0,294	(0,058)	0,745	-0,293	(0,058)	0,746
test matryc Ravena				-0,159	(0,059)	0,852	-0,158	(0,059)	0,854
bezradność intelektualna (mat)				0,294	(0,047)	1,342	0,294	(0,047)	1,342
motywacja niska ^b				-0,002	(0,207)	0,998	0,001	(0,208)	1,001
motywacja wysoka ^b				0,351	(0,124)	1,421	0,352	(0,124)	1,421
motywacja zewnętrzna ^b				0,252	(0,140)	1,287	0,251	(0,140)	1,285
motywacja wewnętrzna ^b				-0,134	(0,382)	0,875	-0,134	(0,153)	0,875
aspiracje rodziców				0,082	(0,024)	1,086	0,082	(0,024)	1,086
zasobność gosp. domowego				0,164	(0,055)	1,179	0,167	(0,055)	1,182
wykształcenie rodziców				0,056	(0,030)	1,057	0,063	(0,024)	1,065
HSIOPS				0,003	(0,006)	1,003			
stała	-1,50	(0,070)	0,223	-4,56	(0,456)	0,012	-4,40	(0,455)	0,012
poziom szkół									
lok.: miasto do 20 tys. ^c				0,457	(0,177)	1,580	0,456	(0,173)	1,578
lok.: miasto od 20 do 100 tys. ^c				0,607	(0,154)	1,834	0,609	(0,155)	1,838
lok.: miasto powyżej 100 tys. ^c				0,810	(0,178)	2,248	0,814	(0,178)	2,257
oszacowanie efektów losowych									
wariancja efektów szkół	0,216			0,135			0,135		
wariancja efektów oddziałów	0,130			0,055			0,056		

liczba uczniów: 4114, liczba oddziałów: 291, liczba szkół: 150

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie o profilu motywacji o przeciętnym nasileniu; c – grupa odniesienia: wieś

Wyniki przeprowadzonej analizy pokazały istnienie silnego efektu płci. Szansa na uczęszczanie na korepetycje z przedmiotów matematyczno-przyrodniczych jest 1,89 razy większa dla dziewczynek niż dla chłopców¹⁰. Zdolności i umiejętności ucznia są negatywnie związane z prawdopodobieństwem pobierania korepetycji. Uczniowie o wyższych uprzednich osiągnięciach (mierzonych wynikami na sprawdzianie w klasie szóstej szkoły podstawowej) rzadziej deklarowali chodzenie na korepetycje. Podobny (choć troszkę słabszy¹¹) związek obserwuje się dla inteligencji. Tak więc to uczniowie słabsi, którzy gorzej radzą sobie w szkole, częściej uczęszczają na korepetycje.

Ważny dla scharakteryzowania zjawiska jest także zaobserwowany istotny związek uczęszczania na korepetycje z bezradnością intelektualną mierzoną za pomocą skali bezradności intelektualnej dla matematyki (Sędek, 1995). Należy pamiętać, że bezradność intelektualna była mierzona na zakończenie nauki w gimnazjum, a pytania dotyczące korepetycji odnosiły się do całego okresu nauki w tej szkole. Z tego powodu nie można wykrytej zależności interpretować w kategoriach wpływu bezradności intelektualnej na uczęszczanie na korepetycje. Zmienna ta została jednak uwzględniona w omawianym modelu w celu ukazania związku korelacyjnego. Uczniowie, którzy korzystali z korepetycji podczas nauki w gimnazjum, wykazują się wyższym poziomem bezradności intelektualnej. Nie wiadomo jednak czy to uczniowie o wyższym poziomie bezradności intelektualnej częściej zdecydowali się skorzystać z korepetycji, czy też korepetycje tę bezradność nasiliły. Nie wiemy także, jakim poziomem bezradności charakteryzowałby się, gdyby z korepetycji nie skorzystali – być może jeszcze wyższym, jeżeli korepetycje pomogły ją złagodzić.

Z różnych wzorców motywacji, testowanych w modelu (2), tylko motywacja wysoka okazała się mieć istotny statystycznie efekt. Uczniowie, których na zakończenie nauki w klasie trzeciej charakteryzował ten wzorec motywacji, częściej niż uczniowie o motywacji o przeciętnej sile uczęszczali na korepetycje podczas nauki w gimnazjum. Efekty dla pozostałych typów motywacji nie są istotne statystycznie.

Dla grupy czynników związanych ze statusem społeczno-ekonomicznym rodziny pochodzenia ucznia wykryto zgodne z oczekiwaniami efekty. Im wyższe aspiracje rodziców (mierzone na początku nauki ucznia w gimnazjum, wyrażone na skali lat nauki potrzebnych do osiągnięcia pożądanego przez rodziców wykształcenia dziecka), tym większe prawdopodobieństwo tego, że uczeń będzie uczęszczał na korepetycje z przedmiotów matematyczno-przyrodniczych. Zasobność gospodarstwa domowego jest także pozytywnie związana z uczęszczaniem na korepetycje. Uczniowie pochodzący z bardziej zamożnych

¹⁰ Co ciekawe, dla przedmiotów humanistycznych obserwuje się odwrotny efekt dla płci (p. aneks do rozdziału).

¹¹ Zarówno wyniki sprawdzianu jak i testu matryc Ravena mierzone były na skali o średniej 0 i odchyleniu standardowym 1.

domów mają większą szansę na to, że będą pobierali korepetycje. Status społeczny rodziny mierzony wykształceniem rodziców lub wskaźnikiem HSIOPS jest także pozytywnie związany z prawdopodobieństwem posłania dziecka na korepetycje, ale zależność ta objawia się tylko wtedy, gdy zmienne te są włączane do modelu rozłącznie (model (3) i (4)). Jest to najwyraźniej skutkiem tego, że zależności te nie są silne.

Wielkość miejscowości, w której znajduje się szkoła, również ma znaczenie. Im większa miejscowość tym większe szanse na to, że uczniowie pobierają korepetycje. W zakresie przedmiotów matematyczno-przyrodniczych są one ponad dwukrotnie większe, jeżeli szkoła znajduje się w mieście liczącym powyżej 100 tysięcy mieszkańców niż wtedy, gdy szkoła znajduje się na terenie wiejskim (przy kontroli wszystkich pozostałych zmiennych uwzględnionych w modelu (2)).

6.3.3. Korelaty korzystania z bezpłatnej pomocy w nauce i odrabianiu lekcji

Kolejne tabele pokazują wyniki analiz korelatów otrzymywania bezpłatnej pomocy w nauce i odrabianiu lekcji od członków rodziny, znajomych lub instytucji innych niż szkoła. Tabela 2 podsumowuje wyniki analiz dla przedmiotów matematyczno-przyrodniczych, natomiast tabela 3 – dla przedmiotów humanistycznych. Przypomnijmy, że w odróżnieniu od korepetycji, pytaliśmy uczniów o otrzymywanie pomocy w nauce w momencie badania, czyli pod koniec trzeciej klasy gimnazjum.

Model (1) to tzw. model pusty uwzględniający tylko pogrupowanie uczniów w szkołach i oddziałach wewnątrz szkół. Oszacowania efektów losowych pokazują istnienie istotnego zróżnicowania międzyszkolnego oraz międzyoddziałowego w zakresie prawdopodobieństwa otrzymywania bezpłatnej pomocy w nauce w ramach przedmiotów matematyczno-przyrodniczych. Warto zwrócić uwagę na to, że – inaczej niż to miało miejsce w przypadku uczęszczania na korepetycje – wariancja efektów na poziomie oddziałów jest większa niż wariancja efektów na poziomie szkół. Oznacza to, że szansa na to, że uczeń otrzymuje bezpłatne wsparcie w nauce od rodziców lub znajomych w większym stopniu zależy od tego, do którego oddziału w szkole uczęszcza, niż od tego, do której szkoły chodzi.

Wyniki analiz uwzględniające predyktory (te same, co w przypadku badania korelatów uczęszczania na korepetycje) pokazały, że bezpłatną pomoc w nauce przedmiotów matematyczno-przyrodniczych otrzymują przede wszystkim uczniowie słabsi, o niższych uprzednich osiągnięciach i niższym poziomie inteligencji (patrz model (2)). Istotne efekty zaobserwowano także w przypadku trzech wzorców motywacji obserwowanych pod koniec nauki w gimnazjum. Motywacja wysoka jest pozytywnie związana z prawdopodobieństwem otrzymania pomocy w nauce, natomiast motywacja niska, a także motywacja zewnętrzna (choć w mniejszym stopniu) negatywnie z nim korelują.

Tabela 2. Korelatory korzystania z bezpłatnej pomocy w nauce i odrabiania lekcji z przedmiotów matematyczno-przyrodniczych. Wyniki trzypoziomowych analiz regresji logistycznych, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: korzystanie z bezpłatnej pomocy w nauce i odrabiania lekcji z przedmiotów matematyczno-przyrodniczych									
	(1)	(s.e.)	iloraz szans	(2)	(s.e.)	iloraz szans	(3)	(s.e.)	iloraz szans
oszacowanie efektów stałych									
poziom ucznia									
plec ^a				0,060 (0,087)		1,062	0,060 (0,087)		1,062
sprawdzian				-0,463 (0,065)		0,629	-0,460 (0,065)		0,631
test matryc Ravena				-0,208 (0,048)		0,812	-0,205 (0,049)		0,815
bezradność intelektualna (mat)				0,047 (0,044)		1,048	0,047 (0,044)		1,048
motywacja niska ^b				-0,818 (0,144)		0,441	-0,817 (0,144)		0,442
motywacja wysoka ^b				0,219 (0,104)		1,246	0,223 (0,104)		1,250
motywacja zewnętrzna ^b				-0,263 (0,117)		0,768	-0,265 (0,117)		0,767
motywacja wewnętrzna ^b				-0,047 (0,135)		0,954	-0,041 (0,135)		0,960
aspiracje rodziców				0,006 (0,018)		1,006	0,010 (0,018)		1,010
zasobność gosp. domowego				0,062 (0,044)		1,064			
wykształcenie rodziców				0,031 (0,020)		1,032	0,040 (0,020)		1,040
HSIOPS				0,013 (0,005)		1,032	0,013 (0,005)		1,013
stała	0,065 (0,045)		1,067	-1,029 (0,325)		0,358	-1,230 (0,311)		0,292
poziom szkół									
lok.: miasto do 20 tys. ^c				0,014 (0,112)		1,013	0,024 (0,111)		1,024
lok.: miasto od 20 do 100 tys. ^c				0,344 (0,097)		1,411	0,354 (0,096)		1,424
lok.: miasto powyżej 100 tys. ^c				0,255 (0,130)		1,291	0,267 (0,130)		1,306
oszacowanie efektów losowych									
wariancja efektów szkół	0,053			0,044			0,044		
wariancja efektów oddziałów	0,104			0,059			0,062		

liczba uczniów: 4186, liczba oddziałów: 291, liczba szkół: 150

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie o profilu motywacji o przeciętnym nasileniu; c – grupa odniesienia: wieś

Warto podkreślić, że bezradność intelektualna – inaczej niż w przypadku uczęszczania na korepetycje – nie jest istotnie związana z prawdopodobieństwem otrzymywania bezpłatnej pomocy w nauce. Jest to ważna różnica. Pozwala bowiem przypuszczać, że nie jest tak, że wsparcia w nauce szukają po prostu uczniowie bezradni intelektualnie (gdyby tak było, obserwowalibyśmy ujemne zależności dla obu rozpatrywanych form wsparcia). Większym nasileniem bezradności intelektualnej charakteryzują się tylko ci uczniowie, którzy uczęszczali na korepetycje podczas trzech lat nauki w gimnazjum.

Dla płci nie zaobserwowano istotnego efektu, co oznacza, że chłopcy i dziewczęta mają takie same prawdopodobieństwo uzyskiwania bezpłatnej pomocy w nauce przedmiotów matematyczno-przyrodniczych.

W przypadku statusu społeczno-ekonomicznego rodziny pochodzenia ucznia, można zauważyć różnice między korelatami otrzymywania korzystania z bezpłatnej pomocy w nauce przedmiotów matematyczno-przyrodniczych a korelatami uczęszczania na korepetycje. Zasobność gospodarstwa domowego nie ma dla otrzymywania bezpłatnego wsparcia w nauce znaczenia. Istotny i niewielki efekt ujawnia się w przypadku wykształcenia rodziców¹² (patrz model (3)), które razem ze wskaźnikiem HSIOPS wyjaśnia część wariancji zmiennej zależnej związanej ze statusem rodziny ucznia. Znaczenia nie mają też aspiracje rodziców.

Wielkość miejscowości, w której znajduje się szkoła, także ma pewne znaczenie dla otrzymywania bezpłatnej pomocy w nauce, nawet jeśli w modelu kontrolujemy status społeczno-ekonomiczny rodzin. Uczniowie z miast o średniej i dużej wielkości mają większą szansę na bezpłatne wsparcie w nauce (przy kontroli zmiennych opisujących status społeczno-ekonomiczny rodziny pochodzenia) niż uczniowie ze szkół wiejskich. Nie obserwuje się w tym względzie różnic między uczniami z małych miast (do 20 tysięcy mieszkańców) i ze wsi. Ponadto warto zauważyć, że znaczenie wielkości miejscowości jest mniejsze niż w przypadku pobierania korepetycji.

Model (4) prezentuje związki tylko z tymi predyktorami, które okazały się istotne po krokowym wyłączeniu z modelu (1) zmiennych niemających znaczenia.

W tabeli 3 zaprezentowano wyniki analogicznych do opisanych powyżej modeli dla otrzymywania bezpłatnej pomocy w nauce z przedmiotów humanistycznych. Większość wyników jest zbieżna z wynikami analiz korelatów korzystania ze wsparcia w zakresie przedmiotów matematyczno-przyrodniczych. Model (1) pokazuje istnienie istotnego zróżnicowania międzyшкольного i międzyoddziałowego ilości otrzymywanego bezpłatnego wsparcia w nauce przedmiotów humanistycznych. Troszkę większe zróżnicowanie jest na poziomie oddziałów

¹² Wyłączenie wykształcenia z modelu (2) nie zwiększyło znaczenia zasobności gospodarstwa domowego – pozostało ono nieistotne statystycznie.

niż szkół. Co ciekawe jednak, uwzględnieniu w modelu dodatkowych zmiennych (model (2)) sprawia, że szkoły i oddziały wewnątrz szkół nie różnią się istotnie między sobą w zakresie tego, jak duży odsetek uczniów otrzymuje bezpłatne wsparcie w nauce w tym zakresie. Pokazuje to, że dodatkowa pomoc w nauce z przedmiotów humanistycznych ma mniejsze znaczenie dla ewentualnego obciążenia wskaźników EWD faktem nieuwzględnienia tej informacji podczas ich wyliczania, gdyż szkoły w mniejszym stopniu różnią się między sobą w tym zakresie, a uwzględnienie w modelu uprzednich osiągnięć i płci pozwala znacząco zmniejszyć obserwowane zróżnicowanie¹³.

Znaczenie uprzednich osiągnięć mierzonych wynikami na sprawdzianie, wzorców motywacji, bezradności intelektualnej i wielkości miejscowości, w której znajduje się szkoła dla wyników egzaminacyjnych okazało się podobne do zaobserwowanego w modelu dla przedmiotów matematyczno-przyrodniczych. Nie będziemy zatem powtarzać interpretacji tych wyników.

Zwrócimy natomiast uwagę na różnice. Dla korzystania z pomocy z przedmiotów humanistycznych nie zanotowano związku z inteligencją. Być może wynika to z faktu, że osiągnięcia humanistyczne (a więc także trudności w nauce), są zazwyczaj w mniejszym stopniu powiązane z inteligencją niż osiągnięcia z przedmiotów ścisłych (Ferrer i McArdle, 2004; Heaven i Ciarrochi, 2012; Teo, Carlson, Mathieu, Egeland i Sroufe, 1996). Zaobserwowano istotny efekt płci – istnieje mniejsze prawdopodobieństwo korzystania z bezpłatnego wsparcia z przedmiotów humanistycznych przez dziewczynki niż przez chłopców. W przypadku czynników statusowych, zasobność gospodarstwa domowego podobnie nie ma większego znaczenia.

Natomiast ujawnia się niewielki efekt dla wskaźnika HSIOPS (istotny dopiero po wyłączeniu z modelu zasobności gospodarstwa domowego¹⁴). Zaskakujący i trudny w interpretacji jest istotny i negatywny efekt dla aspiracji rodziców. Mówi on o tym, że im wyższe aspiracje rodziców (mierzone na początku nauki w gimnazjum), tym mniejsza szansa na to, że rodzice (lub członkowie rodziny i znajomi) będą udzielać dzieciom pomocy w nauce w zakresie przedmiotów humanistycznych. Być może rodzice o wyższych aspiracjach częściej szukają dla uczniów innej formy wsparcia niż własna pomoc. Efekt ten jednak nie jest duży i nie potwierdził się w przypadku przedmiotów matematyczno-przyrodniczych, więc sugerujemy nie nadawać mu dużego znaczenia.

¹³ Kontrola w modelu wyniku sprawdzianu i płci ucznia zmniejsza wariancję międzyszkolną prawdopodobieństwa otrzymywania bezpłatnego wsparcia z przedmiotów humanistycznych do tego stopnia (0,037), że staje się ona nieistotna statystycznie na poziomie istotności $p < 0,05$. Wyniki analiz do wglądu u autorek rozdziału.

¹⁴ Wyłączenie z modelu (2) wykształcenia rodziców nie zwiększyło znaczenia zasobności gospodarstwa domowego – pozostało ono nieistotne statystycznie.

Tabela 3. Korelaty korzystania z bezpłatnej pomocy w nauce i odrabiania lekcji z przedmiotów humanistycznych. Wyniki trzypoziomowych analiz regresji logistycznych, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: korzystania z bezpłatnej pomocy w nauce i odrabiania lekcji z przedmiotów humanistycznych									
	(1)	(s.e.)	iloraz szans	(2)	(s.e.)	iloraz szans	(3)	(s.e.)	iloraz szans
oszacowanie efektów stałych									
poziom ucznia									
pleć ^a				-0,334	(0,090)	0,716	-0,334	(0,090)	0,716
sprawdzian				-0,477	(0,061)	0,621	-0,475	(0,061)	0,622
test matryc Ravena				-0,087	(0,044)	0,917	-0,084	(0,048)	0,919
bezradność intelektualna (mat)				0,007	(0,043)	1,008	0,006	(0,043)	1,006
motywacja niska ^b				-0,702	(0,154)	0,496	-0,701	(0,155)	0,496
motywacja wysoka ^b				0,261	(0,099)	1,298	0,262	(0,099)	1,300
motywacja zewnętrzna ^b				-0,332	(0,135)	0,717	-0,333	(0,135)	0,717
motywacja wewnętrzna ^b				-0,076	(0,122)	0,927	-0,073	(0,123)	0,930
aspiracje rodziców				-0,051	(0,020)	0,950	-0,048	(0,021)	0,953
zasobność gosp. domowego				0,044	(0,047)	1,043	0,010	(0,019)	
wykształcenie rodziców				0,004	(0,020)	1,004			
HSIOPS				0,014	(0,004)	1,014	0,014	(0,004)	1,014
stała	-0,504	(0,045)	0,604	-0,291	(0,295)	0,748	-0,427	(0,297)	0,652
poziom szkół									
lok.: miasto do 20 tys. ^c				0,036	(0,110)	1,036	0,044	(0,110)	1,045
lok.: miasto od 20 do 100 tys. ^c				0,244	(0,115)	1,276	0,251	(0,114)	1,251
lok.: miasto powyżej 100 tys. ^c				0,359	(0,124)	1,432	0,367	(0,124)	1,367
oszacowanie efektów losowych									
wariancja efektów szkół	0,058			0,031			0,030		
wariancja efektów oddziałów	0,067			0,030			0,033		

liczba uczniów: 4186, liczba oddziałów: 291, liczba szkół: 150
 pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie o przeciętnym nasileniu; c – grupa odniesienia: wieś

Podsumowując analizy mające na celu scharakteryzowanie zjawiska pobierania korepetycji i otrzymywania bezpłatnej pomocy w nauce, wypunktujemy najważniejsze wnioski. Przede wszystkim należy zwrócić uwagę na to, że dla trafności szacowania wskaźników EWD większe znaczenie może mieć dodatkowe wsparcie z przedmiotów matematyczno-przyrodniczych niż z przedmiotów humanistycznych. Po pierwsze wynika to z faktu, że zjawisko otrzymywania pomocy w nauce jest częstsze dla tych pierwszych niż drugich (w przypadku korepetycji z przedmiotów humanistycznych zjawisko można nazwać marginalnym). Po drugie, także zróżnicowanie międzyszkolne prawdopodobieństwa otrzymywania takiego wsparcia jest większe dla przedmiotów matematyczno-przyrodniczych.

Z dodatkowego wsparcia edukacyjnego (zarówno w postaci korepetycji, jak i bezpłatnej pomocy w nauce) korzystają przede wszystkim uczniowie słabsi, o niższych uprzednich osiągnięciach i niższym poziomie inteligencji¹⁵. Z drugiej jednak strony, zauważono słaby lecz pozytywny związek ze statusem społecznym rodziny, który oznacza, że częściej dodatkowe wsparcie otrzymują uczniowie z rodzin o wyższym statusie. Status ekonomiczny (mierzony wskaźnikiem zasobności gospodarstwa domowego) był pozytywnie związany tylko z prawdopodobieństwem uczęszczania na korepetycje. Dla bezpłatnej pomocy w nauce nie zano-towano znaczącego związku. Wyniki te pokazują, że z jednej strony wyniki egzaminacyjne szkół pracujących w bardziej uprzywilejowanym środowisku (zatem pracujących z uczniami z rodzin o wyższym statusie społeczno-ekonomicznym) mogą być bardziej obciążone otrzymywaniem przez uczniów edukacji pozaszkolnej, z drugiej jednak strony to wśród uczniów bardziej narażonych na problemy w nauce (niskie uprzednie osiągnięcia, niższa inteligencja), zjawisko otrzymywania wsparcia w nauce jest bardziej powszechne. Tak więc wyniki egzaminacyjne szkół pracujących z młodzieżą z trudnościami w uczeniu się także mogą być w większym stopniu narażone na obciążenie przez wpływ tego, czego nie wypracowała szkoła.

Analiza związku otrzymywania wsparcia edukacyjnego z bezradnością intelektualną pokazała ciekawe różnice. Uczniowie, którzy uczęszczali podczas nauki w gimnazjum na korepetycje, charakteryzowali się wyższym poziomem bezradności intelektualnej pod koniec nauki w szkole (lub inaczej: uczniowie, którzy mieli wyższy poziom bezradności intelektualnej na zakończenie nauki w gimnazjum, częściej deklarowali, że uczęszczali na korepetycje podczas nauki w tej szkole). Nie zaobserwowano jednak istotnego związku bezradności intelektualnej z otrzymywaniem bezpłatnej pomocy w nauce. Trudno orzekać w tym przypadku, co jest przyczyną, a co skutkiem (a być może jest jeszcze inna zmienna, która tłumaczy oba zjawiska i zaobserwowane różnice). Jednak zasadnicza różnica polega na tym, że możemy

¹⁵ Tylko dla otrzymywania bezpłatnej pomocy w nauce z przedmiotów humanistycznych poziom inteligencji nie miał znaczenia.

powiedzieć, że uczniów, którzy pobierali korepetycje od uczniów, którzy ich nie pobierali różni poziom bezradności intelektualnej, a różnic takich nie stwierdzamy, porównując tych, którzy otrzymują bezpłatną pomoc w nauce i tych, którzy jej nie otrzymują. Być może uczęszczanie na korepetycje pogłębia (a w każdym razie nie przeciwdziała) bezradność intelektualną. A może na korepetycje są posyłani ci uczniowie, którzy uprzednio osiągnęli już wyższy poziom bezradności.

Warto także podkreślić rolę wielkości miejscowości, w której znajduje się szkoła. Im większa miejscowość, tym większa szansa na to, że uczeń będzie otrzymywał wsparcie czy to w postaci korepetycji, czy bezpłatnej pomocy w nauce (związek ten jest widoczny także przy kontroli zmiennych opisujących status społeczno-ekonomiczny rodziny pochodzenia). W przypadku korepetycji związek jest silniejszy i bardziej klarowny. Najprawdopodobniej dlatego, że istotną rolę odgrywa tu dostępność korepetycji, która wzrasta wraz ze wzrostem wielkości miejscowości (większa liczba korepetytorów, łatwiejszy dojazd itp.). Związek dla bezpłatnego wsparcia w nauce, słabszy niż dla korepetycji, objawia się istotnie większym nasileniem zjawiska w miastach średnich i dużych w stosunku do szkół wiejskich. Pokazuje to, że nie tylko dostępność pomocy ma znaczenie (bezpłatna pomoc udzielana przez członków rodziny jest technicznie rzecz biorąc tak samo dostępna, niezależnie od wielkości miejscowości). Dla nas jednak największe znaczenie ma fakt, że szkoły z większych miejscowości są bardziej narażone na obciążenie wyników egzaminacyjnych przez pozaszkolne wsparcie w nauce.

6.3.4. Korepetycje a wyniki egzaminacyjne i EWD

Omówienie wyników analiz związku pozaszkolnego wsparcia edukacyjnego z wynikami egzaminacyjnymi i jego znaczenia dla wskaźników EWD rozpoczniemy od problemu korepetycji. W pierwszej kolejności pokażemy związek uczęszczania na korepetycje z wynikami egzaminacyjnymi. W drugim kroku pokażemy, jak zmienia się siła związku, jeśli do modelu włączymy zmienne uwzględnione w modelach EWD. Następnie, aby lepiej zrozumieć badane zjawisko, przeanalizujemy ten związek przy kontroli innych zmiennych mogących mieć znaczenie zarówno dla wyników egzaminacyjnych jak i pobierania korepetycji. W rozdziale zaprezentujemy tylko wyniki dla przedmiotów matematyczno-przyrodniczych, jako że zjawisko uczęszczania na korepetycje z przedmiotów humanistycznych jest marginalne (patrz pierwsza część wyników).

W tabeli 4 przedstawiono wyniki analiz dla sześciu modeli, w których zmienną zależną są wyniki egzaminu gimnazjalnego z przedmiotów matematyczno-przyrodniczych. Pierwszy model, bez dodatkowych zmiennych niezależnych, uwzględnia tylko pogrupowanie uczniów na szkoły i oddziały i stanowi punkt odniesienia dla kolejnych. Model (2) uwzględnia tylko jeden

predyktor: zmienną niezależną zdającą sprawę z tego, czy uczeń podczas nauki w gimnazjum uczęszczał na korepetycje z co najmniej jednego przedmiotu matematyczno-przyrodniczego. Ponieważ kategorią odniesienia są osoby, które nie pobierały korepetycji, ujemny i istotny efekt oznacza, że uczniowie, którzy korzystali z tego typu wsparcia, wypadli gorzej na egzaminie gimnazjalnym niż ci, którzy takiej pomocy nie otrzymywali. Zaobserwowana różnica jest rzędu blisko jednej trzeciej odchylenia standardowego wyników. Oznaczać to może, że albo uczęszczanie na korepetycje przyczyniło się do uzyskania gorszych rezultatów, albo że korepetycje były reakcją na problemy w nauce, które koniec końców także przełożyły się na niższe wyniki egzaminacyjne. Być może też oba mechanizmy mają (niekoniecznie równie silne) znaczenie. Kolejne modele mają na celu przybliżenie zrozumienia znaczenia korepetycji dla wyników egzaminacyjnych. Szczególnie istotne wydaje się wytrącenie wpływu uprzednich osiągnięć, co w pewnym stopniu pozwoli kontrolować znaczenie problemów w nauce.

Model (3) uwzględnia dodatkowe zmienne kontrolowane – te, które są brane pod uwagę podczas wyliczania wskaźników EWD. Najważniejszą spośród nich jest wynik na sprawdzianie, będący wskaźnikiem uprzednich osiągnięć szkolnych. Włączenie ich do modelu spowodowało znaczący spadek siły efektu korepetycji. Jest to wynikiem pokazanej już wcześniej przez nas zależności (patrz tabela 2) mówiącej o tym, że na korepetycje uczęszczają częściej uczniowie słabsi, o niższych wynikach na sprawdzianie. Dlatego kontrola w modelu uprzednich osiągnięć przyczyniła się do zmniejszenia siły efektu korepetycji. Efekt ten jest jednak nadal istotny i ujemny. Oznacza to, że jeśli porównamy ze sobą uczniów o tych samych uprzednich osiągnięciach, ci którzy uczęszczali na korepetycje podczas nauki w gimnazjum osiągnęli średnio o 1,9 punktu mniej na egzaminie gimnazjalnym (na skali o odchyleniu standardowym 15) niż uczniowie, którzy nie korzystali z pomocy korepetytorów. Interpretacje tego efektu pogłębimy, analizując wyniki modeli z większą liczbą zmiennych kontrolowanych.

Zanim do nich przejdziemy, skoncentrujemy się jeszcze na innym ujęciu w modelu zjawiska korepetycji, gdyż być może inny wskaźnik dostarczyłby nam odmiennych interpretacji. Znaczenie korepetycji dla wyników egzaminacyjnych można rozpatrywać też na wyższych poziomach analizy. Możemy zapytać o to, czy oddziały, w których więcej uczniów pobiera korepetycje, mają średnio wyższe wyniki na egzaminie. Inaczej formułując ten problem, moglibyśmy zapytać, czy klasy, w których zjawisko korepetycji jest bardziej popularne, mają wyższe osiągnięcia. Model (4) ma na celu udzielenie odpowiedzi na to pytanie. W modelu tym dodano na poziomie klasy wskaźnik powszechności korepetycji z przedmiotów matematyczno-przyrodniczych, okazał się on jednak nieistotny statystycznie. Również kontrola uprzednich osiągnięć oraz pozostałych zmiennych uwzględnianych w modelach EWD nie przynosi zmian co do siły i istotności tego efektu (patrz model (5)). Model (6) uwzględniał dwa wskaźniki równocześnie, dając tym samym możliwość oszacowania znaczenia tych efektów przy wzajemnej kontroli. Kontrola popularności korepetycji w klasie

Tabela 4. Znaczenie uczęszczania na korepetycje dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej i EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej												
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)
oszacowanie efektów stałych												
poziom ucznia												
korepetycje z mat-przyr. ^a			-4,73	(0,526)	-1,90	(0,397)			11,98	(0,288)	-1,93	(0,419)
sprawdzian					11,87	(0,284)			1,25	(0,122)	11,87	(0,285)
sprawdzian ²					1,22	(0,122)			-0,394	(0,070)	1,22	(0,122)
sprawdzian ³					-0,385	(0,069)			-0,909	(0,305)	-0,385	(0,069)
plec ^b					-0,732	(0,310)			2,03	(0,901)	-0,731	(0,310)
dysleksja na egz. gimnazjalnym ^c					2,14	(0,912)			-3,98	(0,946)	2,13	(0,912)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym					-3,96	(0,955)			100,38	(0,573)	-3,96	(0,956)
stała	100,22	(0,512)	101,12	(0,550)	100,41	(0,366)	100,66	(0,806)	100,31	(0,573)	100,31	(0,573)
poziom oddziałów												
% chodzących na korepetycje							-2,23	(3,44)	-1,34	(2,01)	0,546	(1,90)
oszacowanie efektów losowych												
wariancja efektów szkół	20,75		22,74		4,65		21,43		4,57		4,62	
wariancja efektów oddziałów	18,90		17,76		5,56		18,29		5,55		5,58	
wariancja na poziomie ucznia	159,66		156,35		66,04		159,66		66,57		66,04	
podsumowanie												
deviance	32949,20		32868,82		29229,43		32948,66		29259,43		29229,32	
liczba szacowanych parametrów	4		5		11		5		11		12	

liczba uczniów: 4113, liczba klas: 291, liczba szkół: 150
 pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: uczniowie nie uczęszczający na korepetycje; b – grupa odniesienia: chłopcy; c – grupa odniesienia: uczniowie bez dysleksji

nie zmieniała siły efektu indywidualnego. Dla wskaźnika z poziomu klasy można zauważyć odwrócony znak zależności, jednak efekt nadal nie jest istotny statystycznie, dlatego zmiany tej nie będziemy rozważać.

W kolejnej tabeli przedstawiono wyniki analiz mających na celu ukazanie związku uczęszczania na korepetycje z wynikami egzaminu gimnazjalnego po wytrąceniu znaczenia kolejnych zmiennych niezależnych mających znaczenie dla osiągnięć szkolnych i faktu uczęszczania na korepetycje. W celu zbudowania tych modeli, krokowo włączano kolejne zmienne niezależne. W ostatnim kroku zbadano hipotezę o zróżnicowaniu siły związku korepetycji z wynikami w zależności od oddziały klasowego.

Włączenie do modelu poziomu inteligencji spowodowało spadek siły efektu korepetycji dla wyników egzaminu (patrz model (7)). Jest to wynikiem tego, że uczęszczanie na korepetycje jest negatywnie związane z inteligencją (patrz tabela 1) i nieuwzględnienie jej w modelu jako zmiennej kontrolowanej powodowało, że efekt korepetycji ukazywał częściowo znaczenie inteligencji. Jednak, nawet po wytrąceniu znaczenia uprzednich osiągnięć i inteligencji, obserwujemy istotny i negatywny związek uczęszczania na korepetycje z wynikami testów.

Dodanie do modelu poziomu bezradności intelektualnej (negatywnie związanej z wynikami egzaminu, a pozytywnie związanej z uczęszczaniem na korepetycje) powoduje kolejny spadek siły efektu korepetycji (model (8)). Co ciekawe, kontrola w modelu wzorca motywacji (model (9)) nie zmienia siły efektu korepetycji. Być może dzieje się tak dlatego, że inny typ motywacji jest związany z uczęszczaniem na korepetycje (gdzie tylko motywacja wysoka była istotnie i pozytywnie związana z prawdopodobieństwem chodzenia na korepetycje), a inny z wynikami egzaminu (umiarkowanie silny i pozytywny związek z wzorcem motywacji wewnętrznej).

Warte podkreślenia jest to, że kontrola w modelu zmiennych opisujących status społeczno-ekonomiczny rodziny ucznia¹⁶ powoduje, że efekt korepetycji staje się silniejszy, co sugeruje istnienie efektów supresyjnych – szczególnego rodzaju efektu pośredniczonego (zob. Cichocka i Milewicz, 2010). Zaobserwowany tu efekt wzrostu siły znaczenia korepetycji po dodaniu predyktorów w postaci wskaźników SES zachodzi wówczas, gdy predyktory są odmiennie powiązane ze zmienną zależną (wyniki egzaminu są negatywnie związane z korepetycjami a pozytywnie z SES). Prezentowane wcześniej wyniki pokazujące, że uczęszczanie na korepetycje jest pozytywnie związane z SES rodziny (patrz tabela 1) pozwala zrozumieć,

¹⁶ W modelu uwzględniono wszystkie trzy wskaźniki SES: wykształcenie rodziców, zasobność gospodarstwa domowego oraz HSIOPS dlatego, że każdy z nich oddzielnie włączony do modelu ma istotny wpływ na wyniki egzaminu, mimo że po włączeniu trzech, tylko dla HSIOPS obserwujemy istotny efekt. Ponadto w ten sposób dokładniej wytrącamy tę zmienność wyników, która jest związana ze statusem rodziny ucznia.

Tabela 5. Znaczenie uczęszczania na korepetycje dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej z uwzględnieniem dodatkowych zmiennych kontrolowanych. Wyniki trzypoziomowych analiz regresji, modele z losową stałą (7–11) i losowym nachyleniem (12), odporne (*robust*) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej												
	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)
oszacowanie efektów stałych												
poziom ucznia												
korepetycje z mat.-przyr. ^a	-1,57	(0,363)	-1,16	(0,353)	-1,15	(0,356)	-1,58	(0,355)	-1,53	(0,354)	-1,54	(0,352)
sprawdzian	9,72	(0,295)	9,27	(0,303)	9,24	(0,300)	8,48	(0,230)	8,50	(0,330)	8,51	(0,295)
sprawdzian ²	1,22	(0,123)	1,12	(0,123)	1,13	(0,121)	1,18	(0,116)	1,18	(0,117)	1,18	(0,117)
sprawdzian ³	-0,365	(0,073)	-0,334	(0,076)	-0,330	(0,075)	-0,300	(0,071)	-0,300	(0,071)	-0,300	(0,071)
pleć ^b	-1,33	(0,282)	-1,39	(0,276)	-1,67	(0,292)	-1,84	(0,286)	-1,84	(0,286)	-1,86	(0,288)
dysleksja na egz. gimnazjalnym ^c	1,59	(0,856)	1,45	(0,869)	1,48	(0,857)	1,16	(0,873)	1,16	(0,866)	1,17	(0,878)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym	-3,30	(0,907)	-3,14	(0,918)	-3,20	(0,897)	-2,81	(0,910)	-2,75	(0,898)	-2,83	(0,919)
test matryc Ravena (I pomiar)	3,91	(0,209)	3,58	(0,202)	3,59	(0,199)	3,42	(0,202)	3,42	(0,201)	3,39	(0,150)
bezdarność intelektualna (mat)			-1,85	(0,150)	-1,67	(0,154)	-1,61	(0,150)	-1,61	(0,151)	-1,59	(0,150)
motywacja niska ^d					-0,294	(0,520)	-0,261	(0,492)	-0,216	(0,491)	-0,270	(0,492)
motywacja wysoka ^d					0,610	(0,372)	0,451	(0,359)	0,441	(0,357)	0,478	(0,362)
motywacja zewnętrzna ^d					-0,388	(0,394)	-0,381	(0,385)	-0,365	(0,385)	-0,344	(0,385)
motywacja wewnętrzna ^d					1,93	(0,447)	1,71	(0,439)	1,70	(0,441)	1,73	(0,442)
aspiracje rodziców							0,497	(0,068)	0,501	(0,068)	0,500	(0,068)
zasobność gosp. domowego							0,166	(0,175)	0,187	(0,174)	0,165	(0,175)
wykształcenie rodziców							0,085	(0,082)	0,098	(0,082)	0,091	(0,082)
HSIOPS							0,052	(0,017)	0,053	(0,016)	0,052	(0,017)
stała	100,46	0,353)	100,46	0,358)	100,28	0,393)	89,32	1,28)	89,73	1,31)	89,21	1,27

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej												
	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)
poziom szkół												
lok.: miasto do 20 tys. ^e									-0,187	(0,616)		
lok.: miasto od 20 do 100 tys. ^e									-1,94	(0,564)		
lok.: miasto powyżej 100 tys. ^e									-1,19	(0,737)		
oszacowanie efektów losowych												
wariancja efektów szkół	4,62		4,60		4,55		4,35		3,89		3,68	
wariancja efektów oddziałów	3,82		3,83		3,77		3,33		3,28		4,98	
wariancja na poziomie ucznia	56,98		56,22		55,79		54,22		54,22		53,87	
poziom oddziałów												
losowe nachylenie dla: korepetycje z mat-przyr.											2,55	
podsumowanie												
deviance	28744,427		28556,02		28523,39		28395,12		28385,36		28383,20	
liczba szacowanych parametrów	12		13		17		21		24		23	

liczba uczniów: 4113, liczba klas: 291, liczba szkół: 150
pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: uczniowie nie uczęszczający na korepetycje; b – grupa odniesienia: chłopcy; c – grupa odniesienia: uczniowie bez dysleksji; d – grupa odniesienia: uczniowie o profilu motywacji o przeciętnym nasileniu; e – grupa odniesienia: wieś

dłaczego, jeśli nie kontrolujemy SES rodziny ucznia, efekt korepetycji pozostaje niedoszacowany (zniekształcony przez nieuwzględnione w modelu związki).

W modelu (11) wytrącono znaczenie wielkości miejscowości, w której znajduje się szkoła (predyktor na trzecim poziomie analizy), jako że zmienna ta miała znaczenie dla pobierania korepetycji. Okazało się jednak, że nie zmieniło to znacząco siły efektu korepetycji.

Ostatni (12) model testował, czy związek uczęszczania na korepetycje z wynikami egzaminu gimnazjalnego jest różny w różnych klasach. W modelu tym uwolniono losowe nachylenie na poziomie oddziałów dla (niecentrowanego) wskaźnika pobierania korepetycji. Efekt losowy okazał się nieistotnie statystycznie różny od zera, tak więc nie można stwierdzić, że siła związku uczęszczania na korepetycje z wynikami egzaminu jest różna w różnych oddziałach.

Analizy dotyczące znaczenia uczęszczania na korepetycje dla wyników egzaminacyjnych pokazały istnienie negatywnego związku między korzystaniem z tego typu wsparcia a wynikami egzaminacyjnymi. Uczniowie, którzy zadeklarowali uczęszczanie na korepetycje, uzyskali niższe wyniki na egzaminie gimnazjalnym. Zależność ta jest znacznie słabsza, choć nadal istotna i negatywna, w modelach EWD, kiedy to kontrolujemy uprzednie osiągnięcia szkolne. Kontrola pozostałych zmiennych mających znaczenie zarówno dla prawdopodobieństwa pobierania korepetycji, jak i uzyskiwanych na egzaminie rezultatów, nie zmieniła stwierdzonego kierunku zależności. Oznacza to, że jeśli porównamy ze sobą uczniów o takiej samej inteligencji, płci, poziomie bezradności intelektualnej, motywacji, statusie rodziny pochodzenia, mieszkających w miejscowości o podobnej wielkości i o takich samych uprzednich osiągnięciach, to uczniowie, którzy chodzili na korepetycje podczas nauki w gimnazjum otrzymają o około 1,5 punktu (na skali o odchyleniu standardowym 15) mniej na egzaminie gimnazjalnym niż ci, którzy nie korzystali z takiego wsparcia. Czy oznacza to, że uczęszczanie na korepetycje im zaszkodziło? Zebrane dane nie pozwalają na jednoznaczną odpowiedź na tak postawione pytanie. Możemy jedynie stwierdzić, że nawet jeśli uczniowie ci rozwinęli swoje umiejętności na korepetycjach, to nie na tyle, by dorównać swoim kolegom o takich samych cechach, którzy jednak na korepetycje nie uczęszczali. Na pewno jednak ich nie prześcignęli.

6.3.5. Bezpłatne wsparcie w nauce a wyniki egzaminacyjne i EWD

Druga część analiz koncentruje się na problemie znaczenia otrzymywania bezpłatnego wsparcia w nauce i odrabianiu lekcji dla wyników egzaminacyjnych i EWD. Prezentację i omówienie wyników zaczniemy od przedmiotów matematyczno-przyrodniczych. W pierwszej kolejności pokażemy związek otrzymywania bezpłatnej pomocy w nauce z wynikami egzaminacyjnymi. Następnie pokażemy, jak zmienia się siła związku, jeśli do modelu włączymy zmienne uwzględnione w modelach EWD. W kolejnych krokach analiza problemu zostanie pogłębiona przez pokazanie siły tego

związku, jeśli w modelu kontrolowane są inne zmienne mogące mieć znaczenie zarówno dla wyników egzaminacyjnych, jak i otrzymywania bezpłatnej pomocy w nauce.

Przypomnijmy, że w analizach tych zmiennymi zależnymi są wskaźniki powstałe w wyniku zlogarytmizowania sumy liczby godzin uzyskiwanego wsparcia z określonej grupy przedmiotów. Dla tak przekształconej zmiennej łatwiej w tym przypadku modelować związki, natomiast interpretacja współczynników regresji, ze względu na nieintuicyjność jednostki skali, staje się problematyczna. Dla postawionych pytań badawczych najważniejsze są jednak relacje pomiędzy siłą związku wyliczoną na podstawie różnych modeli, a to daje się odczytać z modeli z tak zbudowanym wskaźnikiem.

W tabeli 6 zaprezentowano wyniki analiz, w których zmienną zależną jest łączny wynik egzaminu gimnazjalnego z przedmiotów matematyczno-przyrodniczych, wyrażony na skali o średniej 100 i odchyleniu standardowym 15 w populacji uczniów, którzy przystąpili do egzaminu. Analizy te miały przede wszystkim na celu znalezienie odpowiedzi na pytanie o to, jak kontrola w modelach EWD zmiennych charakteryzujących ucznia pozwala zmodyfikować zależność wyników egzaminacyjnych od ilości otrzymywanego bezpłatnego wsparcia w nauce. Model (1), bez dodatkowych zmiennych wyjaśniających, stanowi punkt odniesienia dla kolejnych modeli.

Model (2), w którym uwzględniono tylko jedną zmienną niezależną – ilość otrzymywanego bezpłatnego wsparcia, pokazuje istnienie ujemnej, statystycznie istotnej zależności: im więcej uczeń otrzymał bezpłatnej pomocy w nauce i odrabianiu lekcji od rodziny i znajomych, tym niższe uzyskał rezultaty na egzaminie gimnazjalnym. Nie interpretujemy jednak tego związku w kategoriach przyczynowo-skutkowych. Wcześniejsze analizy pokazały, że pomoc w nauce otrzymują przede wszystkim te osoby, które jej potrzebują – czyli uczniowie słabsi. Dlatego kontrola w modelu uprzednich osiągnięć powinna osłabić związek wyników egzaminacyjnych z ilością otrzymywanego wsparcia. I tak się faktycznie dzieje, co pokazują wyniki modelu (3), w którym dołączono te zmienne, które są brane pod uwagę w modelach EWD. Widzimy znaczący spadek siły zależności. Nadal jednak obserwujemy istotny i negatywny efekt otrzymywania bezpłatnej pomocy w nauce. Oznacza to, że jeśli porównujemy ze sobą uczniów o takich samych uprzednich wynikach na sprawdzianie (płci i statusu dotyczącego dysleksji), uczniowie, którzy korzystali z dodatkowego bezpłatnego wsparcia, uzyskali średnio niższe wyniki na egzaminie gimnazjalnym.

Problem ten możemy też analizować z innej perspektywy, pytając czy oddziały, w których uczniowie otrzymują średnio więcej bezpłatnej pomocy w nauce, uzyskują lepsze rezultaty na egzaminach. Model (4) odpowiada na tak postawione pytanie, jeśli nie kontrolujemy uprzednich osiągnięć. W modelu tym uwzględniono jeden predyktor na poziomie klasy: średnią oddziałową ze zlogarytmizowanych indywidualnych wskaźników ilości otrzymywanego

Tabela 6. Znaczenie bezpłatnej pomocy w nauce dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej i EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej												
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)
Oszacowanie efektów stałych												
poziom uczenia												
bezpłatne wsparcie z mat-przyr.			-2,23	(0,184)	-0,650	(0,117)					-0,606	(0,121)
sprawdzian					11,71	(0,278)			11,93	(0,278)	11,70	(0,279)
sprawdzian ²					1,24	(0,120)			1,23	(0,120)	1,24	(0,120)
sprawdzian ³					-0,385	(0,069)			-0,407	(0,068)	-0,388	(0,069)
plec ^a					-0,927	(0,298)			-0,913	(0,301)	-0,912	(0,298)
dysleksja na egz. gimnazjalnym ^b					2,19	(0,879)			2,09	(0,878)	2,18	(0,876)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym					-4,01	(0,930)			-4,00	(0,935)	-3,97	(0,927)
stała	100,24	(0,515)	98,04	(0,482)	99,46	(0,355)	94,68	(0,810)	98,43	(0,608)	98,32	(0,604)
poziom oddziałów												
średnia z bezpłatnego wsparcia z mat-przyr.							-5,75	(0,791)	-1,74	(0,450)	-1,20	(0,519)
oszacowanie efektów losowych												
wariancja efektów szkół	20,85		19,63		4,42		18,95		4,27		4,32	
wariancja efektów oddziałów	19,26		16,92		5,45		14,11		5,15		5,32	
wariancja na poziomie ucznia	159,66		151,59		65,82		159,61		66,48		65,80	
podsumowanie												
deviance	33534,56		33307,15		29727,45		33488,06		29758,19		29721,50	
liczba szacowanych parametrów	4		5		11		5		11		12	

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji;

wsparcia (trzeci typ wskaźnika korzystania z bezpłatnego wsparcia w nauce). Wyniki wskazują na to, że oddziały, w których jest bardziej nasilone zjawisko otrzymywania przez uczniów bezpłatnej pomocy w nauce, mają niższe wyniki. Znow jednak konieczne jest wytrącenie znaczenia uprzednich osiągnięć, jako że są one istotnie związane z faktem otrzymywania pomocy. W modelu (5) uwzględniono te zmienne, które są brane pod uwagę w modelach EWD. Efekt dla wskaźnika zagregowanego do poziomu klasy znacząco zmaleł, jednak nadal pozostał ujemny i istotny.

Uwzględnienie obu tych wskaźników w jednym modelu (patrz model (6)) pozwoliło zwerifikować, czy zarówno czynnik na poziomie indywidualnym, jak i na poziomie grupowym, równocześnie mają znaczenie dla wyników egzaminacyjnych. Okazało się, że zarówno uczęszczanie do klasy, w której uczniowie częściej korzystają z bezpłatnej pomocy w nauce (niezależnie od tego, czy samemu się z niej korzysta), jak i otrzymywanie takiego wsparcia z przedmiotów matematyczno-przyrodniczych jest negatywnie związane z wynikami egzaminacyjnymi. Wyniki te pozwalają wysnuć przypuszczenie, że fakt otrzymywania pomocy w nauce może być także rozpatrywany jako wskaźnik trudności w uczeniu się. Jednak tego typu pomoc nie przewyższa tych trudności na tyle, by uczniowie co najmniej dorównali tym, którzy mieli takie same wyniki na sprawdzianie, a ze wsparcia nie korzystali (bo być może go nie potrzebowali). Dodatkowo, negatywny efekt zagregowanego wskaźnika na poziomie klas dopełnia obraz trudności w nauce i nieskutecznej edukacji w klasach, w których rodzina próbuje przejąć rolę nauczyciela. Przypuszczenia te nabiorą jednak większej mocy, jeśli z modelu wytrącimy wpływ innych zmiennych, które są związane z wynikami egzaminacyjnymi. Wyniki te zaprezentowano w tabeli 7. W modelach tych uwzględniono dwa wskaźniki ilości otrzymywanej pomocy w nauce: wskaźnik na poziomie indywidualnym i zagregowany uwzględniony na poziomie oddziału.

Dodanie do modelu inteligencji (model (7)) spowodowało spadek siły zależności wyników egzaminacyjnych od ilości otrzymywanego wsparcia, jednak efekt nadal pozostał istotny. Zaobserwowaliśmy także spadek siły efektu dla wskaźnika z poziomu oddziału, który w tym modelu spadł poniżej progu istotności (jest on szacowany z większym błędem niż wskaźnik z poziomu ucznia). Natomiast kontrola w modelu bezradności intelektualnej i wzorca motywacji nie powoduje zmiany w zakresie relacji między wynikami egzaminu a otrzymywaną pomocą w nauce wyrażoną w postaci wskaźnika indywidualnego (modele (8) i (9)). W modelach tych jednak wytrącenie części wariancji wyników na poziomie indywidualnym spowodowało, że średnia ilość wsparcia w nauce w oddziale znow jest istotnie i negatywnie związana z wynikami. Kontrola w modelu zmiennych związanych ze statusem społeczno-ekonomicznym rodziny ucznia (model (10)) powoduje podobny efekt, jak w przypadku analizy związku z korepetycjami, tzn. delikatny wzrost siły interesującej nas zależności dla wskaźnika z poziomu ucznia. Siła

Tabela 7. Znaczenie bezpłatnej pomocy w nauce dla wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej z uwzględnieniem dodatkowych zmiennych kontrolowanych. Wyniki trzypoziomowych analiz regresji, modele z losową stałą (7–11) i losowym nachyleniem (12), odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej												
	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)
oszacowanie efektów stałych												
poziom uczni												
bezpłatne wsparcie z mat-przyr.	-0,448	(0,116)	-0,459	(0,122)	-0,498	(0,111)	-0,573	(0,111)	-0,570	(0,111)	-0,466	(0,111)
sprawdzian	9,65	(0,284)	9,15	(0,296)	9,12	(0,293)	8,37	(0,288)	8,39	(0,287)	8,38	(0,283)
sprawdzian ²	1,227	(0,119)	1,13	(0,118)	1,14	(0,117)	1,19	(0,112)	1,20	(0,113)	1,19	(0,112)
sprawdzian ³	-0,374	(0,073)	-0,336	(0,078)	-0,334	(0,075)	-0,303	(0,073)	-0,304	(0,073)	-0,306	(0,072)
plec ^a	-1,47	(0,271)	-1,48	(0,264)	-1,80	(0,280)	-1,99	(0,276)	-1,99	(0,276)	-2,02	(0,276)
dysleksja na egz. gimnazjalnym ^b	1,61	(0,827)	1,49	(0,838)	1,53	(0,821)	1,18	(0,835)	1,18	(0,828)	1,19	(0,838)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym	-3,26	(0,890)	-3,08	(0,898)	-3,14	(0,872)	-2,76	(0,880)	-2,72	(0,869)	-2,78	(0,881)
test matryc Ravena (I pomiar)	3,83	(0,208)	3,49	(0,197)	3,48	(0,193)	3,33	(0,195)	3,33	(0,194)	3,32	(0,196)
bezradność intelektualna (mat)			-1,93	(0,144)	-1,73	(0,149)	-1,69	(0,144)	-1,68	(0,145)	-1,68	(0,147)
motywacja niska ^c					-0,561	(0,526)	-0,660	(0,503)	-0,512	(0,500)	-0,553	(0,510)
motywacja wysoka ^c					0,726	(0,369)	0,566	(0,356)	0,557	(0,354)	0,554	(0,355)
motywacja zewnętrzna ^c					-0,448	(0,390)	-0,462	(0,383)	-0,449	(0,438)	-0,438	(0,323)
motywacja wewnętrzna ^c					1,91	(0,443)	1,69	(0,436)	1,68	(0,438)	1,69	(0,435)
aspiracje rodziców							0,484	(0,066)	0,489	(0,066)	0,484	(0,067)
zasobność gosp. domowego							0,157	(0,176)	0,179	(0,174)	0,159	(0,177)
wykształcenie rodziców							0,089	(0,082)	0,102	(0,083)	0,086	(0,083)
HSIOPS							0,052	(0,016)	0,052	(0,016)	0,053	(0,016)
stała	98,93	(0,549)	98,76	(0,537)	98,53	(0,580)	97,70	(1,25)	88,20	(1,26)	87,74	(1,27)

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej												
	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)
poziom oddziałów												
średnia z bezpłatnego wsparcia z mat-przyr.	-0,858	(0,549)	-1,08	(0,469)	-1,10	(0,475)	-0,98	(0,468)	-0,876	(0,444)	-0,971	(0,473)
poziom szkół												
lok.: miasto do 20 tys. ^d									-1,31	(0,557)		
lok.: miasto od 20 do 100 tys. ^d									-1,76	(0,531)		
lok.: miasto powyżej 100 tys. ^d									-1,11	(0,668)		
oszacowanie efektów losowych												
wariancja efektów szkół	4,30		4,16		4,03		3,83		3,44		3,89	
wariancja efektów oddziałów	3,50		3,92		3,88		3,41		3,33		2,47	
wariancja na poziomie ucznia	58,81		55,83		55,36		53,83		53,84		53,56	
poziom oddziałów												
losowe nachylenie dla korepetycji z mat-przyr.											0,16	
podsumowanie												
deviance	29238,91		29027,08		28990,03		28861,04		28851,42		28856,95	
liczba szacowanych parametrów	13		14		18		22		25		24	

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji; c – grupa odniesienia: uczniowie o profilu motywacji o przeciętnym nasileniu; d – grupa odniesienia: wieś

związku dla średniej oddziałowej nie uległa zmianie. Podobnie, dodanie do modelu informacji o wielkości miejscowości, w której znajduje się szkoła, nie zmieniło oszacowań związków dla interesujących nas wskaźników (model (11)). W modelu (12) testowano hipotezę o istnieniu różnej siły zależności wyników egzaminacyjnych od ilości otrzymywanego wsparcia w różnych oddziałach klasowych. Hipoteza ta została odrzucona, jako że wariancja efektów losowych dla nachylenia na poziomie oddziałów jest nieistotnie statystycznie różna od zera.

Najważniejszy wniosek z opisanych powyżej modeli z dodatkowymi zmiennymi niezależnymi jest taki, że kontrola charakterystyk ucznia i szkoły mających znaczenie dla wyników osiągnięć (i częściowo będących powiązanych z ilością otrzymywanej pomocy w nauce) nie zmieniła kierunku i istotności zaobserwowanych efektów. Aby lepiej zinterpretować siłę opisanych zależności, przeliczmy je na różnice między grupami. Korzystając z informacji o rozkładzie wartości dla zbudowanych wskaźników, możemy obliczyć, że różnica między uczniami, którzy zadeklarowali, że w ogóle nie uzyskują pomocy w nauce w ramach przedmiotów matematyczno-przyrodniczych, a tymi, którzy otrzymują jej najwięcej spośród badanej próby uczniów wynosi około 2 punkty na egzaminie gimnazjalnym (wyrażonym na skali o odchyleniu standardowym 15). Natomiast różnica między uczniami, którzy nie otrzymują wsparcia i tymi, którzy otrzymują średnią jego ilość (spośród tych, którzy zadeklarowali, że jakąś pomoc dostają), wynosi 1,1 punktu. Różnice te są wyznaczane przy założeniu o niezmienności wartości dla pozostałych zmiennych uwzględnionych w modelu (11). Dla wskaźnika zagregowanego do poziomu oddziału, różnica w średnich wynikach egzaminacyjnych między klasami o skrajnych wartościach średniej ilości otrzymywanego wsparcia będzie wynosiła 2,6 punktu. Natomiast oddziały, w których średnia ilość otrzymywanej pomocy jest równa pierwszemu i dziewiątemu percentylowi, różnią się średnio wynikami egzaminacyjnymi o 1,1 punktu. Widzimy więc zatem, że wykryte efekty, mimo że istotne statystycznie, są słabe.

Analizy dla części humanistycznej egzaminu gimnazjalnego prowadzą do takich samych wniosków, co analogiczne analizy przeprowadzone dla części matematyczno-przyrodniczej. By nie powtarzać interpretacji, ich wyniki modeli przedstawiono w aneksie. Dodamy tylko słowo komentarza do wielkości efektów. Wydawać by się bowiem mogło, że zaobserwowane zależności są silniejsze dla przedmiotów humanistycznych (wielkości współczynników regresji są co do wartości bezwzględnej większe). Pamiętać jednak należy o tym, że nasilenie zjawiska pomocy w nauce z przedmiotów humanistycznych jest mniejsze, więc relatywnie siła efektów jest podobna.

Opisane powyżej analizy znaczenia bezpłatnej pomocy w nauce dla wyników egzaminów zewnętrznych pokazały istnienie negatywnego związku korzystania z tego typu wsparcia z wynikami egzaminacyjnymi. Negatywną zależność obserwuje się zarówno na poziomie indywidualnym, jak i ze zmienną z poziomu oddziału, która mówi o nasileniu

zjawiska otrzymywania bezpłatnej pomocy w nauce wśród uczniów z jednej klasy. Związki te się utrzymują nawet jeśli w modelu są kontrolowane zmienne ważne dla korzystania ze wsparcia w nauce oraz związane z wynikami egzaminacyjnymi. Jednak kontrola zmiennych uwzględnianych w modelowaniu EWD (w szczególności zaś uprzednich osiągnięć) pozwala znacząco zredukować ich siłę. Wyniki te są zbliżone do wyników analizy problemu uczęszczania na korepetycje. Uczniowie, którzy otrzymują dodatkową nieodpłatną pomoc w nauce, mają średnio niższe wyniki na egzaminie gimnazjalnym w porównaniu do uczniów, którzy takiej pomocy nie otrzymują, a nie różnią się pod względem uprzednich osiągnięć, inteligencji, statusu społeczno-ekonomicznego rodziny, motywacji, poziomu bezradności intelektualnej, płci i miejsca zamieszkania. Dodatkowo zaobserwowany negatywny związek ze wskaźnikiem na poziomie oddziału pokazuje, że próba przejścia przez rodzinę roli nauczyciela idzie w parze z nieefektywnym nauczaniem, prowadzącym do niższych rezultatów na egzaminie. Zaobserwowane związki nie są jednak silne.

6.4. Podsumowanie

Przeprowadzone analizy wykazały, że dla trafności wskaźników EWD większe znaczenie ma korzystanie z dodatkowego wsparcia w nauce z przedmiotów matematyczno-przyrodniczych niż z przedmiotów humanistycznych. Odsetek uczniów korzystających z korepetycji z przedmiotów humanistycznych jest niewielki, natomiast korzystających z bezpłatnego wsparcia w tym zakresie – mniejszy niż odsetek korzystających z niego w zakresie przedmiotów matematyczno-przyrodniczych. Dodatkowo obserwujemy większe międzyszkolne zróżnicowanie korzystania z bezpłatnego wsparcia z przedmiotów matematyczno-przyrodniczych.

Ponadto ujawniła się negatywna zależność pomiędzy korzystaniem z dodatkowego wsparcia w nauce poza lekcjami w szkole a osiąganymi wynikami egzaminacyjnymi. Uczniowie, którzy zadeklarowali korzystanie z korepetycji z przedmiotów matematyczno-przyrodniczych, uzyskują w tej części egzaminu gimnazjalnego wyniki niższe niż uczniowie, którzy z korepetycji nie korzystali. Związek ten utrzymuje się przy kontroli charakterystyk ucznia i jego rodziny istotnych zarówno dla uczęszczania na korepetycje, jak i dla wyników egzaminacyjnych. Podobne rezultaty otrzymano w przypadku korzystania z bezpłatnego wsparcia w nauce – wiąże się ono z niższymi osiągnięciami na egzaminie. Schemat badania nie pozwala na wnioskowanie, że jest to zależność przyczynowo-skutkowa (lecz jednocześnie jej nie wyklucza). Negatywna zależność wynikać może z faktu korzystania z pozaszkolnego wsparcia w nauce przez uczniów o określonych cechach, m.in. o niższych wcześniejszych osiągnięciach uczniów i niższej inteligencji.

Uczniowie korzystający z dodatkowych zajęć poza lekcjami w szkole mogli także rozwinąć swoją wiedzę, jednak nie na tyle, by była ona większa niż wiedza tych uczniów, którzy nie potrzebowali pomocy w nauce.

Uzyskane wyniki skłaniają do sformułowania kolejnych pytań dotyczących relacji między korzystaniem z obu form wsparcia w nauce a osiągnięciami szkolnymi. Ich przypuszczalny reaktywny w stosunku do kłopotów w szkole charakter (korzystają z nich uczniowie o niższych wynikach na sprawdzianie i niższej inteligencji) nasuwa pytania dotyczące siły i kierunku związku w zależności od poziomu wcześniejszych osiągnięć i poziomu inteligencji. Ponadto w przyszłości warto przeprowadzić analizy uwzględniające nie tylko, czy uczeń korzystał z korepetycji, lecz ile czasu im poświęcał.

Podsumowując, wbrew potocznej opinii nie możemy powiedzieć, że szkoły, w których uczniowie otrzymują dodatkowe wsparcie w nauce (płatne lub bezpłatne) zyskują we wskaźnikach EWD.

Widoczne jest także, że w modelach EWD negatywna zależność między korzystaniem z dodatkowego wsparcia w nauce pozostaje istotna, jednak znacznie słabnie. To bardzo ważny rezultat, gdyż dane dotyczące korzystania przez uczniów ze wsparcia w nauce nie są gromadzone w systemie egzaminów zewnętrznych i z dużym prawdopodobieństwem nigdy się to nie zmieni. Nie ma więc możliwości uwzględniania ich w modelach populacyjnych. Mimo że omawiana zależność pozostaje w nich istotna, pozostają one znacznie lepszą miarą efektywności pracy szkoły niż wyniki egzaminacyjne.

Literatura

Black, A. E. i Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, 84(6), 740–756.
doi: 10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3

Bray, M. (2007). *The shadow education system: private tutoring and its implications for planners*. Paryż: UNESCO, International Institute for Educational Planning.

Briggs, D. C. (2001). The Effect of Admissions Test Preparation: Evidence from NELS:88. *CHANCE*, 14(1), 10–18. doi:10.1080/09332480.2001.10542245

CBOS (2012). *Wydatki rodziców na edukację dzieci w roku szkolnym 2012/2013* (No. BS/141/2012). Warszawa: Fundacja Centrum Badania Opinii Społecznej. Pobrano z http://www.cbos.pl/SPISKOM.POL/2012/K_141_12.PDF

Cichocka, A. i Milewicz, M. (2010). Co się kryje w nieistotnych efektach statystycznych? Możliwości zastosowania analizy supresji w psychologii społecznej. *Psychologia Spo*, 5(14), 191–198.

Dang, H.-A. (2007). The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review*, 26(6), 683–698.
doi:10.1016/j.econedurev.2007.10.003

Dolata, R. (2008). *Szkoła, segregacje, nierówności*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.

Domanski, H. i Pokropek, A. (2011). *Podziały terytorialne, globalizacja a nierówności społeczne: wprowadzenie do modeli wielopoziomowych*. Warszawa: Wydawnictwo IFiS PAN.

Driessen, G., Smit, F. i Sleegers, P. (2005). Parental Involvement and Educational Achievement. *British Educational Research Journal*, 31(4), 509–532.
doi:10.1080/01411920500148713

Ferrer, E. i McArdle, J. J. (2004). An Experimental Analysis of Dynamic Hypotheses About Cognitive Abilities and Achievement From Childhood to Early Adulthood. *Developmental Psychology*, 40(6), 935–952. doi:10.1037/0012-1649.40.6.935

Grolnick, W. S., Benjet, C., Kurowski, C. O. i Apostoleris, N. H. (1997). Predictors of parent involvement in children's schooling. *Journal of Educational Psychology*, 89(3), 538–548.

GUS. (2012a). *Rocznik statystyczny Rzeczypospolitej Polskiej 2012*. Warszawa: Zakład Wydawnictw Statystycznych.

GUS. (2012b). *Powierzchnia i ludność w przekroju terytorialnym w 2012 r.* Warszawa: Główny Urząd Statystyczny. Pobrano z http://www.stat.gov.pl/cps/rde/xbcr/gus/L_powierzchnia_ludnosc_teryt_2012.pdf

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hawrot, A. i Pisarek, J. (2013). *Inteligencja a wyniki nauczania. Niepublikowana praca na podstawie wyników badania podłużnego SUEK*. Warszawa: IBE.

Heaven, P. C. L. i Ciarrochi, J. (2012). When IQ is not everything: Intelligence, personality and academic performance at school. *Personality and Individual Differences*, 53(4), 518–522. doi:10.1016/j.paid.2012.04.024

Holloway, S. D., Yamamoto, Y., Suzuki, S. i Mindnich, J. (2008). Determinants of Parental Involvement in Early Schooling: Evidence from Japan. *Early Childhood Research & Practice*, 10(1). Pobrano z <http://ecrp.uiuc.edu/v10n1/holloway.html>

Hong, S. i Ho, H.-Z. (2005). Direct and Indirect Longitudinal Effects of Parental Involvement on Student Achievement: Second-Order Latent Growth Modeling Across Ethnic Groups. *Journal of Educational Psychology*, 97(1), 32–42. doi:10.1037/0022-0663.97.1.32

Hoover-Dempsey, K. V., Battiato, A. C., Walker, J. M. T., Reed, R. P., DeJong, J. M. i Jones, K. P. (2001). Parental Involvement in Homework. *Educational Psychologist*, 36(3), 195–209.

IBE. (2012). *Raport o stanie edukacji 2011. Kontynuacja zmian*. Warszawa: IBE.

Ireson, J. (2004). Private Tutoring: how prevalent and effective is it? *London Review of Education*, 2(2), 109–122. doi:10.1080/1474846042000229458

Konarski, R. (2009). *Modele równań strukturalnych: teoria i praktyka*. Warszawa: Wydawnictwo Naukowe PWN.

Mischo, C. i Haag, L. (2002). Expansion and effectiveness of private tutoring. *European Journal of Psychology of Education*, 17(3), 263–273. doi:10.1007/BF03173536

Muthén, L. K. i Muthén, B. O. (1998–2010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

OECD. (2010). *PISA 2009 Results: Overcoming social background equity in learning opportunities and outcomes (Volume II)*. Paryż: OECD.

OECD. (2011). *Quality Time for Students: Learning In and Out of School*. Paryż: OECD Publishing.

Putkiewicz, E. (2006). Private tutoring in Poland. [w:] Brejc, M. (red.), *Co-operative Partnerships in Teacher Education. Proceedings of the 31st Annual ATEE Conference* (s. 185–197). Portorož, Slovenia: ATEE. Pobrano z <http://www.pef.uni-lj.si/atee/978-961-6637-06-0/185-197.pdf>

Rosenzweig, C. (2001). A Meta-Analysis of Parenting and School Success: The Role of Parents in Promoting Students' Academic Performance. Zaprezentowano na the Annual Meeting of the American Educational Research Association, Seattle, WA. Pobrano z <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED452232>

Ryan, R. M. i Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57(5), 749–761. doi:10.1037//0022-3514.57.5.749

Ryan, R. M. i Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67. doi:10.1006/ceps.1999.1020

Scientific Software International. (2012). Pobrano 19 marca 2012, z <http://www.ssicentral.com/hlm/example6-2.html>

Sędek, G. (1995). *Bezradność intelektualna w szkole*. Warszawa: Wydawnictwo Instytutu Psychologii.

Shaughnessy, J., Zechmeister, E. i Zechmeister, J. (2008). *Metody badawcze w psychologii*. Gdańsk: GWP.

Singh, K., Bickley, P. G., Trivette, P., Keith, T. Z., Keith, P. B. i Anderson, E. (1995). The Effects of Four Components of Parental Involvement on Eighth-Grade Student Achievement: Structural Analysis of NELS-88 Data. *School Psychology Review*, 24(2), 299–317.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75(3), 417–453. doi:10.3102/00346543075003417

Snijders, T. A. B. i Bosker, R. J. (2012). *Multilevel analysis: an introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles: Sage.

Stanisz, A. (2007). *Przystępny kurs statystyki: z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3*. Kraków: StatSoft.

Tansel, A. i Bircan, F. (2006). Demand for education in Turkey: A tobit analysis of private tutoring expenditures. *Economics of Education Review*, 25(3), 303–313. doi:10.1016/j.econedurev.2005.02.003

Teo, A., Carlson, E., Mathieu, P., Egeland, B. i Sroufe, L. A. (1996). A Prospective Longitudinal Study of Psychosocial Predictors of Achievement. *Journal of School Psychology*, 34(3), 285–306. doi:10.1016/0022-4405(96)00016-7

Trusty, J. (1998). Family Influences on Educational Expectations of Late Adolescents. *The Journal of Educational Research*, 91(5), 260–271. doi:10.1080/00220679809597553

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K. i Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology*, 101(3), 671–688. doi:10.1037/a0015083

Zhang, Y. (2013). Does private tutoring improve students' National College Entrance Exam performance? – A case study from Jinan, China. *Economics of Education Review*, 32, 1–28. doi:10.1016/j.econedurev.2012.09.008

Rozdział 7

Społeczno-gospodarcze charakterystyki gmin

W rozdziale przeanalizowane zostaną związki społeczno-gospodarczych charakterystyk gmin, w których ulokowane są gimnazja, z wynikami egzaminacyjnymi i wskaźnikami EWD. Czy takie czynniki jak zamożność gminy, infrastruktura kulturalna lub wielkość wydatków na oświatę wyjaśniają egzaminacyjne sukcesy szkoły? Gdyby wyniki końcowe, a przede wszystkim miary postępu czynionego przez uczniów w danym gimnazjum, były determinowane przez czynniki społeczno-gospodarcze, trafne byłoby zarzuty, że to ekonomia, a nie starania szkół decydują o sukcesach egzaminacyjnych. Przyjrzymy się zatem, czy układ korelacji potwierdza ekonomiczny determinizm. Wyniki analiz przedstawionych w tym rozdziale dostarczą też przesłanek do ewentualnego wykorzystania w modelach EWD istniejących danych, pozwalających scharakteryzować lokalne środowisko szkoły.

7.1. Społeczno-gospodarcze uwarunkowania osiągnięć szkolnych

Wykorzystanie w prowadzonych analizach modelowania hierarchicznego daje możliwość użycia danych ze źródeł wtórnych, opisujących środowisko działania szkoły, jako zmiennych z poziomu ponadindywidualnego. Ze względu na sposób doboru próby badawczej dane na temat gmin, w których znajdują się gimnazja, mogą być analizowane na poziomie efektów szkół (praktycznie każda z badanych szkół wylosowana została z terenu innej gminy). Rozszerzenie analiz o dane zastane jest korzystne z dwóch powodów. Po pierwsze, pozwala sprawdzić, na ile modele EWD „oczyszczają” miarę efektywności pracy szkół z czynników zewnętrznych o charakterze lokalnym w stosunku do wykorzystywania jako miar jedynie osiągnięć na egzaminach. Po drugie, pozwala ocenić na ile modele opisujące zróżnicowanie szkół, bazujące jedynie na danych wtórnych oraz na agregatach wyników egzaminów, trafnie odzwierciedlają faktycznie występujące zależności.

Znaczna część analiz dotyczących czynników oddziałujących na osiągnięcia szkolne opiera się wyłącznie lub częściowo na danych gromadzonych w ramach regionalnych lub krajowych systemów statystycznych. Zaletą tego typu podejścia jest możliwość czerpania z szerokiej puli danych zastanych, w tym z powszechnie interpretowanych wskaźników opisujących rynek pracy, lokalną/regionalną gospodarkę, poziom zamożności poszczególnych jednostek administracyjnych oraz ich mieszkańców, struktury wydatków na różnego rodzaju polityki publicznej. Z kolei jego wadą może być próba wyciągania wniosków substancjalnych na podstawie relacji pozornych, których zrozumienie i wyjaśnienie możliwe jest po włączeniu do analizy danych na poziomie indywidualnym (tzw. błąd ekologiczny).

7.1.1. Zamożność

Wyniki międzynarodowych, porównawczych badań umiejętności zestawiane są z poziomem życia mierzonym w zróżnicowany sposób. Wnioski z tych analiz wskazują, że istnieje silna korelacja pomiędzy PKB kraju/regionu a wynikami testów umiejętności. Badacze wskazują na przyczynowy związek zachodzący pomiędzy umiejętnościami a zamożnością. Na podstawie przeglądu analiz prowadzonych na danych międzynarodowych Eric Hanushek i Ludger Woessmann (2011) stwierdzili, że zmiany w wiedzy i umiejętnościach mierzonych za pomocą międzynarodowych testów porównawczych, są źródłem wzrostu dochodów i zamożności na poziomie krajów. Do podobnych wniosków prowadzi również raport OECD przygotowany na podstawie dotychczasowych edycji badań PISA (OECD, 2010). W analizach wskazuje się również na możliwość istnienia odwrotnej relacji przyczynowej pomiędzy zamożnością społeczności a osiągnięciami. Na podstawie danych dla Polski, Jan Herczyński i Mikołaj Herbst (2002) wskazują na wpływ zamożności mieszkańców gmin (mierzonej wpływami z PIT do budżetu samorządu) na osiągnięcia uczniów na testach gimnazjalnych. Najnowsze badania sugerują, że pomiar zamożności, poziomu dobrobytu społeczeństw/społeczności powinien opierać się na bardziej złożonych miarach niż PKB czy dochody (Yu, 2012).

7.1.2. Wydatki na oświatę

Z zamożnością wiązą się bezpośrednio wydatki na oświatę. Analizy prowadzone były zarówno na poziomie krajowym, jak i międzynarodowym. Przykładowo, na danych dla Polski Mikołaj Herbst wykazał brak związku pomiędzy wydatkami na edukację a osiągnięciami uczniów (Herbst, 2004). Do podobnych wniosków prowadzą również porównania międzynarodowe. Hanushek i Woessmann (2011), na podstawie przeglądu literatury w tym zakresie stwierdzili, że wzrost wydatków na edukację w przeliczeniu na ucznia nie przekłada się na podniesienie poziomu osiągnięć uczniów. Według wspomnianych autorów również analiza zmian w wydatkach na edukację w przeliczeniu na ucznia prowadzi do wniosków, że zwiększenie wydatków nie ma związku ze zmianami w wynikach uzyskiwanych na międzynarodowych testach umiejętności. Z tego względu w literaturze przedmiotu (Costrell, Hanushek, i Loeb, 2008; Hanushek, 2001; 2013) wskazuje się na konieczność uwzględniania w analizach wydatków dodatkowych czynników kontrolnych, zwłaszcza w zakresie jakości i organizacji nauczania.

7.1.3. Bezrobocie

Na podstawie danych ogólnokrajowych (USA) Elizabeth Ananat, Anna Gassman-Pines, Diana Francis i Christina Gibson-Davis (2011) wykazały, że wzrost stopy bezrobocia wpływa na obniżenie osiągnięć szkolnych mierzonych przede wszystkim wynikami zewnętrznych egzaminów z matematyki. Również w Polsce wyniki egzaminów zewnętrznych zostały poddane analizie pod kątem oddziaływania lokalnego poziomu bezrobocia na ich średni wynik na

terenach gminy. Jak wskazują Herczyński i Herbst (Herbst, 2004; Herczynski i Herbst, 2002), wzrost lokalnej stopy bezrobocia (na poziomie powiatu), przy kontroli innych czynników lokalnych, wpływa na obniżenie wyników uzyskiwanych przez gimnazjalistów w testach zewnętrznych.

Z kolei analizy prowadzone z uwzględnieniem informacji o indywidualnych osiągnięciach uczniów z dołączonymi danymi na temat lokalnej stopy bezrobocia, nie dają tak jednoznacznych rezultatów. Leon Feinstein i James Symons (1999) stwierdzili, że bezrobocie na poziomie lokalnym nie wpływa na wyniki uzyskiwane w testach, a znaczenie mają zmienne indywidualne, grupowe i interakcje rodzic-nauczyciel. Z kolei Phillip B. Levine (2011) na podstawie analizy danych NLSY79 wskazał, że negatywne oddziaływanie lokalnej stopy bezrobocia na gorsze wyniki w testach osiągnięć dotyczy jedynie dzieci z rodzin o niskim statusie społeczno-ekonomicznym. Do jeszcze innych wniosków doszły Ann Stevens i Jessamyn Schaller (2009) wskazując, że lokalny poziom bezrobocia zwiększa ryzyko powtarzania klasy przez uczniów, także przy kontroli bezrobocia rodziców, również oddziałującego negatywnie.

7.1.4. Struktura lokalnej gospodarki

W analizach poziomu umiejętności/wiedzy młodzieży wskazuje się także na wpływ struktury lokalnej gospodarki. Jest to temat relatywnie rzadko eksplorowany. W badaniach (Broomhall i Johnson, 1994; James, 2001) pozytywnie zweryfikowano hipotezę, że na obszarach, na których jest niewielka liczba podmiotów gospodarczych zgłaszających zapotrzebowanie na wysoko wykwalifikowanych pracowników, młodzi ludzie nie znajdują uzasadnienia dla potrzeby kształcenia. Brak przykładów transmisji wyższego wykształcenia na wyższe dochody prowadzi do przykładania mniejszej wagi do zdobywania wiedzy i umiejętności. W efekcie uczniowie ci szybciej niż ich rówieśnicy z innych rejonów rezygnują z dalszej nauki.

7.1.5. Zasoby kulturowe

Źródeł różnicowania osiągnięć szkolnych uczniów można również poszukiwać w dostępie do zasobów kulturowych reprezentowanych najczęściej przez dostęp do bibliotek szkolnych. Jak wskazują liczne badania zebrane w przeglądowym raporcie Dorothy Williams, Caroline Wavell i Louisy Coles (2001) oraz w analizie Chana (2008) dostępność tego typu instytucji oraz jakość świadczonych w nich usług, wpływa na osiągnięcia uczniów również przy kontroli czynników indywidualnych. W Polsce efekty działania bibliotek na terenach wiejskich zostały przedstawione w raporcie *Po co Polakom biblioteki* (Giza-Poleszczuk i in., 2012). Jednym z wniosków autorów jest zwrócenie uwagi na rolę bibliotek publicznych w podnoszeniu wiedzy dużej grupy użytkowników.

7.1.6. Przedszkola

Osobnym problemem wymagającym wnikliwej analizy są efekty opieki przedszkolnej. Istnieje wiele badań w tym zakresie. Dwie duże metaanalizy badań eksperymentalnych i quasi-eksperymentalnych (Camilli, Vargas, Ryan, Barnett, 2010; Leak i in., 2010) wskazują na istotny, lecz malejący z czasem wpływ edukacji przedszkolnej na osiągnięcia szkolne dzieci. Najmocniejszy wpływ miała edukacja przedszkolna na wiedzę, mniejszy zaś na umiejętności szkolne i społeczne. W zakresie znaczenia wychowania przedszkolnego w Polsce, do odmiennych wniosków doszedł Krzysztof Konarzewski (2012). W analizach wykorzystał dane z polskiej edycji badania osiągnięć matematycznych i przyrodniczych (TIMSS) oraz osiągnięć w rozumieniu tekstu (PIRLS). Przy kontroli płci i SES rodziny dziecka nie wystąpiły różnice w osiągnięciach dziesięciolatków uczęszczających i nieuczęszczających do przedszkola.

Z kolei badanie Marie-Claude Geoffroya (2010) wskazuje na występowanie interakcji pomiędzy wykształceniem rodziców a opieką przedszkolną dzieci. Analizy przeprowadzone zostały na kanadyjskich danych kohortowych z badania *Quebec Longitudinal Study of Child Development* (QLSCD), obejmującego wszystkie dzieci urodzone w prowincji Quebec pomiędzy październikiem 1997 a lipcem 1998 roku. Na ich podstawie można stwierdzić, że pozytywne efekty uczęszczania dzieci do przedszkoli widoczne są wśród dzieci, których matki nie mają wykształcenia średniego lub wyższego.

Poziom upowszechnienia opieki przedszkolnej może również wpływać na późniejsze osiągnięcia edukacyjne. Wykazali to pośrednio Gabriela Schütz, Heinrich Ursprung i Ludger Woessmann (2005) na podstawie porównań międzynarodowych. Stwierdzili, że istnieje zależność pomiędzy poziomem powszechności nauczania przedszkolnego a siłą wpływu „zaplecza rodzinnego” (*family background*) na osiągnięcia edukacyjne. Wzrastający udział dzieci w nauczaniu przedszkolnym do poziomu około 60% powoduje wzrost nierówności w osiągnięciach ze względu na zaplecze rodzinne. Powyżej tego odsetka zróżnicowanie spada. Jednakże ze względu na brak w literaturze bezpośrednich analiz dotyczących związku pomiędzy poziomem upowszechnienia opieki przedszkolnej a późniejszymi osiągnięciami, prezentowane poniżej analizy będą miały przede wszystkim charakter eksploracyjny.

7.1.7. Problem uwzględniania czynników lokalnych z punktu widzenia modeli EWD

Analiza wpływu czynników lokalnych, o których informacja gromadzona jest w ramach statystyki publicznej, na osiągnięcia edukacyjne uczniów bez uwzględnienia oraz przy kontroli modelu EWD, pozwoli podzielić je na trzy podstawowe grupy. Pierwszą stanowią te, które nie mają znaczenia dla osiągnięć szkolnych. Drugą te, które oddziałują przy kontroli czynników indywidualnych, a przestają oddziaływać w modelach EWD. Ostatnią grupę stworzą zmienne, które mają wpływ na

osiągnięcia szkolne również przy kontroli zmiennych analizowanych w ramach modeli EWD (wynik sprawdzianu, posiadanie zaświadczenia o specyficznych trudnościach w uczeniu się, płeć). O ile istnieją czynniki, które zaliczyć można do grupy trzeciej, należałoby uwzględnić je w modelach EWD, lub przynajmniej mieć je na uwadze przy interpretacji wskaźników.

7.2. Uwarunkowania lokalne: konstrukcja wskaźników i ich rozkłady w próbie

Informacje o czynnikach lokalnych pobrane zostały z Banku Danych Lokalnych Głównego Urzędu Statystycznego. BDL GUS jest „największym w Polsce uporządkowanym zbiorem informacji o sytuacji społeczno-gospodarczej, demograficznej, społecznej oraz stanie środowiska, opisującym województwa, powiaty oraz gminy jako podmioty systemu organizacji społecznej i administracyjnej państwa” (GUS, 2013). Dane do analiz zostały pobrane dla wszystkich gmin w Polsce, łącznie z kodem jednostki samorządu terytorialnego. Do analiz wykorzystano najświeższe dostępne w chwili kwerendy bazy danych (za 2010 rok). Informacje, po odpowiednim przygotowaniu, zostały dołączone do zbioru danych indywidualnych uczniów.

7.2.1. Dane na temat zamożności

Jako lokalne miary zamożności wykorzystano dwa wskaźniki. Pierwszy to poziom wydatków gminy na osobę, świadczący o ogólnej zamożności samorządu lokalnego. Drugim jest syntetyczny wskaźnik ubóstwa, powstały jako efekty analizy głównych składowych z uwzględnieniem następujących zmiennych: udział osób w gospodarstwach domowych korzystających ze środowiskowej pomocy społecznej w ludności ogółem, świadczenia na rzecz osób fizycznych na osobę, wpływy z PIT na osobę w gminie (wpływ negatywny) oraz wpływy z CIT na osobę w gminie (wpływ negatywny). Wszystkie dane dotyczyły roku 2010. Liczba czynników do ekstrakcji określona została na podstawie wartości własnej większej od 1. Wartości czynnikowe dla poszczególnych gmin zapisane zostały z wykorzystaniem metod regresyjnych. Szczegóły analizy zaprezentowano w aneksie.

7.2.2. Dane na temat lokalnej gospodarki

Do skonstruowania wskaźników opisujących strukturę lokalnej gospodarki wykorzystane zostały informacje na temat liczby podmiotów ujętych w rejestrze REGON według sekcji i działów PKD 2007 (Sekcje od A do U, działy od 1 do 98) na poziomie gmin. Są to najdokładniejsze dane dostępne na tym poziomie podziału administracyjnego. Należy mieć na uwadze, że rejestr REGON nie jest idealnym odzwierciedleniem rzeczywistości. Obarczony jest błędem wynikającym z faktu, że funkcjonują w nim podmioty, które zaprzestały już prowadzenia działalności. Na potrzeby prowadzonych analiz przyjęto założenie, że wielkość tego błędu jest niezależna od gminy.

Na podstawie wyżej wymienionych danych skonstruowano sześć wskaźników odnoszących się do struktury lokalnej gospodarki. Pierwsze trzy to udział trzech podstawowych sektorów gospodarki (rolnictwo i wydobywanie; przemysł i budownictwo; usługi) w jej lokalnej strukturze. Czwarty wskaźnik to liczba podmiotów gospodarczych na 100 mieszkańców gminy, będąca jednocześnie wskaźnikiem możliwości prowadzenia działalności gospodarczej oraz lokalnej kultury w tym zakresie.

Piąty i szósty wskaźnik odnoszą się do przemysłu wysokich technik. Jest to ta część przemysłu, w której podmioty charakteryzują się relatywnie wysokimi nakładami na badania i rozwój, wysokim poziomem zatrudnienia pracowników naukowo-technicznych, intensywną współpracą z ośrodkami naukowo-badawczymi, szybkim procesem starzenia się opracowanych produktów i technologii oraz efektami działalności innowacyjnej w postaci patentów i licencji (Świdurska, 2009). Na tej podstawie wyróżnione zostały sekcje PKD, będące składową przemysłu wysokich technologii w Polsce. Opierając się na klasyfikacji PKD 2007 do podmiotów wysokich technik (*high-tech*), zaliczono w Polsce te, które były ujęte w następujących sekcjach i działach klasyfikacji: Sekcja C dział 20, Sekcja C dział 21, Sekcja C dział 26, Sekcja C dział 27, Sekcja C dział 28, Sekcja C dział 29, Sekcja C dział 30, Sekcja J, Sekcja M, Sekcja P. Na tej podstawie obliczone zostały wskaźniki: udziału podmiotów *high-tech* (HT) w lokalnej gospodarce oraz liczba podmiotów HT na 100 mieszkańców.

7.2.3. Dane dotyczące poziomu bezrobocia

GUS publikuje informację na temat stopy bezrobocia jedynie na poziomie powiatów lub wyższym. Na poziomie gmin dostępna jest jedynie statystyka udziału zarejestrowanych bezrobotnych w liczbie osób w wieku produkcyjnym. Ta ostatnia przyjmuje znacznie niższe wartości. Dla przykładu stopa bezrobocia w Polsce w 2011 roku wyniosła 12,5%, natomiast odsetek bezrobotnych wśród osób w wieku produkcyjnym – 8,0%. W skali kraju relacja wskaźnika udziału bezrobotnych do stopy bezrobocia na przestrzeni lat 2004–2011 była stabilna, w granicach 0,64 +/-0,01. Ze względu na możliwość skorzystania z informacji na niższym poziomie agregacji – a więc precyzyjniej opisującej lokalne uwarunkowania funkcjonowania szkół – w analizach jako wskaźnik sytuacji na gminnym rynku pracy wykorzystana została informacja na temat udziału zarejestrowanych bezrobotnych w ludności w wieku produkcyjnym.

7.2.4. Dane o wydatkach na oświatę

Jako wskaźniki w tym zakresie wykorzystane zostały dwie informacje o odmiennym charakterze. Pierwszy to wydatki samorządu gminnego na wychowanie i oświatę na mieszkańca gminy. Informacje te są ogólne, nie dotyczą jedynie uczniów gimnazjów, wskazują jednak na zróżnicowanie w poziomie finansowego zaangażowania samorządu w realizację zadań oświatowych. Jako drugi wskaźnik wykorzystano wydatki bieżące samorządu na edukację gimnazjalną (wydatki samorządu w rozdziale 80110 budżetu – Gimnazja) przeliczone na liczbę

uczniów w gimnazjach prowadzonych przez samorząd gminny. Wskaźnik ten opisuje bezpośrednio sytuację finansową gimnazjów objętych badaniem. Ze względu na znaczną skośność rozkładu przed wprowadzeniem do analiz został on zlogarytmowany.

7.2.5. Dane na temat dostępności zasobów kulturowych

W Polsce każda gmina do końca 2011 roku miała obowiązek prowadzenia przynajmniej jednej biblioteki publicznej (niebędącej biblioteką szkolną ani niepełniącą jednocześnie funkcji innej instytucji kultury¹). Gmina pełni również rolę mecenasa i odpowiada za funkcjonowanie bibliotek w sposób, który zapewni mieszkańcom dogodny dostęp do materiałów bibliotecznych i informacji. Ze względu na te ustawowe obowiązki liczba bibliotek na 1000 mieszkańców oraz relatywna wielkość księgozbioru zostały wykorzystane jako wskaźniki wagi przywiązywanej przez samorząd gminny do udostępniania mieszkańcom zasobów kulturowych.

7.2.6. Dane dotyczące opieki przedszkolnej

W celu kontroli czynnika, jakim jest upowszechnienie nieobowiązkowej opieki przedszkolnej, w analizach uwzględniony został odsetek dzieci w wieku 3–5 lat objętych taką opieką, w stosunku do liczby rówieśników w gminie w roku szkolnym 2003/2004. Dane te stanowią kontekst dla informacji o uczęszczaniu uczniów, objętych badaniem do przedszkola, zebranych w ramach I etapu badań w pierwszej klasie gimnazjum.

7.2.7. Dane kontekstowe

W analizach społecznych ważnym czynnikiem jest również typ analizowanej gminy. W praktyce stosowany jest podział administracyjny na gminy miejskie, wiejskie oraz miejsko-wiejskie. Jego zasadniczą wadą jest fakt, że niezbyt dobrze odzwierciedla zróżnicowanie gmin w zakresie składu społecznego mieszkańców. W klasyfikacji tej, jako gminy wiejskie traktowane są zarazem gminy ościenne wielkich aglomeracji miejskich, jak i gminy peryferyjne bez ośrodka miejskiego. W tym kontekście lepszym wskaźnikiem zróżnicowania wydaje się wielkość gminy, zwłaszcza ze względu na prowadzenie analiz łącznie z omówionymi powyżej czynnikami lokalnymi. Ze względu na fakt, że zmienna ta nie ma rozkładu normalnego, w analizie został wykorzystany logarytm naturalny wielkości gminy².

Tabela 1 przedstawia podstawowe charakterystyki zmiennych wykorzystywanych w modelach trzypoziomowych do opisu otoczenia szkoły.

¹ Ustawa z 27 czerwca 1997 r. o bibliotekach; Dz.U. nr 85, poz. 539 z późn. zm.

² W efekcie interpretacja współczynnika regresji przy tej zmiennej (po podzieleniu przez 100) będzie mówiła, o ile zmieni się wartość zmiennej zależnej, jeśli wielkość gminy zmieni się o 1%.

Tabela 1. Statystyki opisowe dla zmiennych z poziomu lokalnego

zmienna	średnia	odch. stand.	min	max	skośność
liczba mieszkańców gminy	116033,04	274237,16	2550,00	1720398,00	4,419
wydatki w gminie ogółem na osobę w 2010 r.	3417,74	922,67	2249,07	7048,18	1,653
poziom ubóstwa	0,00	1,00	-3,38	3,20	-0,332
udział I sektora w lokalnej gospodarce	3,65	4,41	0,18	29,97	2,427
udział II sektora w lokalnej gospodarce	23,81	6,43	12,44	47,10	0,891
udział III sektora w lokalnej gospodarce	72,54	8,75	45,28	86,88	-0,655
udział sektora HT w lokalnej gospodarce	9,33	4,32	2,02	22,84	0,757
I. podmiotów gosp. na 100 mieszkańców	0,10	0,04	0,03	0,21	0,892
I. podmiotów HT na 100 mieszkańców	1,02	0,84	0,07	4,57	1,991
udział bezrobotnych w liczbie osób w wieku produkcyjnym	8,00	3,36	2,80	21,90	0,923
wydatki na osobę na oświatę i wychowanie	1112,16	223,41	586,86	1758,78	0,379
wydatki bieżące na gimnazja na ucznia	8043,59	3000,80	4734,81	31311,12	5,032
liczba bibliotek na osobę	0,29	0,24	0,00	1,11	1,521
księgozbiór bibliotek na 1000 ludności	3671,08	1553,61	0,00	8991,50	1,018
odsetek dzieci objętych wychowaniem przedszkolnym w wieku 3–5 lat w 2003 r.	36,59	19,79	0,00	74,30	-0,246

Źródło: BDL GUS

7.3. Pytania i hipotezy badawcze

W tym podrozdziale zasadnicze pytanie badawcze dotyczy tego, jak czynniki lokalne, stanowiące kontekst funkcjonowania gimnazjów oraz życia rodzin uczniów tych szkół, wpływają na osiągnięcia szkolne gimnazjalistów oraz, na ile uwzględnienie w analizach informacji o wcześniejszych osiągnięciach edukacyjnych pozwala ograniczyć wpływ tych czynników.

Można zatem postawić następujące pytania badawcze: na ile osiągnięcia edukacyjne gimnazjalistów związane są z:

- zamożnością gminy i jej mieszkańców,
- strukturą lokalnej gospodarki,
- problemem bezrobocia na terenie gminy (mierzonym udziałem zarejestrowanych osób bezrobotnych wśród osób w wieku produkcyjnym),
- wydatkami gminy na oświatę,
- dostępem do zasobów kulturowych (mierzonym relatywną liczbą bibliotek oraz wielkością księgozbioru) na terenie gminy,
- dostępem do opieki przedszkolnej na terenie gminy,
- wielkością gminy,

Ponadto w rozdziale tym podjęta zostanie próba odpowiedzi na pytania:

- Czy i jaki związek z osiągnięciami szkolnymi gimnazjalistów ma wcześniejsze objęcie ich zorganizowaną opieką przedszkolną?
- Czy i jaka jest interakcja pomiędzy wykształceniem rodziców i wcześniejszym uczęszczaniem do przedszkola a osiągnięciami szkolnymi gimnazjalistów?

Te pytania badawcze stawiane są w powiązaniu z pytaniami o związek poziomu objęcia opieką przedszkolną na terenie gminy i osiągnięciami gimnazjalistów.

- W przypadku wszystkich wymienionych wyżej pytań badawczych interesować nas również będzie, na ile wpływ poszczególnych czynników daje się kontrolować, uwzględniając informację o wcześniejszych osiągnięciach edukacyjnych uczniów.

Próba odpowiedzi na postawione wyżej pytania badawcze pozwoli nie tylko zrozumieć wpływ czynników lokalnych na osiągnięcia uczniów, ale da również możliwość sprawdzenia, czy modele EWD pozwalają kontrolować wpływ tych czynników. Jest to zatem element ewaluacji trafności metody EWD.

Na podstawie przeglądu literatury można postawić następujące hipotezy badawcze:

- H1) Zamożność gminy jest pozytywnie związana z osiągnięciami gimnazjalistów.
- H2) Wyższy poziom bezrobocia w gminie wiąże się z niższymi osiągnięciami edukacyjnymi gimnazjalistów.
- H3) W gminach, w których gospodarka zgłasza popyt na dobrze wykształconych pracowników, uczniowie będą mieć wyższe osiągnięcia edukacyjne.
- H4) Wydatki oświatowe nie mają związku z osiągnięciami uczniów.
- H5) Większa dostępność zasobów kulturowych podnosi poziom osiągnięć edukacyjnych.
- H6) [eksploracyjna] Większy dostęp do opieki przedszkolnej poprawia osiągnięcia edukacyjne uczniów.
- H7) [eksploracyjna] W większych gminach poziom osiągnięć uczniów jest wyższy.
- H8) Uczniowie objęci edukacją przedszkolną, mają wyższy poziom osiągnięć na egzaminach gimnazjalnych.

Można również sformułować ogólną hipotezę dotyczącą wpływu czynników lokalnych na wyniki egzaminu gimnazjalnego, przy kontroli wcześniejszych osiągnięć szkolnych.

- H9) Ze względu na relatywnie niewielką dynamikę zmian w zakresie czynników lokalnych, uwzględnienie w analizach wcześniejszych osiągnięć edukacyjnych pozwoli wyłączyć lub znacznie ograniczyć wpływ tych czynników na osiągnięcia na egzaminie gimnazjalnym.

7.3.1. Schemat analizy danych

W celu weryfikacji hipotez wyspecyfikowano szereg hierarchicznych modeli liniowych, które odzwierciedlały złożony schemat doboru próby (Domański i Pokropek, 2011). Wykorzystano modele trzypoziomowe z losową stałą uwzględniające poziom ucznia, klasy i szkoły. Podobnie jak w pozostałych rozdziałach, hipotezy weryfikowano, wykorzystując dwie zmienne zależne: łączny wynik egzaminu gimnazjalnego z matematyki i przedmiotów przyrodniczych oraz łączny wynik egzaminu gimnazjalnego z języka polskiego oraz WOS-u i historii.

Modelowanie, mające na celu weryfikację znaczenia poszczególnych czynników lokalnych oraz uczęszczania do przedszkola dla osiągnięć szkolnych i EWD, przeprowadzono w czterech etapach. W pierwszym pokazano związek interesującej zmiennej z wynikami egzaminu, bez kontroli zmiennych na poziomie jednostkowym. W tym celu zbudowano trzypoziomowy model regresji, w którym stanowiła ona predyktor na poziomie szkoły. W drugim etapie do modelu włączane były zmienne uwzględniane w modelach EWD. W trzecim etapie kontroli poddawane były zmienne z poziomu indywidualnego wyróżnione jako istotne w rozdziale 1, w części dotyczącej modeli podstawowych stosowanych w analizach (poziom inteligencji, wiek uczniów, wyższy z poziomów wykształcenia rodziców, wyższy ze wskaźników prestiżu zawodów, zamożność gospodarstwa domowego, wyposażenie gospodarstwa w sprzęt elektroniczny, kolejność urodzenia ucznia, typ rodziny). W czwartym zbudowano trzypoziomowy model EWD uwzględniający tę zmienną, pokazując jej znaczenie dla wyników egzaminów, jeśli kontroli podlegają zmienne indywidualne oraz zmienne uwzględniane w modelach EWD. Strategia ta pozwoli sprawdzić, w jakim stopniu w modelach EWD pośrednio kontrolujemy czynniki lokalne mogące oddziaływać na szkoły i uczniów. Umożliwi to również stwierdzenie, czy kontrola zmiennych indywidualnych wnosi dodatkowe informacje do modeli mierzących efektywność szkół. W prezentacji wyników pokazane zostaną także modele EWD bez uwzględnienia dodatkowego interesującego nas predyktora, które stanowiły punkt odniesienia w zakresie weryfikacji hipotez o poprawie dopasowania modeli.

Ostateczna weryfikacja oddziaływania zmiennych lokalnych przeprowadzona zostanie analogicznie do opisanych powyżej czterech etapów. Poprzedzone one zostaną wyłonieniem czynników lokalnych oddziałujących równocześnie na osiągnięcia na egzaminie gimnazjalnym (osobno dla części matematyczno-przyrodniczej i części humanistycznej). Czynniki zostały wyłonione metodą krokowej eliminacji wstecznej. W pierwszym kroku do modelu ze zmienną zależną wprowadzono wszystkie analizowane osobno zmienne lokalne. W kolejnych krokach eliminowano poszczególne zmienne, kierując się wielkością poziomu istotności. W modelu ostatecznym pozostawiono jedynie te czynniki, których istotność była poniżej 0,05.

Analizy wykonano za pomocą oprogramowania HLM 6.06, wykorzystując metodę estymacji *full maximum likelihood*. Do oszacowania istotności efektów stałych wykorzystano odporne

(robust) błędy standardowe³. Wartości istotne na poziomie istotności $p < 0,05$ zaznaczono w tabelach pogrubionym drukiem. Analizy wykonano z uwzględnieniem wag warunkowych (patrz: rozdział 1) na wszystkich trzech poziomach, odpowiednio przeskalowanych za pomocą programu HLM (Scientific Software International, 2012).

7.4. Wyniki

7.4.1. Zamożność

Zmienna opisująca wydatki samorządu gminnego na osobę nie jest powiązana z poziomem uzyskiwanych przez gimnazjalistów wyników na egzaminie, tak w części humanistycznej (tabela 2, model 2), jak i matematyczno-przyrodniczej (tabela 3, model 2). Z kolei wzrost poziomu ubóstwa mieszkańców gmin wiąże się ze spadkiem osiągnięć na obu częściach egzaminu gimnazjalnego (tabela 2, model 6 oraz tabela 3, model 6). Analiza modeli, do których poza miarami zamożności wprowadzono również zmienne z modeli EWD i/lub z poziomu indywidualnego, prowadzi do wniosku o braku związku zamożności gminy z osiągnięciami gimnazjalistów (tabela 2, modele: 3, 4, 7, 8, 9 oraz tabela 3, modele: 3, 4, 5, 7, 8, 9). Jedynie w części humanistycznej przy jednoczesnej kontroli zmiennych z poziomu indywidualnego oraz zmiennych wykorzystywanych w modelach EWD, wzrost wydatków o 1000 złotych na mieszkańca gminy prowadził do spadku osiągnięć na egzaminie o 0,6 punktu na skali o średniej 100 i odchyleniu 15 (tabela 2, model 5). W przypadku analizowanych gmin prowadzi to do różnicy niespełna 3 punktów (około 1/5 odchylenia standardowego) pomiędzy gminami o najniższych i najwyższych wydatkach.

Podsumowując, można stwierdzić, że zamożność samorządów oraz mieszkańców gmin nie wpływa na osiągnięcia uczniów na egzaminach gimnazjalnych. Brak tego typu zmiennych w modelach EWD nie wpływa na obciążenie oceny efektywności poszczególnych placówek. Porównanie modeli dla części humanistycznej egzaminu przedstawia tabela 2, zaś dla części matematyczno-przyrodniczej – tabela 3.

7.4.2. Bezrobocie

Wskaźnik bezrobocia, rozumiany jako udział zarejestrowanych bezrobotnych w liczbie osób w wieku produkcyjnym na poziomie gminy – zgodnie z przewidywaniami, ma negatywny wpływ na osiągnięcia uczniów w obu częściach egzaminu gimnazjalnego, tak samodzielnie (tabela 4, model 2; tabela 5, model 2), jak i przy kontroli zmiennych indywidualnych (tabela 4, model 4; tabela 5, model 4), w tym bezrobocia rodziców (tabela 4, model 5; tabela 5, model

³ Są to oszacowania błędów standardowych odporne na naruszenie założeń dotyczących normalności rozkładu reszt (patrz np. Konarski, 2009 s. 318–324).

5). Podobny efekt występuje również w modelach analogicznych do modeli EWD (tabela 4, model 3; tabela 5, model 3) oraz w modelach uwzględniających zarówno czynniki indywidualne (w tym bezrobocie rodziców), jak i zmienne z modeli EWD (tabela 4, modele 6–7; tabela 5, modele 6–7).

W modelach, w których poza zmienną zależną uwzględniony został jedynie udział zarejestrowanych osób bezrobotnych w liczbie osób w wieku produkcyjnym wśród mieszkańców gminy, wpływ tego czynnika na osiągnięcia jest bardzo silny. Wzrost wskaźnika o jeden punkt procentowy skutkuje spadkiem wyniku na obu częściach egzaminu o 0,52 punktu (dla części humanistycznej o 0,524, a na części matematyczno-przyrodniczej o 0,516). Daje to różnicę pomiędzy gminą o najniższym bezrobociu, a gminą o najwyższym bezrobociu wartości 10,00 punktów dla egzaminu z części humanistycznej oraz różnicę 9,85 punktu dla części matematyczno-przyrodniczej (na skali o średniej 100 i odchyleniu standardowym 15). Jest to około 2/3 odchylenia standardowego dla obu części egzaminu.

W modelach, w których kontrolowane są czynniki indywidualne oraz bezrobocie rodziców, wpływ bezrobocia na poziomie gminy na osiągnięcia na obu częściach egzaminu gimnazjalnego spada o ponad połowę (tabela 4, model 5; tabela 5, model 5). Kontrolowanie zmiennych standardowo ujmowanych w modelach EWD jeszcze bardziej ogranicza wpływ gminnego poziomu bezrobocia na wyniki w obu częściach egzaminu.

Jednakże lepsze rezultaty daje kontrola zarówno czynników indywidualnych, w tym bezrobocia rodziców, jak i zmiennych z modeli EWD (tabela 4, model 7; tabela 5, model 7). W takich modelach zmiana udziału bezrobotnych w liczbie osób w wieku produkcyjnym o jeden punkt procentowy powoduje spadek osiągnięć uczniów w części humanistycznej egzaminu o 0,12 punktu, zaś w części matematyczno-przyrodniczej o 0,13 punktu.

Warto również zauważyć, że w modelach tych jedynie bezrobocie matki ma negatywny wpływ na osiągnięcia uczniów na części humanistycznej egzaminu. Gimnazjaliści, których matki pozostawały bez zatrudnienia w trakcie ich nauki w pierwszej klasie, osiągnęły przeciętnie wyniki gorsze o 1,23 punktu (na skali o średniej 100 i odchyleniu standardowym 15).

Podsumowując, im wyższy poziom wskaźnika bezrobocia w gminie, tym niższe osiągnięcia gimnazjalistów. Dzięki stosowaniu modeli EWD wpływ ten można znacznie ograniczyć. Jeszcze lepsze efekty, dla kontroli tego zjawiska, dałoby wprowadzenie do modeli EWD zmiennych indywidualnych ucznia, takich jak inteligencja, status społeczno-ekonomiczny czy kompozycja rodziny, te jednak nigdy nie będą dostępne. Nawet w takiej sytuacji wpływ ten pozostałby prawdopodobnie znaczny i istotny. Szczegółowe zestawienie modeli zostały przedstawione w tabelach 4 i 5.

Tabela 2. Zamożność a wynik egzaminu gimnazjalnego w części humanistycznej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)
oszacowanie efektów stałych							
poziom ucznia							
sprawdzian					11,964	(0,250)	
sprawdzian ²					-0,412	(0,132)	
sprawdzian ³					-0,371	(0,082)	
pleć					1,786	(0,331)	
dysleksja na egz. gimnazja.					2,480	(0,624)	
dysleksja na spr. szk. podst.					-3,756	(0,915)	
test matryc Ravena (I pomiar)							7,065
wykształcenie rodziców							0,953
HTREI							0,043
zamożność gosp. domowego							1,905
wyposażenie w sprzęt elektro,							-0,472
wiek ucznia							-0,060
starszy niż główna kohorta							-8,255
młodszy niż główna kohorta							4,334
jedynak (ref.)							
najmłodszy							-1,687
średni							-3,040
najstarszy							-0,015
brak danych							-3,495
stała	100,771	(0,497)	95,537	(2,735)	101,497	(1,007)	103,357
poziom oddziału							
poziom szkoły							
wyd. gminy w 1000 na osobę			1,592	(0,865)	-0,163	(0,284)	-0,081
poziomo ubóstwa							
oszacowanie efektów losowych							
wariancja efektów szkół	19,31		17,36		5,59		5,71
wariancja efektów oddziałów	16,87		16,97		2,41		5,36
wariancja na poziomie ucznia	174,44		174,50		76,67		113,04
poziom oddziału							
współczynniki korelacji wewnątrzgrupowej							
poziom szkół	0,092		0,083		0,066		0,046
poziom oddziałów	0,080		0,081		0,028		0,043
poziom szkół i oddziałów łącznie	0,172		0,164		0,094		0,089
podsumowanie							
deviance	37115,95		37109,00		33201,26		34985,55
liczba szacowanych parametr.	4		5		11		17
liczba uczniów	4590		4590		4590		4590
liczba klas	291		291		291		291
liczba szkół	150		150		150		150

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
	9,967	(0,266)			11,949	(0,251)			9,966	(0,265)
	-0,395	(0,127)			-0,406	(0,130)			-0,397	(0,126)
	-0,351	(0,082)			-0,374	(0,083)			-0,357	(0,083)
	1,332	(0,311)			1,769	(0,330)			1,312	(0,310)
	1,571	(0,663)			1,976	(0,605)			1,250	(0,615)
	-2,890	(0,931)			-3,059	(0,781)			-2,401	(0,777)
(0,254)	2,107	(0,194)					7,065	(0,253)	2,115	(0,195)
(0,111)	0,428	(0,084)					0,951	(0,110)	0,432	(0,083)
(0,022)	0,032	(0,018)					0,043	(0,022)	0,030	(0,018)
(0,223)	1,133	(0,186)					1,906	(0,224)	1,147	(0,186)
(0,187)	-0,364	(0,151)					-0,781	(0,309)	-0,599	(0,254)
(0,055)	0,078	(0,041)					-0,060	(0,055)	0,075	(0,041)
(1,866)	-3,613	(1,586)					-8,257	(1,866)	-3,667	(1,599)
(1,992)	2,245	(2,072)					4,338	(1,999)	2,261	(2,075)
									-1,176	(0,439)
(0,484)	-1,196	(0,439)					-1,681	(0,487)	-2,010	(0,555)
(0,743)	-1,989	(0,544)					-3,032	(0,753)	-0,098	(0,463)
(0,519)	-0,105	(0,461)					-0,009	(0,523)	-0,437	(1,232)
(1,742)	-0,475	(1,244)					-3,494	(1,740)		
(1,585)	103,733	(1,065)	101,254	(0,502)	101,047	(0,360)	103,874	(0,640)	101,994	(0,596)
(0,435)	-0,593	(0,250)								
			-2,257	(0,461)	-0,220	(0,350)	0,001	(0,310)	0,428	(0,315)
	4,92		14,31		5,43		5,68		5,01	
	1,92		17,52		2,59		5,51		1,96	
	70,00		174,73		76,77		113,15		70,09	
	0,064		0,069		0,064		0,046		0,065	
	0,025		0,085		0,031		0,044		0,025	
	0,089		0,154		0,095		0,090		0,090	
	32771,78		32823,82		29367,73		30948,28		28980,44	
	23		5		11		17		23	
	4590		4590		4590		4590		4590	
	291		291		291		291		291	
	150		150		150		150		150	

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 3. Zamożność a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)
Oszacowanie efektów stałych							
Poziom ucznia							
sprawdzian					11,973	(0,269)	
sprawdzian ²					1,225	(0,116)	
sprawdzian ³					-0,399	(0,061)	
pleć					-0,861	(0,288)	
dysleksja na egz. gimnazjalnym					2,100	(0,839)	
interakcja dys. na spr. i egz.					-4,294	(0,882)	
test matryc Ravena (I pomiar)							8,140
wykształcenie rodziców							0,717
HTREI							0,063
zamożność gosp. domowego							0,956
wiek ucznia							-0,009
starszy niż główna kohorta							-3,215
młodszy niż główna kohorta							6,262
struktura rodziny (ref. pełna 2p)							
pełna 3 pokoleniowa (3p)							-0,397
niepełna dwupokoleniowa							-1,463
niepełna trzypokoleniowa							-1,700
inna							-0,964
brak danych							-8,370
stała	100,029	(0,513)	94,084	(3,212)	100,097	(1,199)	100,530
poziom oddziału							
poziom szkoły							
wyd. gminy w 1000 na osobę			1,808	(1,034)	-0,022	(0,362)	0,377
poziom ubóstwa							
oszacowanie efektów losowych							
wariancja efektów szkół	21,801		19,089		4,502		7,980
wariancja efektów oddziałów	20,978		21,185		5,429		5,634
wariancja na poziomie ucznia	159,830		159,892		66,335		97,948
poziom oddziału							
współczynniki korelacji wewnątrzgrupowej							
poziom szkół	0,108		0,095		0,059		0,072
poziom oddziałów	0,104		0,106		0,071		0,050
poziom szkół i oddziałów łącznie	0,212		0,201		0,130		0,122
podsumowanie							
deviance	34436,196		34428,408		30545,872		32198,544
liczba szacowanych parametrów	4		5		11		17
liczba uczniów	4591		4591		4591		4591
liczba klas	291		291		291		291
liczba szkół	150		150		150		150

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
	9,537	(0,289)			12,003	(0,279)			9,578	(0,300)
	1,235	(0,112)			1,236	(0,126)			1,266	(0,120)
	-0,373	(0,063)			-0,400	(0,063)			-0,372	(0,067)
	-1,479	(0,258)			-0,883	(0,305)			-1,496	(0,270)
	1,295	(0,786)			2,161	(0,899)			1,497	(0,872)
	-3,317	(0,846)			-4,347	(0,953)			-3,472	(0,938)
(0,256)	3,698	(0,192)					8,126	(0,252)	3,662	(0,199)
(0,100)	0,212	(0,078)					0,746	(0,106)	0,221	(0,080)
(0,019)	0,044	(0,018)					0,063	(0,020)	0,048	(0,019)
(0,184)	0,336	(0,164)					0,895	(0,191)	0,347	(0,173)
(0,052)	0,133	(0,042)					0,007	(0,055)	0,151	(0,044)
(1,262)	-2,164	(1,071)					-2,402	(1,248)	-1,623	(1,105)
(2,302)	4,726	(2,060)					6,538	(2,384)	4,851	(2,136)
(0,419)	0,281	(0,317)					-0,375	(0,440)	0,237	(0,330)
(0,737)	-1,148	(0,572)					-1,515	(0,772)	-1,259	(0,582)
(0,996)	-1,282	(0,869)					-1,835	(1,015)	-1,411	(0,892)
(1,667)	-0,125	(1,222)					-0,921	(1,685)	-0,150	(1,237)
(1,914)	-5,148	(1,637)					-8,284	(2,078)	-4,907	(1,793)
(1,909)	100,741	(1,054)	100,458	(0,552)	100,044	(0,344)	101,691	(0,486)	99,783	(0,426)
(0,592)	-0,239	(0,313)								
			-2,028	(0,527)	0,051	(0,298)	-0,094	(0,340)	0,384	(0,258)
	4,208		18,820		4,560		8,430		4,086	
	3,383		19,961		5,483		5,075		3,324	
	58,009		160,123		66,657		98,761		58,410	
					0,869		0,880		0,887	
	0,064		0,095		0,059		0,075		0,062	
	0,052		0,100		0,071		0,045		0,051	
	0,116		0,195		0,130		0,120		0,113	
	29939,042		32515,424		28869,418		30436,648		28297,880	
	23		5		11		17		23	
	4591		4591		4591		4591		4591	
	291		291		291		291		291	
	150		150		150		150		150	

Tabela 4. Bezrobocie a wynik egzaminu gimnazjalnego w części humanistycznej

oszacowanie efektów stałych		(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)
poziom uczenia															
sprawdzian						11,940	(0,249)					9,952	(0,265)	9,943	(0,280)
sprawdzian ²						-0,412	(0,132)					-0,403	(0,128)	-0,349	(0,134)
sprawdzian ³						-0,371	(0,082)					-0,354	(0,083)	-0,345	(0,087)
pleć						1,789	(0,330)					1,325	(0,310)	1,164	(0,311)
dysleksja na egz. gimnazjalnym						2,466	(0,626)					1,561	(0,666)	1,490	(0,656)
dysleksja na spr. szk. podst.						-3,769	(0,918)					-2,900	(0,932)	-2,989	(0,964)
test matryc Ravena (I pomiar)								7,049	(0,252)	6,930	(0,260)	2,104	(0,195)	2,041	(0,203)
wykształcenie rodziców								0,950	(0,109)	0,935	(0,109)	0,422	(0,083)	0,394	(0,088)
HTREI								0,043	(0,022)	0,041	(0,022)	0,030	(0,018)	0,029	(0,019)
zamożność gosp. domowego								1,874	(0,225)	1,866	(0,222)	1,117	(0,185)	1,187	(0,192)
wyposażenie w sprzęt elektro.								-0,505	(0,187)	-0,616	(0,198)	-0,383	(0,152)	-0,412	(0,160)
wiek ucznia								-0,060	(0,055)	-0,051	(0,057)	0,078	(0,041)	0,087	(0,042)
starszy niż główna kohorta								-8,317	(1,856)	-6,993	(1,956)	-3,651	(1,585)	-4,000	(1,894)
młodszy niż główna kohorta								4,353	(1,958)	4,695	(2,127)	2,286	(2,044)	2,585	(2,138)
jedynak (ref)														-1,095	(0,471)
najmłodszy								-1,668	(0,482)	-1,559	(0,520)	-1,154	(0,437)	-1,095	(0,471)
średni								-2,991	(0,743)	-2,800	(0,738)	-1,924	(0,543)	-1,818	(0,537)
najstarszy								0,012	(0,522)	0,022	(0,527)	-0,061	(0,462)	-0,089	(0,479)
brak danych								-3,479	(1,739)	-3,536	(1,943)	-0,422	(1,249)	-0,582	(1,351)
bezrobocie ojca										-0,922	(0,916)			-0,088	(0,711)
bezrobocie matki										-2,018	(0,755)			-1,271	(0,611)
stała		100,771	(0,497)	105,075	(1,003)	102,482	(0,590)	104,933	(0,908)	104,754	(0,974)	102,772	(0,737)	102,652	(0,779)

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)
poziom oddziały														
poziom szkoły														
udział bezrobotnych			-0,524	(0,106)	-0,186	(0,061)	-0,228	(0,069)	-0,214	(0,071)	-0,125	(0,058)	-0,117	(0,058)
oszacowanie efektów losowych														
wariancja efektów szkół	19,31		15,94		5,15		5,10		4,73		4,88		4,57	
wariancja efektów oddziałów	16,87		16,82		2,41		5,30		5,96		1,93		2,20	
wariancja na poziomie ucznia	174,44		174,48		76,69		113,06		111,68		70,01		69,30	
poziom oddziały														
współczynniki korelacji wewnątrzgrupowej														
poziom szkół	0,092		0,077		0,061		0,041		0,039		0,064		0,060	
poziom oddziałów	0,080		0,081		0,029		0,043		0,049		0,025		0,029	
poziom szkół i oddziałów łącznie	0,172		0,158		0,090		0,084		0,087		0,089		0,089	
podsumowanie														
deviance	37115,95		37100,61		33194,79		34977,86		32660,26		32772,23		30602,9	
liczba szacowanych parametr.	4		5		11		17		19		23		25	
liczba uczniów	4590		4590		4590		4590		4291		4590		4291	
liczba klas	291		291		291		291		291		291		291	
liczba szkół	150		150		150		150		150		150		150	

Tabela 5. Bezrobocie a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

oszacowanie efektów stałych		(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)
poziom uczenia															
sprawdzian						11,955	(0,266)					9,525	(0,288)	9,674	(0,288)
sprawdzian ²						1,226	(0,116)					1,233	(0,112)	1,253	(0,117)
sprawdzian ³						-0,398	(0,061)					-0,373	(0,063)	-0,407	(0,069)
pleć						-0,861	(0,288)					-1,479	(0,259)	-1,562	(0,258)
dysleksja na egz. gimnazjalnym						2,089	(0,836)					1,287	(0,790)	1,291	(0,753)
interakcja dys. na spr. i egz.						-4,311	(0,882)					-3,327	(0,849)	-3,462	(0,847)
test matryc Ravena (I pomiar)								8,132	(0,254)	8,107	(0,255)	3,694	(0,191)	3,702	(0,199)
wykształcenie rodziców								0,720	(0,099)	0,720	(0,098)	0,210	(0,077)	0,206	(0,078)
HTREI								0,064	(0,019)	0,063	(0,019)	0,044	(0,018)	0,042	(0,019)
zamożność gosp. domowego								0,930	(0,183)	0,925	(0,191)	0,319	(0,162)	0,374	(0,165)
wiek ucznia								-0,009	(0,052)	-0,010	(0,054)	0,132	(0,042)	0,137	(0,044)
starszy niż główna kohorta								-3,244	(1,256)	-2,275	(1,484)	-2,183	(1,069)	-1,804	(1,470)
młodszy niż główna kohorta								6,275	(2,302)	6,274	(2,389)	4,757	(2,054)	4,684	(2,065)
struktura rodziny (ref. pełna 2p)															
pełna 3 pokoleniowa (3p)								-0,429	(0,418)	-0,548	(0,414)	0,282	(0,316)	0,249	(0,325)
niepełna dwupokoleniowa								-1,461	(0,741)	-1,257	(1,145)	-1,159	(0,574)	-1,304	(0,838)
niepełna trzypokoleniowa								-1,735	(0,990)	-2,692	(1,318)	-1,315	(0,871)	-3,085	(1,116)
inna								-0,947	(1,669)	-1,039	(1,998)	-0,170	(1,211)	-0,736	(1,500)
brak danych								-8,391	(1,925)	-6,751	(2,138)	-5,156	(1,637)	-4,211	(1,857)

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)
bezrobocie ojca									-0,614	(0,786)			0,458	(0,573)
bezrobocie matki									-0,945	(0,593)			-0,614	(0,490)
stała	100,029	(0,513)	104,175	(1,137)	101,493	(0,747)	103,688	(0,874)	103,904	(0,878)	101,055	(0,706)	101,020	(0,706)
poziom oddziały														
poziom szkoły														
udział bezrobotnych			-0,505	(0,112)	-0,179	(0,077)	-0,232	(0,084)	-0,213	(0,082)	-0,134	(0,073)	-0,113	(0,070)
oszacowanie efektów losowych														
wariancja efektów szkół	21,801		18,701		4,138		7,562		7,435		3,999		3,788	
wariancja efektów oddziałów	20,978		20,889		5,398		5,532		5,603		3,389		3,672	
wariancja na poziomie ucznia	159,830		159,851		66,340		97,949		99,011		58,014		58,287	
poziom oddziały														
współczynniki korelacji wewnątrzgrupowej														
poziom szkół	0,108		0,094		0,055		0,068		0,066		0,061		0,058	
poziom oddziałów	0,104		0,105		0,071		0,050		0,050		0,052		0,056	
poziom szkół i oddziałów łącznie	0,212		0,199		0,126		0,118		0,116		0,113		0,114	
podsumowanie														
deviance	34436,196		34423,92		30539,89		32192,66		30113,45		29935,69		27985,50	
liczba szacowanych parametrów	4		5		11		17		19		23		25	
liczba uczniów	4591		4591		4591		4591		4292		4591		4292	
liczba klas	291		291		291		291		291		291		291	
liczba szkół	150		150		150		150		150		150		150	

7.4.3. Gospodarka

Zgodnie z wynikami wcześniejszych badań (Broomhall i Johnson, 1994; James, 2001), struktura lokalnej gospodarki, bez kontroli zmiennych na poziomie indywidualnym oraz zmiennych z modeli EWD, ma wpływ na poziom osiągnięć edukacyjnych na egzaminie gimnazjalnym w obu jego częściach (tabela 6, modele 3–8; tabela 7, modele 2–7). Wyższy udział pierwszego (rolnictwo, leśnictwo, łowiectwo i zbieractwo oraz wydobywanie) oraz drugiego sektora (przemysł i budownictwo) wpływają na obniżenie osiągnięć edukacyjnych. W przypadku wzrostu udziału sektora pierwszego w lokalnej gospodarce o 1%, następuje spadek osiągnięć na teście matematyczno-przyrodniczym o 0,18 punktu a na teście humanistycznym o 0,30 punktu (tabela 6, model 3; tabela 7, model 2). Wzrost udziału sektora drugiego w lokalnej gospodarce o 1% prowadzi do spadku osiągnięć na części matematyczno-przyrodniczej egzaminu o 0,22 punktu, a w części humanistycznej o 0,20 punktu (tabela 6, model 4; tabela 7, model 3).

Z kolei wzrost udziału trzeciego sektora oraz sektora HT w lokalnej gospodarce, wzrost liczby podmiotów gospodarczych w przeliczeniu na mieszkańca oraz wzrost liczby podmiotów HT na mieszkańca wpływa na podniesienie osiągnięć w obu częściach egzaminu. Przy wzroście udziału trzeciego sektora o 1%, wzrost osiągnięć na teście matematyczno-przyrodniczym można szacować na 0,17 punktu a w części humanistycznej o 0,21 punktu (tabela 6, model 5; tabela 7, model 4). Wzrost udziału sektora HT w lokalnej gospodarce przekłada się na wzrost osiągnięć szkolnych w obu testach odpowiednio o 0,51 punktu oraz 0,53 punktu (tabela 6, model 6; tabela 7, model 5). Wzrost liczby podmiotów gospodarczych w gminie o 1 na 100 mieszkańców przekłada się na wzrost osiągnięć na egzaminie odpowiednio o 0,62 punktu i 0,64 punktu (tabela 6, model 7; tabela 7, model 6). A wzrost liczby podmiotów HT o 1 na 100 mieszkańców gminy daje wzrost osiągnięć odpowiednio o 3,26 punktu i 3,40 punktu (tabela 6, model 8; tabela 7, model 7).

Wprowadzenie zmiennych kontrolnych z modeli EWD sprawia, że żaden z omówionych powyżej efektów nie jest istotny (tabela 6, modele 9–14; tabela 7, modele 8–13).

Kontrola zmiennych indywidualnych (IQ, SES, struktura rodziny) daje taki sam efekt – żaden ze wskaźników struktury lokalnej gospodarki nie wpływa na oszacowania osiągnięć na teście gimnazjalnym. Uwzględnienie w analizie jednocześnie zmiennych z modeli EWD oraz zmiennych indywidualnych powoduje, że jedynie liczba podmiotów HT na osobę w gminie związana jest istotnie z osiągnięciami na części humanistycznej egzaminu – wzrost liczby podmiotów o 1 na 100 mieszkańców obniża osiągnięcia o 0,70 punktu (tabela 6, model 15).

Przeprowadzone analizy prowadzą do wniosku, że struktura lokalnej gospodarki – bez kontroli zmiennych na poziomie jednostkowym – istotnie kształtuje osiągnięcia edukacyjne gimnazjalistów. Jednakże wprowadzenie zmiennych z modeli EWD lub zmiennych indywidualnych sprawia, że związek ten przestaje być istotny statystycznie. Modele EWD szacują efektywność gimnazjów, uwzględniając wpływ struktury lokalnej gospodarki. Dalszej analizy wymaga stwierdzone negatywny wpływ liczby podmiotów HT na osiągnięcia w części humanistycznej przy

kontroli zmiennych na poziomie jednostkowym. Zestawienie wyników omawianych modeli przedstawiają tabele 6 i 7.

7.4.4. Wydatki na oświatę

Bezpośrednie wydatki na oświatę, pomimo oczywistego związku ze stanem edukacji na danym terenie, zgodnie z dotychczasowym stanem wiedzy są słabo lub w ogóle niezwiązane z osiągnięciami edukacyjnymi. Biorąc pod uwagę wykorzystywany do tej pory w polskich badaniach wskaźnik wydatków na oświatę i wychowanie w budżecie samorządu terytorialnego (zrelatywizowany do liczby mieszkańców), można potwierdzić brak zależności między wydatkami na edukację a osiągnięciami gimnazjalistów. Brak jest również wpływu tego czynnika na różnice w osiągnięciach uczniów w modelach EWD (tabela 8, modele 6–7; tabela 9, modele 6–7).

Wykorzystując z kolei wskaźnik oparty na informacjach o wydatkach bieżących ponoszonych przez samorządy gminne na gimnazja (zrelatywizowane liczbą uczniów w szkołach tego typu), dla których organem założycielskim jest gmina (lub miasto na prawach powiatu), można zauważyć, że kontrolując zmienne indywidualne wraz ze wzrostem wydatków o 1000 złotych spada wynik uczniów na egzaminie z części humanistycznej przeciętnie o 0,13 punktu, a na części matematyczno-przyrodniczej o 0,14 punktu na skalach o średniej 100 i odchyleniu 15 (tabela 8, model 4; tabela 9, model 4). Przy zróżnicowaniu obecnym w badanej próbie różnica między wynikiem gminy o najwyższych i najniższych wydatkach będzie sięgać dla części humanistycznej 3,35 punktu (0,22 odchylenia standardowego), a dla części matematyczno-przyrodniczej 3,69 punktu (0,25 odchylenia standardowego).

W modelach analogicznych do EWD z uwzględnieniem zmiennych indywidualnych wpływ wydatków na gimnazja w przeliczeniu na uczniów nie ma znaczenia dla ich osiągnięć w części humanistycznej (tabela 8, model 5). Z kolei w części matematyczno-przyrodniczej wpływ ten jest negatywny – jednak nieco słabszy niż bez kontroli wcześniejszych osiągnięć. Każdy 1000 złotych więcej na gimnazjalistę wydane przez samorząd to 0,12 punktu mniej (tabela 9, model 5). Różnica pomiędzy gminą o najwyższych i najniższych wydatkach na osobę to 3,08 punktu (0,21 odchylenia standardowego).

W efekcie przeprowadzonych analiz można stwierdzić, że wydatki na oświatę nie mają zasadniczego znaczenia dla wyników uzyskiwanych przez uczniów na egzaminie gimnazjalnym. Nie wpływają one również modyfikująco na szacowane w modelach EWD wskaźniki efektywności pracy szkół. Przy obecnie wykorzystywanych informacjach nie ma konieczności ujmowania wydatków na oświatę tak w analizach osiągnięć gimnazjalistów, jak i w modelach EWD. Dalszej analizy wymaga negatywna zależność wydatków na gimnazja z osiągnięciami uczniów przy kontroli zmiennych indywidualnych (IQ, SES, struktura rodziny) oraz zmiennych z modeli EWD. Szczegółowe zestawienie wyników omówionych analiz przedstawiają tabele 8 i 9.

Tabela 6. Struktura lokalnej gospodarki a wynik egzaminu gimnazjalnego w części humanistycznej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)
oszacowanie efektów stałych																
poziom uczenia																
stała	100,463	(0,497)	100,771	(0,497)	102,183	(0,668)	105,801	(2,070)	86,228	(3,931)	96,475	(1,102)	95,104	(1,507)	97,985	(0,707)
poziom oddziały																
poziom szkoły																
udział i sektora w gospodarce					-0,308	(0,067)										
udział ii sektora w gospodarce							-0,200	(0,080)								
udział iii sektora w gospodarce									0,207	(0,056)						
udział sektora ht w gospodarce											0,527	(0,129)				
I, podm. gospod./100 mieszk.													0,644	(0,173)		
I, podm. HT/100 mieszk.															3,403	(0,803)
oszacowanie efektów losowych																
wariancja efektów szkół	17,96		19,31		16,75		17,34		15,50		14,29		13,32		12,45	
wariancja efektów oddziałów	20,20		16,87		16,94		16,90		16,96		17,09		17,02		17,18	
wariancja na poziomie ucznia	173,99		174,44		174,49		174,50		174,54		174,56		174,57		174,60	
poziom oddziały																
współczynniki korelacji wewnątrzgrupowej																
poziom szkół	0,085		0,092		0,080		0,083		0,075		0,069		0,065		0,061	
poziom oddziałów	0,095		0,080		0,081		0,081		0,082		0,083		0,083		0,084	
poziom szkół i oddziałów	0,180		0,172		0,162		0,164		0,157		0,152		0,148		0,145	
łącznie																
podsumowanie																
deviance	42415		37116		37106		37108		37101		37096		37091		37088	
liczba szacowanych parametr.	4		4		5		5		5		5		5		5	
liczba uczniów	5248		4590		4590		4590		4590		4590		4590		4590	
liczba klas	291		291		291		291		291		291		291		291	
liczba szkół	150		150		150		150		150		150		150		150	

oszacowanie efektów statycznych													
	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)	(13)	(s.e.)	(14)	(s.e.)	(s.e.)
poziom uczenia													
sprawdzian	11,953	(0,252)	11,959	(0,249)	11,954	(0,251)	11,951	(0,249)	11,951	(0,249)	11,948	(0,250)	
sprawdzian ²	-0,415	(0,132)	-0,415	(0,132)	-0,415	(0,132)	-0,416	(0,132)	-0,417	(0,133)	-0,417	(0,133)	
sprawdzian ³	-0,372	(0,082)	-0,372	(0,082)	-0,372	(0,082)	-0,372	(0,082)	-0,372	(0,082)	-0,372	(0,082)	
pleć	1,785	(0,333)	1,786	(0,333)	1,786	(0,333)	1,784	(0,332)	1,785	(0,333)	1,783	(0,333)	
dysleksja na egz. gimnazjalnym	2,534	(0,626)	2,526	(0,618)	2,528	(0,620)	2,506	(0,616)	2,498	(0,624)	2,497	(0,618)	
dysleksja na spr. szk. podst.	-3,834	(0,920)	-3,816	(0,911)	-3,825	(0,916)	-3,808	(0,909)	-3,800	(0,908)	-3,803	(0,909)	
stała	101,130	(0,431)	101,098	(1,105)	99,971	(2,651)	100,601	(0,806)	100,549	(0,865)	100,729	(0,522)	
poziom oddziały													
poziom szkoły													
udział I sektora w gospodarce	-0,036	(0,066)											
udział II sektora w gospodarce			-0,005	(0,045)									
udział III sektora w gospodarce					0,014	(0,036)							
udział sektora HT w gospodarce							0,045	(0,078)					
I. podm. gospod./100 mieszk.									0,048	(0,081)			
I. podm. HT/100 mieszk.											0,294	(0,371)	
oszacowanie efektów losowych													
wariancja efektów szkół	5,57		5,62		5,57		5,56		5,57		5,54		
wariancja efektów oddziałów	2,48		2,47		2,49		2,49		2,48		2,49		
wariancja na poziomie ucznia	76,62		76,62		76,62		76,62		76,62		76,62		
poziom oddziały													
współczynniki korelacji wewnątrzgrupowej													
poziom szkół	0,066		0,066		0,066		0,066		0,066		0,065		
poziom oddziałów	0,029		0,029		0,029		0,029		0,029		0,029		
poziom szkół i oddziałów													
łącznie	0,095		0,095		0,095		0,095		0,095		0,095		
podsumowanie													
deviance	33135		33136		33136		33135		33135		33135		
liczba szacowanych parametr.	11		11		11		11		11		11		
liczba uczniów	4590		4590		4590		4590		4590		4590		
liczba klas	291		291		291		291		291		291		
liczba szkół	150		150		150		150		150		150		

Tabela 6. Struktura lokalnej gospodarki a wynik egzaminu gimnazjalnego w części humanistycznej – cd.

	(15)	(s.e.)
oszacowanie efektów stałych		
poziom ucznia		
sprawdzian	9,972	(0,266)
sprawdzian ²	-0,397	(0,128)
sprawdzian ³	-0,353	(0,083)
pleć	1,332	(0,310)
dysleksja na egz, gimnazjalnym	1,624	(0,661)
dysleksja na spr, szk, podst,	-2,914	(0,921)
test matryc Ravena (I pomiar)	2,109	(0,194)
wykształcenie rodziców	0,436	(0,084)
HTREI	0,031	(0,018)
zamożność gosp, domowego	1,150	(0,185)
wyposażenie w sprzęt elektro,	-0,350	(0,153)
wiek ucznia	0,078	(0,041)
starszy niż główna kohorta	-3,564	(1,591)
młodszy niż główna kohorta	2,295	(2,080)
jedynak (ref.)		
najmłodszy	-1,212	(0,438)
średni	-2,023	(0,544)
najstarszy	-0,121	(0,462)
brak danych	-0,486	(1,238)
stała	102,384	(0,744)
poziom oddziału		
poziom szkoły		
I. podm. HT/100 mieszk.	-0,705	(0,341)
oszacowanie efektów losowych		
wariancja efektów szkół	4,93	
wariancja efektów oddziałów	1,92	
wariancja na poziomie ucznia	69,99	
poziom oddziału		
współczynniki korelacji wewnątrzgrupowej		
poziom szkół	0,064	
poziom oddziałów	0,025	
poziom szkół i oddziałów łącznie	0,089	
podsumowanie		
deviance	32771,30	
liczba szacowanych parametr.	23	
liczba uczniów	4590	
liczba klas	291	
liczba szkół	150	

Tabela 7. Struktura lokalnej gospodarki a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

oszacowanie efektów statych		(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)
poziom ucznia		100,029	(0,513)	100,764	(0,731)	105,880	(2,135)	87,534	(4,060)	96,070	(1,136)	94,593	(1,734)	97,384	(0,736)
poziom oddziału															
poziom szkoły															
udział I sektora w gospodarce				-0,160	(0,099)										
udział II sektora w gospodarce						-0,233	(0,077)								
udział III sektora w gospodarce								0,178	(0,059)						
udział sektora HT w gospodarce										0,485	(0,143)				
I. podm. gospod./100 mieszk.												0,612	(0,208)		
I. podm. HT/100 mieszk.														3,226	(0,934)
oszacowanie efektów losowych															
wariancja efektów szkół		21,801		20,995		19,225		18,934		17,384		16,268		15,589	
wariancja efektów oddziałów		20,978		21,021		20,913		20,977		21,161		21,023		21,126	
wariancja na poziomie ucznia		159,830		159,846		159,893		159,899		159,921		159,929		159,961	
poziom oddziału															
współczynniki korelacji wewnątrzgrupowej															
poziom szkół		0,108		0,104		0,096		0,095		0,088		0,082		0,079	
poziom oddziałów		0,104		0,104		0,105		0,105		0,107		0,107		0,107	
poziom szkół i oddziałów łącznie		0,212		0,208		0,201		0,200		0,195		0,189		0,186	
podsumowanie															
deviance		34436,196		34433,806		34427,252		34426,638		34421,738		34416,206		34414,624	
liczba szacowanych parametrów		4		5		5		5		5		5		5	
liczba uczniów		4591		4591		4591		4591		4591		4591		4591	
liczba klas		291		291		291		291		291		291		291	
liczba szkół		150		150		150		150		150		150		150	

Tabela 7. Struktura lokalnej gospodarki a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej – cd.

	(8)	(s.e.)	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)	(13)	(s.e.)
oszacowanie efektów stałych												
poziom ucznia												
sprawdzian	11,995	(0,268)	11,964	(0,269)	11,978	(0,268)	11,969	(0,266)	11,970	(0,269)	11,967	(0,268)
sprawdzian ²	1,226	(0,117)	1,223	(0,116)	1,225	(0,116)	1,224	(0,117)	1,224	(0,116)	1,223	(0,116)
sprawdzian ³	-0,400	(0,061)	-0,399	(0,061)	-0,399	(0,061)	-0,399	(0,061)	-0,399	(0,061)	-0,399	(0,061)
pleć	-0,858	(0,288)	-0,861	(0,288)	-0,861	(0,288)	-0,862	(0,288)	-0,861	(0,288)	-0,862	(0,289)
dysleksja na egz. gimnazjalnym	2,081	(0,833)	2,095	(0,836)	2,098	(0,836)	2,091	(0,837)	2,094	(0,850)	2,087	(0,843)
interakcja dys. na spr. i egz.	-4,246	(0,880)	-4,303	(0,881)	-4,286	(0,879)	-4,292	(0,882)	-4,292	(0,890)	-4,290	(0,885)
stała	99,564	(0,474)	100,930	(0,972)	100,875	(2,464)	99,872	(0,747)	99,936	(0,893)	99,917	(0,508)
poziom oddziału												
poziom szkoły												
udział I sektora w gospodarce	0,100	(0,098)										
udział II sektora w gospodarce			-0,036	(0,038)								
udział III sektora w gospodarce					-0,012	(0,033)						
udział sektora HT w gospodarce							0,019	(0,075)				
I. podm. gospod./100 mieszk.									0,010	(0,087)		
I. podm. HT/100 mieszk.											0,132	(0,394)
oszacowanie efektów losowych												
wariancja efektów szkół	4,363		4,397		4,513		4,473		4,487		4,461	
wariancja efektów oddziałów	5,347		5,450		5,418		5,441		5,437		5,446	
wariancja na poziomie ucznia	66,338		66,341		66,334		66,337		66,336		66,338	
poziom oddziału												
współczynniki korelacji wewnątrzgrupowej												
poziom szkół	0,057		0,058		0,059		0,059		0,059		0,059	
poziom oddziałów	0,070		0,072		0,071		0,071		0,071		0,071	
poziom szkół i oddziałów łącznie	0,127		0,130		0,130		0,130		0,130		0,130	
podsumowanie												
deviance	30542,166		30545,062		30545,704		30545,8		30545,858		30545,75	
liczba szacowanych parametrów	11		11		11		11		11		11	
liczba uczniów	4591		4591		4591		4591		4591		4591	
liczba klas	291		291		291		291		291		291	
liczba szkół	150		150		150		150		150		150	

7.4.5. Zasoby kulturowe

Przeprowadzone analizy wskazują, że wzrost liczby bibliotek na 1000 mieszkańców o jedną, bez kontroli innych zmiennych, powoduje spadek osiągnięć gimnazjalistów na testach zewnętrznych o 4,91 punktu w części humanistycznej oraz o 5,06 punktu w części matematyczno-przyrodniczej (tabela 10, model 2; tabela 11, model 2). Z kolei wielkość księgozbioru w przeliczeniu na 1000 mieszkańców nie ma związku z osiągnięciami uczniów (tabela 10, model 6; tabela 11, model 6). Kontrola wcześniejszych osiągnięć uczniów oraz płci i posiadania orzeczenia o dysleksji powoduje, że oba wskaźniki dostępności dóbr kulturowych nie mają znaczenia dla wyników gimnazjalnych (tabela 10, modele 3 i 7; tabela 11, modele 3 i 7).

Uwzględnienie w analizach zmiennych indywidualnych (IQ, SES, struktura rodziny), bez uwzględniania zmiennych z modeli EWD, powoduje, że wielkość księgozbioru ma pozytywny związek z wynikami uczniów w części humanistycznej egzaminu. Wzrost liczby książek w bibliotekach publicznych o jedną na mieszkańca związany jest ze wzrostem osiągnięć o 0,445 punktu na skali o średniej 100 i odchyleniu standardowym 15 (tabela 10, model 8).

Z kolei poddanie jednoczesnej kontroli tak zmiennych indywidualnych, jak i zmiennych z modeli EWD prowadzi do sytuacji, w której zarówno wielkość księgozbioru, jak i liczba bibliotek w przeliczeniu na mieszkańca gminy związana jest z osiągnięciami młodzieży. Siła relacji pomiędzy wskaźnikami dostępności kulturowej a testami jest większa dla części humanistycznej. Jedna biblioteka więcej na 1000 mieszkańców to wzrost wyniku uzyskiwanego przez gimnazjalistów w tej części o 2,69 punktu (tabela 10, model 5). Dla części matematyczno-przyrodniczej wzrost ten wynosi 1,53 punktu (tabela 11, model 5). Wzrost liczby książek w bibliotekach publicznych o 1 na mieszkańca powoduje wzrost wyników osiąganych przez uczniów w części matematyczno-przyrodniczej o 0,288 punktu, a w części humanistycznej o 0,529 punktu (tabela 10, model 9; tabela 11, model 9).

Z powyższych analiz można wyciągnąć kilka istotnych wniosków. Uwzględnianie w analizach liczby bibliotek na 1000 mieszkańców (na poziomie gmin) może prowadzić do błędnych wniosków. Kontrola zmiennych, które uwzględniane są w modelach EWD, pozwala na wyłączenie wpływu wskaźników dostępności kulturowej na osiągnięcia szkolne, a tym samym wskazuje na przewagę szacowania efektywności szkół modelami EWD w stosunku do modeli nieuwzględniających informacji o osiągnięciach na wejściu. Ponadto, przy kontroli zarówno zmiennych indywidualnych, jak i zmiennych z modeli EWD, istotny statystycznie pozostaje pozytywny związek pomiędzy dostępnością zasobów kulturowych a osiągnięciami gimnazjalistów, szczególnie w części matematyczno-przyrodniczej. Powinno być to przedmiotem dalszych analiz, gdyż może wskazywać na rosnące znaczenie tych zasobów w kształceniu wraz z wiekiem uczniów. Zestawienie omówionych analiz prezentują tabele 10 i 11.

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 8. Wydatki na oświatę a wynik egzaminu gimnazjalnego w części humanistycznej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	
oszacowanie efektów stałych									
poziom ucznia									
sprawdzian					11,960	(0,251)			
sprawdzian ²					-0,413	(0,132)			
sprawdzian ³					-0,371	(0,082)			
pleć					1,780	(0,331)			
dysleksja na egz. gimnazjalnym					2,470	(0,623)			
dysleksja na spr. szk. podst.					-3,754	(0,914)			
test matryc Ravena (I pomiar)							7,069	(0,255)	
wykształcenie rodziców							0,954	(0,110)	
HTREI							0,042	(0,022)	
zamożność gosp. domowego							1,911	(0,223)	
wyposażenie w sprzęt elektro.							-0,468	(0,187)	
wiek ucznia							-0,060	(0,055)	
starszy niż główna kohorta							-8,280	(1,864)	
młodszy niż główna kohorta							4,365	(1,995)	
jedynak (ref.)									
najmłodszy							-1,663	(0,487)	
średni							-3,038	(0,741)	
najstarszy							-0,004	(0,521)	
brak danych							-3,441	(1,747)	
stała	100,771	(0,497)	100,845	(0,893)	101,275	(0,572)	104,098	(0,841)	
poziom oddziału									
poziom szkoły									
wydat. w tys. na gimnazja/ucznia			-0,009	(0,102)	-0,038	(0,046)	-0,126	(0,057)	
wydat. na oświatę (w 100 zł/osobę)									
oszacowanie efektów losowych									
wariancja efektów szkół	19,31		19,32		5,58		5,64		
wariancja efektów oddziałów	16,87		16,86		2,41		5,24		
wariancja na poziomie ucznia	174,44		174,45		76,67		113,04		
poziom oddziału									
współczynniki korelacji wewnątrzgrupowej									
poziom szkół	0,092		0,092		0,066		0,045		
poziom oddziałów	0,080		0,080		0,028		0,042		
poziom szkół i oddziałów łącznie	0,172		0,172		0,094		0,088		
podsumowanie									
deviance	37115,95		37115,94		33201,20		34983,01		
liczba szacowanych parametr.	4		5		11		17		
liczba uczniów	4590		4590		4590		4590		
liczba klas	291		291		291		291		
liczba szkół	150		150		150		150		

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
	9,956	(0,267)			11,960	(0,250)			9,961	(0,266)
	-0,402	(0,128)			-0,417	(0,132)			-0,407	(0,128)
	-0,353	(0,083)			-0,373	(0,082)			-0,355	(0,083)
	1,312	(0,311)			1,783	(0,332)			1,321	(0,311)
	1,552	(0,661)			2,470	(0,629)			1,559	(0,669)
	-2,889	(0,927)			-3,788	(0,918)			-2,920	(0,932)
	2,111	(0,196)					7,057	(0,254)	2,104	(0,195)
	0,423	(0,084)					0,951	(0,109)	0,421	(0,083)
	0,030	(0,018)					0,043	(0,022)	0,031	(0,018)
	1,136	(0,185)					1,899	(0,224)	1,129	(0,186)
	-0,365	(0,152)					-0,471	(0,187)	-0,367	(0,152)
	0,078	(0,041)					-0,060	(0,055)	0,078	(0,041)
	-3,637	(1,582)					-8,300	(1,865)	-3,646	(1,586)
	2,290	(2,069)					4,393	(1,981)	2,311	(2,056)
	-1,149	(0,438)					-1,698	(0,481)	-1,172	(0,437)
	-1,947	(0,543)					-3,038	(0,741)	-1,947	(0,542)
	-0,069	(0,461)					-0,023	(0,521)	-0,080	(0,462)
	-0,398	(1,246)					-3,459	(1,758)	-0,406	(1,258)
	102,407	(0,800)	105,448	(3,187)	103,168	(1,289)	105,789	(1,690)	103,727	(1,216)
	-0,080	(0,051)								
			-0,423	(0,281)	-0,199	(0,108)	-0,002	(0,001)	-0,002	(0,001)
	5,01		18,65		5,43		5,41		4,94	
	1,94		16,75		2,39		5,34		1,93	
	70,00		174,46		76,68		113,06		70,01	
	0,065		0,089		0,064		0,044		0,064	
	0,025		0,080		0,028		0,043		0,025	
	0,090		0,169		0,093		0,087		0,089	
	32773,983		37112,438		33198,726		34982,38		32773,09	
	23		5		11		17		23	
	4590		4590		4590		4590		4590	
	291		291		291		291		291	
	150		150		150		150		150	

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 9. Wydatki na oświatę a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)
oszacowanie efektów stałych								
poziom ucznia								
sprawdzian					11,973	(0,269)		
sprawdzian ²					1,225	(0,117)		
sprawdzian ³					-0,399	(0,061)		
pleć					-0,870	(0,287)		
dysleksja na egz. gimnazjalnym					2,088	(0,835)		
interakcja dys. na spr. i egz.					-4,294	(0,880)		
test matryc Ravena (I pomiar)							8,148	(0,257)
wykształcenie rodziców							0,723	(0,101)
HTREI							0,063	(0,019)
zamożność gosp. domowego							0,964	(0,184)
wiek ucznia							-0,009	(0,052)
starszy niż główna kohorta							-3,240	(1,262)
młodszy niż główna kohorta							6,267	(2,310)
struktura rodziny (ref. pełna 2p)								
pełna 3 pokoleniowa (3p)							-0,427	(0,418)
niepełna dwupokoleniowa							-1,451	(0,736)
niepełna trzypokoleniowa							-1,711	(0,993)
inna							-0,945	(1,666)
brak danych							-8,404	(1,913)
stała	100,029	(0,513)	100,136	(0,821)	100,662	(0,677)	102,868	(0,637)
poziom oddziału								
poziom szkoły								
wydat. w tys. na gimnazja/ucznia			-0,013	(0,087)	-0,079	(0,064)	-0,135	(0,052)
wydat. w tys. na oświatę/osobę								
oszacowanie efektów losowych								
wariancja efektów szkół	21,801		21,808		4,411		8,034	
wariancja efektów oddziałów	20,978		20,97		5,429		5,486	
wariancja na poziomie ucznia	159,830		159,83		66,338		97,952	
poziom oddziału								
współczynniki korelacji wewnątrzgrupowej								
poziom szkół	0,108		0,108		0,058		0,072	
poziom oddziałów	0,104		0,104		0,071		0,049	
poziom szkół i oddziałów łącznie	0,212		0,212		0,129		0,121	
podsumowanie								
deviance	34436,20		34434,02		30545,01		32197,90	
liczba szacowanych parametrów	4		5		11		17	
liczba uczniów	4591		4591		4591		4591	
liczba klas	291		291		291		291	
liczba szkół	150		150		150		150	

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
	9,529	(0,288)			11,972	(0,268)			9,534	(0,289)
	1,234	(0,112)			1,223	(0,116)			1,230	(0,112)
	-0,373	(0,063)			-0,400	(0,061)			-0,374	(0,063)
	-1,496	(0,255)			-0,862	(0,289)			-1,482	(0,258)
	1,272	(0,784)			2,096	(0,839)			1,290	(0,791)
	-3,312	(0,844)			-4,312	(0,885)			-3,331	(0,850)
	3,704	(0,193)					8,139	(0,256)	3,696	(0,192)
	0,211	(0,078)					0,720	(0,100)	0,210	(0,078)
	0,043	(0,018)					0,064	(0,019)	0,044	(0,018)
	0,342	(0,164)					0,954	(0,184)	0,334	(0,164)
	0,133	(0,042)					-0,009	(0,052)	0,133	(0,042)
	-2,190	(1,067)					-3,242	(1,261)	-2,176	(1,070)
	4,763	(2,066)					6,276	(2,318)	4,753	(2,064)
	0,282	(0,315)					-0,421	(0,416)	0,288	(0,314)
	-1,152	(0,571)					-1,455	(0,735)	-1,155	(0,571)
	-1,305	(0,871)					-1,689	(0,995)	-1,284	(0,872)
	-0,177	(1,214)					-0,943	(1,655)	-0,166	(1,210)
	-5,169	(1,636)					-8,418	(1,914)	-5,166	(1,635)
	100,929	(0,643)	103,981	(3,696)	101,283	(1,335)	103,798	(2,110)	100,970	(1,174)
	-0,119	(0,051)								
			-3,577	(3,250)	-1,137	(1,160)	-1,827	(1,836)	-0,913	(0,999)
	4,037		21,374		4,465		8,025		4,169	
	3,373		20,876		5,415		5,563		3,400	
	58,017		159,833		66,335		97,949		58,012	
	0,062		0,106		0,059		0,072		0,064	
	0,052		0,103		0,071		0,050		0,052	
	0,114		0,209		0,130		0,122		0,116	
	29939,00		34436,19		30544,64		32196,88		29936,16	
	23		5		11		17		23	
	4591		4591		4591		4591		4591	
	291		291		291		291		291	
	150		150		150		150		150	

Tabela 10. Dostęp do dóbr kultury a wynik egzaminu gimnazjalnego w części humanistycznej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	
oszacowanie efektów stałych									
poziom ucznia									
sprawdzian					11,970	(0,249)			
sprawdzian ²					-0,413	(0,132)			
sprawdzian ³					-0,372	(0,082)			
pleć					1,781	(0,331)			
dysleksja na egz. gimnazjalnym					2,498	(0,620)			
dysleksja na spr. szk. podst.					-3,767	(0,908)			
test matryc Ravena (I pomiar)							7,072	(0,254)	
wykształcenie rodziców							0,960	(0,110)	
HTREI							0,042	(0,022)	
zamożność gosp. domowego							1,920	(0,223)	
wyposażenie w sprzęt elektro.							-0,457	(0,187)	
wiek ucznia							-0,059	(0,055)	
starszy niż główna kohorta							-8,217	(1,865)	
młodszy niż główna kohorta							4,378	(1,996)	
jedynak (ref.)									
najmłodszy							-1,703	(0,486)	
średni							-3,075	(0,744)	
najstarszy							-0,038	(0,522)	
brak danych							-3,537	(1,739)	
stała	100,771	0,497	102,412	0,824	100,673	(0,557)	102,583	(0,693)	
poziom oddziału									
poziom szkoły									
biblioteki/1000 mieszkańców			-4,908	1,485	0,872	(1,100)	1,584	(1,083)	
księgozbiór na mieszkańca									
oszacowanie efektów losowych									
wariancja efektów szkół	19,31		17,67		5,58		5,59		
wariancja efektów oddziałów	16,87		16,87		2,40		5,34		
wariancja na poziomie ucznia	174,44		174,48		76,67		113,03		
poziom oddziału									
współczynniki korelacji wewnątrzgrupowej									
poziom szkół	0,092		0,085		0,066		0,045		
poziom oddziałów	0,080		0,081		0,028		0,043		
poziom szkół i oddziałów łącznie	0,172		0,165		0,094		0,088		
podsumowanie									
deviance	37115,95		37109,25		33200,78		34983,76		
liczba szacowanych parametr.	4		5		11		17		
liczba uczniów	4590		4590		4590		4590		
liczba klas	291		291		291		291		
liczba szkół	150		150		150		150		

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
	9,972	(0,266)			11,976	(0,248)			9,968	(0,263)
	-0,399	(0,128)			-0,416	(0,132)			-0,406	(0,129)
	-0,353	(0,083)			-0,372	(0,082)			-0,354	(0,083)
	1,315	(0,311)			1,791	(0,330)			1,335	(0,310)
	1,622	(0,647)			2,558	(0,607)			1,664	(0,624)
	-2,924	(0,910)			-3,776	(0,901)			-2,913	(0,896)
	2,113	(0,195)					7,066	(0,254)	2,105	(0,193)
	0,434	(0,083)					0,963	(0,110)	0,432	(0,083)
	0,030	(0,018)					0,041	(0,022)	0,029	(0,018)
	1,155	(0,185)					1,923	(0,224)	1,149	(0,185)
	-0,345	(0,153)					-0,458	(0,187)	-0,355	(0,153)
	0,079	(0,041)					-0,060	(0,055)	0,078	(0,041)
	-3,562	(1,588)					-8,269	(1,869)	-3,629	(1,584)
	2,342	(2,074)					4,394	(1,992)	2,346	(2,064)
	-1,190	(0,439)					-1,693	(0,486)	-1,167	(0,440)
	-2,004	(0,545)					-3,064	(0,744)	-1,969	(0,545)
	-0,111	(0,462)					-0,037	(0,525)	-0,095	(0,463)
	-0,495	(1,239)					-3,577	(1,745)	-0,516	(1,238)
	100,894	(0,660)	100,705	(1,505)	99,370	(0,976)	101,386	(0,884)	99,715	(0,861)
	2,693	(1,039)								
			0,017	(0,386)	0,409	(0,222)	0,445	(0,189)	0,529	(0,183)
	4,73		19,31		5,27		5,32		4,51	
	1,93		16,87		2,35		5,27		1,89	
	69,99		174,44		76,67		113,04		69,99	
	0,062		0,092		0,063		0,043		0,059	
	0,025		0,080		0,028		0,043		0,025	
	0,087		0,172		0,090		0,086		0,084	
	32767,672		37115,945		33194,631		34979,40		32762,52	
	23		5		11		17		23	
	4590		4590		4590		4590		4590	
	291		291		291		291		291	
	150		150		150		150		150	

Tabela 11. Dostęp do dóbr kultury a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)
oszacowanie efektów stałych								
poziom ucznia								
sprawdzian					11,975	(0,268)		
sprawdzian ²					1,225	(0,117)		
sprawdzian ³					-0,399	(0,061)		
pleć					-0,862	(0,288)		
dysleksja na egz. gimnazjalnym					2,104	(0,835)		
interakcja dys. na spr. i egz.					-4,297	(0,881)		
test matryc Ravena (I pomiar)							8,143	(0,257)
wykształcenie rodziców							0,721	(0,100)
HTREI							0,063	(0,019)
zamożność gosp. domowego							0,957	(0,183)
wiek ucznia							-0,009	(0,052)
starszy niż główna kohorta							-3,217	(1,262)
młodszy niż główna kohorta							6,243	(2,307)
struktura rodziny (ref. pełna 2p)								
pełna 3 pokoleniowa (3p)							-0,414	(0,420)
niepełna dwupokoleniowa							-1,455	(0,736)
niepełna trzypokoleniowa							-1,694	(0,993)
inna							-0,923	(1,666)
brak danych							-8,378	(1,915)
stała	100,029	(0,513)	101,93	(0,846)	99,949	(0,443)	101,805	(0,589)
poziom oddziału								
poziom szkoły								
biblioteki/1000 mieszkańców			-5,685	(1,409)	0,223	(0,895)	-0,085	(1,004)
księgozbiów na mieszkańca								
oszacowanie efektów losowych								
wariancja efektów szkół	21,801		19,566		4,505		8,161	
wariancja efektów oddziałów	20,978		20,964		5,426		5,587	
wariancja na poziomie ucznia	159,830		159,867		66,335		97,943	
poziom oddziału								
współczynniki korelacji wewnątrzgrupowej								
poziom szkół	0,108		0,098		0,059		0,073	
poziom oddziałów	0,104		0,105		0,071		0,050	
poziom szkół i oddziałów łącznie	0,212		0,203		0,13		0,123	
podsumowanie								
deviance	34436,20		34428,35		30545,83		32199,40	
liczba szacowanych parametrów	4		5		11		17	
liczba uczniów	4591		4591		4591		4591	
liczba klas	291		291		291		291	
liczba szkół	150		150		150		150	

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
	9,539	(0,288)			11,981	(0,268)			9,539	(0,287)
	1,233	(0,112)			1,223	(0,116)			1,230	(0,112)
	-0,373	(0,063)			-0,399	(0,061)			-0,374	(0,063)
	-1,487	(0,257)			-0,858	(0,289)			-1,478	(0,259)
	1,318	(0,781)			2,147	(0,827)			1,345	(0,771)
	-3,332	(0,841)			-4,308	(0,876)			-3,330	(0,831)
	3,700	(0,192)					8,144	(0,257)	3,697	(0,192)
	0,216	(0,078)					0,727	(0,102)	0,217	(0,078)
	0,043	(0,018)					0,063	(0,019)	0,043	(0,018)
	0,346	(0,164)					0,966	(0,183)	0,346	(0,164)
	0,133	(0,042)					-0,008	(0,052)	0,133	(0,042)
	-2,150	(1,073)					-3,232	(1,267)	-2,177	(1,076)
	4,763	(2,058)					6,272	(2,31)	4,775	(2,062)
	0,271	(0,317)					-0,433	(0,419)	0,272	(0,317)
	-1,137	(0,571)					-1,446	(0,737)	-1,143	(0,573)
	-1,286	(0,874)					-1,688	(0,995)	-1,281	(0,873)
	-0,117	(1,221)					-0,864	(1,674)	-0,086	(1,230)
	-5,160	(1,639)					-8,324	(1,918)	-5,087	(1,645)
	99,538	(0,469)	100,191	(1,559)	98,939	(0,684)	100,674	(0,953)	98,726	(0,620)
	1,266	(0,788)								
			-0,042	(0,419)	0,279	(0,159)	0,285	(0,233)	0,318	(0,130)
	4,155		21,783		4,413		8,043		4,063	
	3,377		20,983		5,362		5,536		3,339	
	58,009		159,831		66,331		97,94		58,006	
	0,063		0,108		0,058		0,072		0,062	
	0,052		0,104		0,070		0,050		0,051	
	0,115		0,212		0,128		0,122		0,113	
	29937,89		34436,18		30542,84		32197,29		29934,93	
	23		5		11		17		23	
	4591		4591		4591		4591		4591	
	291		291		291		291		291	
	150		150		150		150		150	

7.4.6. Opieka przedszkolna

Analizując modele, do których poza zmienną niezależną (wyniki na matematyczno-przyrodniczej lub humanistycznej części egzaminu) wprowadzono jedynie informację o udziale dzieci w wieku 3–5 lat objętych opieką przedszkolną na poziomie gminy w roku 2003 (rok po opuszczeniu przez gimnazjalistów zerówki – najwcześniejsze dane dostępne w Banku Danych Lokalnych GUS), można wskazać, że wzrost udziału dzieci objętych opieką o jeden punkt procentowy wiąże się z podwyższeniem osiągnięć na egzaminie gimnazjalnym w części matematyczno-przyrodniczej i humanistycznej o odpowiednio 0,10 i 0,11 punktu (tabela 12, model 2; tabela 13, model 2). Po wprowadzeniu zmiennych kontrolnych z modeli EWD wpływ ten okazuje się nieistotny statystycznie dla obu części egzaminu (tabela 12, modele 3–4; tabela 13, modele 3–4). W modelu, w którym kontroli poddano zmienne indywidualne (IQ, SES, struktura rodziny), związek ten także jest nieistotny statystycznie (tabela 12, modele 5–6; tabela 13, modele 5–6).

Brak go również w modelach, w których kontrolowane były zarówno zmienne indywidualne, jak i zmienne z modeli EWD (tabela 12, modele 7–8; tabela 13, modele 7–8).

Wprowadzenie do analizy informacji o tym, ile lat uczniów uczęszczało do przedszkola przed rozpoczęciem nauki w szkole podstawowej (łącznie ze zmiennymi indywidualnymi oraz informacją o poziomie objęcia opieką przedszkolną w gminie), pozwala stwierdzić, że każdy rok w przedszkolu dłużej wiąże się z osiągnięciami w części humanistycznej niższymi o 0,45 punktu na skali o średniej 100 i odchyleniu standardowym 15 (tabela 12, model 6). Brak jest takiej relacji dla części matematyczno-przyrodniczej egzaminu (tabela 13, model 6). Kontrolując dodatkowo zmienne z modeli EWD, nadal można stwierdzić istnienie negatywnego związku, jednak jego siła jest słabsza – jeden rok w przedszkolu wiąże się z osiągnięciami niższymi o 0,32 punktu w części humanistycznej (tabela 12, model 9).

Lepsze zrozumienie tego zjawiska jest możliwe po wprowadzeniu do modelu interakcji pomiędzy wykształceniem rodziców a przebywaniem dziecka w przedszkolu. Interakcja jest istotna w części humanistycznej, zarówno bez uwzględniania zmiennych z modeli EWD, jak i przy ich kontroli. W modelu bez zmiennych zawartych w modelu EWD uczęszczanie do przedszkola nie ma znaczenia, a interakcja wskazuje, że wyniki na teście rosną wraz ze spadkiem wykształcenia rodziców i jednocześnie wraz z dłuższym uczęszczaniem do przedszkola. Jest to spójne z omawianymi wcześniej wynikami badań kanadyjskich. Relacja ta utrzymuje się również w modelach, w których kontroli poddane są zmienne z modeli EWD. W modelach tych dodatkowo na granicy istotności znajduje się uczęszczanie do przedszkola. Każdy rok dłużej w przedszkolu to 1,39 punktu więcej na humanistycznej części egzaminu gimnazjalnego (tabela 12, model 9).

Podsumowując, można stwierdzić, że powszechność opieki przedszkolnej na poziomie gminy ma wpływ na osiągnięcia gimnazjalistów. Wpływ ten jest kontrolowany w modelach EWD. Dzięki czemu wskaźniki EWD gimnazjów nie są obciążone tym czynnikiem.

Innym zagadnieniem jest związek pomiędzy uczęszczaniem do przedszkola a osiągnięciami na egzaminie gimnazjalnym. Przeprowadzone analizy pozwalają stwierdzić, że uczęszczanie do przedszkola ma pozytywny wpływ na osiągnięcia uczniów z rodzin, których rodzice są gorzej wykształceni, a którzy dłużej uczęszczali do przedszkola. Można zatem mówić o wyrównującym oddziaływaniu przedszkola w stosunku do dzieci objętych tego typu opieką w zakresie umiejętności humanistycznych. Co więcej, kontrola zmiennych z modeli EWD pozwala stwierdzić, że dłuższe uczęszczanie do przedszkola podnosi osiągnięcia uczniów w części humanistycznej egzaminu, przy kontroli interakcji uczęszczania do przedszkola z wykształceniem rodziców. Zestawienie analizowanych modeli przedstawiają tabele 12 i 13.

7.4.7. Czynniki kontrolne

Wielkość gminy może być uznana za wskaźnik przede wszystkim na skali metropolia – peryferia. Większe gminy mają ogólnie większy potencjał, są miastami a gminy bardzo duże – miastami na prawach powiatu. Jeśli do modelu wprowadzimy jedynie informację o wielkości gminy, w której znajduje się szkoła badanych uczniów, to relacja z osiągnięciami jest pozytywna (tabela 14, model 2; tabela 15, model 2). Im w gminie mieszka więcej osób, tym wyższe osiągnięcia mają uczniowie w obu częściach egzaminu gimnazjalnego (w przypadku części matematyczno-przyrodniczej zależność jest na granicy istotności). Wzrost wielkości o jeden procent przekłada się na wzrost osiągnięć na części humanistycznej niemal o 0,009 punktu, zaś w części matematyczno przyrodniczej o 0,008 punktu. Wprowadzenie do modeli zmiennych EWD wytrąca związek z wielkością gminy obu części egzaminu (tabela 14, model 3; tabela 15, model 3).

Przy kontroli zmiennych indywidualnych wielkość gminy nie ma znaczenia (tabela 14, model 4; tabela 15, model 4). Kontrolowanie dodatkowo zmiennych z modeli EWD wskazuje, że wielkość gminy związana jest z wynikami na części humanistycznej (tabela 14, model 5). Związek ten jest negatywny. Wzrost wielkości gminy o jeden procent to spadek osiągnięć gimnazjalistów na tej części testu o 0,005 punktu na skali o średniej 100 i odchyleniu standardowym 15.

Można zatem stwierdzić, że bez kontroli zmiennych z poziomu jednostkowego wielkość gminy wpływa pozytywnie na osiągnięcia uczniów. Zastosowanie modelu EWD pozwala kontrolować ten czynnik, dzięki czemu wskaźniki EWD szkół nie są obciążone ze względu na wielkość gmin, w których się znajdują. Jednocześnie, przy kontroli zmiennych indywidualnych w modelach, w których uwzględnione zostały zmienne EWD, zwiększenie gminy powoduje spadek osiągnięć uczniów w części humanistycznej. Szczegółowe zestawienie modeli prezentują tabele 14 i 15.

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 12. Powszechność opieki przedszkolnej i uczęszczanie do przedszkola a wynik egzaminu gimnazjalnego w części humanistycznej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)
oszacowanie efektów stałych								
poziom ucznia								
sprawdzian					11,953	(0,250)	11,920	(0,248)
sprawdzian ²					-0,415	(0,132)	-0,412	(0,132)
sprawdzian ³					-0,372	(0,082)	-0,370	(0,082)
pleć					1,787	(0,334)	1,778	(0,333)
dysleksja na egz. gimnazjalnym					2,522	(0,614)	2,488	(0,625)
dysleksja na spr. szk. podst.					-3,814	(0,910)	-3,801	(0,919)
test matryc Ravena (I pomiar)								
wykształcenie rodziców								
HTREI								
zamożność gosp. domowego								
wyposażenie w sprzęt elektro.								
wiek ucznia								
starszy niż główna kohorta								
młodszy niż główna kohorta								
jedynak (ref.)								
najmłodszy								
średni								
najstarszy								
brak danych								
liczba lat w przedszkolu							0,255	(0,124)
interakcja wyksz. i l. lat w przedsz.								
stała	100,793	(0,497)	97,374	(0,986)	100,756	(0,739)	100,692	(0,738)
poziom oddziału								
poziom szkoły								
udział dzieci 3–5 lat w przedszk.			0,107	(0,026)	0,007	(0,016)	0,000	(0,017)
oszacowanie efektów losowych								
wariancja efektów szkół	18,97		14,24		5,59		5,56	
wariancja efektów oddziałów	17,24		17,34		2,48		2,49	
wariancja na poziomie ucznia	174,06		174,12		76,62		76,55	
poziom oddziału								
współczynniki korelacji wewnątrzgrupowej								
poziom szkół	0,090		0,069		0,066		0,066	
poziom oddziałów	0,082		0,084		0,029		0,029	
poziom szkół i oddziałów łącznie	0,172		0,154		0,095		0,095	
podsumowanie								
deviance	37034,90		37014,26		33135,51		33131,37	
liczba szacowanych parametr.	4		5		11		12	
liczba uczniów	4581		4581		4581		4581	
liczba klas	291		291		291		291	
liczba szkół	150		150		150		150	

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
					9,977	(0,265)	9,969	(0,268)	9,951	(0,266)
					-0,404	(0,127)	-0,407	(0,128)	-0,388	(0,127)
					-0,353	(0,083)	-0,354	(0,083)	-0,348	(0,082)
					1,322	(0,314)	1,336	(0,314)	1,353	(0,314)
					1,609	(0,658)	1,616	(0,648)	1,704	(0,658)
					-2,935	(0,920)	-2,926	(0,908)	-3,019	(0,918)
	7,057	(0,256)	7,073	(0,256)	2,106	(0,194)	2,121	(0,194)	2,125	(0,195)
	0,951	(0,110)	0,993	(0,112)	0,432	(0,084)	0,462	(0,084)	0,599	(0,113)
	0,042	(0,022)	0,046	(0,022)	0,030	(0,018)	0,032	(0,018)	0,037	(0,018)
	1,900	(0,225)	1,918	(0,226)	1,142	(0,185)	1,154	(0,186)	1,135	(0,185)
	-0,465	(0,189)	-0,424	(0,186)	-0,341	(0,154)	-0,313	(0,152)	-0,320	(0,153)
	-0,058	(0,056)	-0,066	(0,056)	0,078	(0,041)	0,072	(0,041)	0,072	(0,041)
	-8,231	(1,891)	-8,196	(1,880)	-3,629	(1,622)	-3,604	(1,606)	-3,513	(1,600)
	4,348	(1,993)	4,409	(2,058)	2,277	(2,073)	2,320	(2,111)	2,285	(2,103)
	-1,666	(0,482)	-1,748	(0,483)	-1,192	(0,436)	-1,249	(0,436)	-1,212	(0,435)
	-3,014	(0,731)	-3,142	(0,735)	-2,024	(0,541)	-2,115	(0,544)	-2,085	(0,542)
	-0,003	(0,519)	-0,048	(0,517)	-0,111	(0,462)	-0,142	(0,462)	-0,115	(0,459)
	-3,483	(1,744)	-3,711	(1,738)	-0,479	(1,235)	-0,646	(1,231)	-0,672	(1,218)
			-0,453	(0,180)			-0,320	(0,131)	1,162	(0,594)
									-0,114	(0,046)
	103,109	(1,023)	103,451	(1,039)	102,539	(0,929)	102,776	(0,937)	102,852	(0,935)
	-0,001	(0,018)	0,008	(0,019)	-0,023	(0,016)	-0,016	(0,016)	-0,017	(0,016)
	5,61		5,72		4,98		5,00		4,93	
	5,43		5,40		1,94		1,92		1,94	
	113,04		112,81		70,00		69,90		69,80	
	0,045		0,046		0,065		0,065		0,064	
	0,044		0,044		0,025		0,025		0,025	
	0,089		0,090		0,090		0,090		0,090	
	34917,22		34909,24		32709,35		32702,91		32695,93	
	17		18		23		24		25	
	4581		4581		4581		4581		4581	
	291		291		291		291		291	
	150		150		150		150		150	

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 13. Powszechność opieki przedszkolnej i uczęszczanie do przedszkola a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)
oszacowanie efektów stałych								
poziom ucznia								
sprawdzian					11,981	(0,269)	11,971	(0,269)
sprawdzian ²					1,225	(0,117)	1,223	(0,117)
sprawdzian ³					-0,399	(0,061)	-0,399	(0,061)
pleć					-0,860	(0,288)	-0,854	(0,288)
dysleksja na egz. gimnazjalnym					2,106	(0,838)	2,107	(0,848)
interakcja dys. na spr. i egz.					-4,295	(0,882)	-4,303	(0,891)
test matryc Ravena (I pomiar)								
wykształcenie rodziców								
HTREI								
zamożność gosp. domowego								
wiek ucznia								
starszy niż główna kohorta								
młodszy niż główna kohorta								
struktura rodziny (ref. pełna 2p)								
pełna 3 pokoleniowa (3p)								
niepełna dwupokoleniowa								
niepełna trzypokoleniowa								
inna								
brak danych								
liczba lat w przedszkolu							0,181	(0,122)
interakcja wyksz. i l. lat w przedsz.								
stała	100,029	(0,513)	97,129	(1,012)	100,269	(0,727)	100,217	(0,729)
poziom oddziału								
poziom szkoły								
udział dzieci 3–5 lat w przedszk.			0,090	(0,029)	-0,008	(0,017)	-0,012	(0,017)
oszacowanie efektów losowych								
wariancja efektów szkół	21,801		18,221		4,540		4,635	
wariancja efektów oddziałów	20,978		21,127		5,363		5,209	
wariancja na poziomie ucznia	159,830		159,864		66,342		66,414	
poziom oddziału								
współczynniki korelacji wewnątrzgrupowej								
poziom szkół	0,108		0,091		0,060		0,061	
poziom oddziałów	0,104		0,106		0,070		0,068	
poziom szkół i oddziałów łącznie	0,212		0,197		0,130		0,129	
podsumowanie								
deviance	34436,20		34423,686		30545,526		30484,634	
liczba szacowanych parametrów	4		5		11		12	
liczba uczniów	4582		4582		4582		4582	
liczba klas	291		291		291		291	
liczba szkół	150		150		150		150	

SPOŁECZNO-GOSPODARCZE CHARAKTERYSTYKI GMIN

	(5)	(s.e.)	(6)	(s.e.)	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)
					9,546	(0,288)	9,552	(0,286)	9,549	(0,287)
					1,233	(0,112)	1,227	(0,112)	1,230	(0,113)
					-0,373	(0,063)	-0,375	(0,063)	-0,374	(0,063)
					-1,484	(0,258)	-1,473	(0,258)	-1,470	(0,258)
					1,303	(0,782)	1,309	(0,778)	1,323	(0,780)
					-3,314	(0,841)	-3,306	(0,836)	-3,322	(0,838)
	8,145	(0,258)	8,154	(0,258)	3,703	(0,192)	3,709	(0,192)	3,709	(0,192)
	0,723	(0,100)	0,759	(0,100)	0,220	(0,078)	0,240	(0,078)	0,262	(0,095)
	0,063	(0,019)	0,066	(0,019)	0,044	(0,018)	0,045	(0,018)	0,046	(0,018)
	0,960	(0,184)	0,971	(0,183)	0,352	(0,164)	0,357	(0,163)	0,354	(0,164)
	-0,010	(0,052)	-0,013	(0,052)	0,134	(0,042)	0,131	(0,042)	0,131	(0,042)
	-3,214	(1,264)	-3,253	(1,280)	-2,136	(1,078)	-2,178	(1,084)	-2,164	(1,079)
	6,239	(2,306)	6,284	(2,339)	4,736	(2,055)	4,754	(2,066)	4,750	(2,065)
	-0,423	(0,417)	-0,468	(0,417)	0,258	(0,313)	0,219	(0,314)	0,214	(0,313)
	-1,451	(0,737)	-1,471	(0,734)	-1,115	(0,572)	-1,125	(0,573)	-1,123	(0,572)
	-1,697	(0,993)	-1,682	(0,997)	-1,286	(0,878)	-1,278	(0,880)	-1,285	(0,877)
	-0,914	(1,672)	-0,842	(1,657)	-0,072	(1,233)	-0,043	(1,234)	-0,063	(1,232)
	-8,377	(1,909)	-8,444	(1,875)	-5,082	(1,623)	-5,128	(1,611)	-5,110	(1,609)
			-0,363	(0,165)			-0,190	(0,127)	0,048	(0,568)
									-0,018	(0,043)
	101,903	(0,811)	102,120	(0,814)	100,748	(0,713)	100,864	(0,714)	100,883	(0,715)
	-0,003	(0,019)	0,004	(0,020)	-0,024	(0,015)	-0,020	(0,016)	-0,020	(0,016)
	8,253		8,366		4,074		4,154		4,162	
	5,449		5,475		3,310		3,275		3,277	
	97,963		97,936		58,024		58,068		58,063	
	0,074		0,075		0,062		0,063		0,064	
	0,049		0,049		0,051		0,050		0,050	
	0,123		0,124		0,113		0,113		0,114	
	32199,35		32132,742		29935,6		29877,258		29877,056	
	17		18		23		24		25	
	4582		4582		4582		4582		4582	
	291		291		291		291		291	
	150		150		150		150		150	

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 14. Wielkość gminy a wynik egzaminu gimnazjalnego w części humanistycznej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)
oszacowanie efektów stałych										
poziom ucznia										
sprawdzian					11,963	(0,251)			9,966	(0,266)
sprawdzian ²					-0,414	(0,132)			-0,403	(0,128)
sprawdzian ³					-0,372	(0,082)			-0,353	(0,083)
pleć					1,784	(0,331)			1,329	(0,310)
dysleksja na egz. gimnazjalnym					2,480	(0,622)			1,575	(0,653)
dysleksja na spr. szk. podst.					-3,753	(0,916)			-2,859	(0,920)
test matrycy Ravena (I pomiar)							7,067	(0,254)	2,106	(0,194)
wykształcenie rodziców							0,968	(0,111)	0,441	(0,084)
HTREI							0,043	(0,022)	0,030	(0,018)
zamożność gosp. domowego							1,923	(0,224)	1,154	(0,185)
wyposażenie w sprzęt elektro.							-0,451	(0,187)	-0,342	(0,153)
wiek ucznia							-0,060	(0,055)	0,078	(0,041)
starszy niż główna kohorta							-8,223	(1,869)	-3,576	(1,593)
młodszy niż główna kohorta							4,382	(1,998)	2,333	(2,077)
jedynak (ref)										
najmłodszy							-1,742	(0,486)	-1,233	(0,440)
średni							-3,127	(0,742)	-2,056	(0,546)
najstarszy							-0,068	(0,521)	-0,140	(0,464)
brak danych							-3,583	(1,737)	-0,541	(1,239)
stała	100,771	(0,497)	91,938	(3,485)	101,443	(2,057)	106,582	(2,291)	106,459	(2,098)
poziom oddziału										
poziom szkoły										
wielkość gminy (logarytm)			0,903	(0,352)	-0,049	(0,201)	-0,350	(0,207)	-0,472	(0,190)
oszacowanie efektów losowych										
wariancja efektów szkół	19,31		17,14		5,61		5,56		4,75	
wariancja efektów oddziałów	16,87		16,97		2,40		5,26		1,87	
wariancja na poziomie ucznia	174,44		174,50		76,67		113,05		70,01	
poziom oddziału										
współczynniki korelacji wewnątrzgrupowej										
poziom szkół	0,092		0,082		0,066		0,045		0,062	
poziom oddziałów	0,080		0,081		0,028		0,042		0,024	
poziom szkół i oddziałów										
łącznie	0,172		0,164		0,095		0,087		0,086	
podsumowanie										
deviance	37115,95		37108,07		33201,44		34982,54		32767,29	
liczba szacowanych parametr.	4		5		11		17		23	
liczba uczniów	4590		4590		4590		4590		4590	
liczba klas	291		291		291		291		291	
liczba szkół	150		150		150		150		150	

Tabela 15. Wielkość gminy a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)
oszacowanie efektów stałych										
poziom ucznia										
sprawdzian					11,989	(0,269)			9,541	(0,288)
sprawdzian ²					1,225	(0,117)			1,233	(0,112)
sprawdzian ³					-0,399	(0,061)			-0,373	(0,063)
pleć					-0,861	(0,288)			-1,481	(0,258)
dysleksja na egz. gimnazjalnym					2,115	(0,834)			1,306	(0,778)
interakcja dys. na spr. i egz.					-4,283	(0,881)			-3,291	(0,84)
test matryc Ravena (I pomiar)							8,144	(0,257)	3,696	(0,192)
wykształcenie rodziców							0,735	(0,101)	0,229	(0,078)
HTREI							0,064	(0,019)	0,044	(0,018)
zamożność gosp. domowego							0,974	(0,184)	0,358	(0,164)
wiek ucznia							-0,008	(0,052)	0,133	(0,042)
starszy niż główna kohorta							-3,215	(1,268)	-2,158	(1,078)
młodszy niż główna kohorta							6,275	(2,31)	4,787	(2,069)
struktura rodziny (ref. pełna 2p)										
pełna 3 pokoleniowa (3p)							-0,467	(0,419)	0,226	(0,317)
niepełna dwupokoleniowa							-1,413	(0,736)	-1,097	(0,571)
niepełna trzypokoleniowa							-1,690	(0,995)	-1,285	(0,877)
inna							-0,825	(1,673)	-0,026	(1,241)
brak danych							-8,338	(1,902)	-5,094	(1,624)
stała	100,03	(0,513)	94,14	(4,235)	102,77	(2,101)	105,40	(2,442)	104,51	(1,821)
poziom oddziału										
poziom szkoły										
wielkość gminy (logarytm)			0,602	(0,431)	-0,281	(0,204)	-0,369	(0,239)	-0,464	(0,174)
oszacowanie efektów losowych										
wariancja efektów szkół	21,801		20,631		4,464		8,015		3,941	
wariancja efektów oddziałów	20,978		21,068		5,361		5,503		3,297	
wariancja na poziomie ucznia	159,83		159,87		66,33		97,94		58,00	
poziom oddziału										
współczynniki korelacji wewnątrzgrupowej										
poziom szkół	0,108		0,102		0,059		0,072		0,06	
poziom oddziałów	0,104		0,105		0,07		0,049		0,051	
poziom szkół i oddziałów	0,212		0,207		0,129		0,121		0,111	
łącznie										
podsumowanie										
deviance	34436,20		34433,20		30543,30		32196,52		29931,40	
liczba szacowanych parametrów	4		5		11		17		23	
liczba uczniów	4591		4591		4591		4591		4591	
liczba klas	291		291		291		291		291	
liczba szkół	150		150		150		150		150	

7.4.8. Model z kontrolą kilku czynników na poziomie gmin

Ostateczna weryfikacja oddziaływania czynników lokalnych na osiągnięcia uczniów nastąpiła poprzez wyodrębnienie metodą krokowej eliminacji wstecznej czynników oddziałujących wspólnie. W pierwszym kroku wprowadzone zostały: wielkość gminy (logarytm), wydatki gminy na mieszkańca, wskaźnik ubóstwa, udział zarejestrowanych bezrobotnych w liczbie mieszkańców w wieku produkcyjnym, udział 1 i 3 sektora gospodarki w gospodarce ogółem (udział sektora 2 został usunięty z analizy ze względu na współliniowość), udział sektora HT w lokalnej gospodarce, liczba podmiotów gospodarczych na liczbę mieszkańców, liczba podmiotów HT na liczbę mieszkańców, wydatki gminy na oświatę i wychowanie na mieszkańca, wydatki bieżące na gimnazja na ucznia (logarytm), udział dzieci w wieku 3–5 lat objętych opieką w przedszkolach publicznych, liczba bibliotek publicznych na liczbę mieszkańców, wielkość księgozbioru w bibliotekach publicznych na liczbę mieszkańców. Macierz korelacji Pearsona pomiędzy zmiennymi z poziomu lokalnego przedstawia tabela 16.

W części humanistycznej w modelu jako czynniki istotne pozostały: wielkość gminy (wzrost wielkości o 1% obniża osiągnięcia o 0,02 punktu), udział zarejestrowanych bezrobotnych w liczbie mieszkańców w wieku produkcyjnym (wzrost udziału o 1 punkt procentowy obniża osiągnięcia o 0,25 punktu), liczba podmiotów HT na liczbę mieszkańców (wzrost o podmiot na 100 mieszkańców podnosi osiągnięcia o 4,95 punktu), wydatki gminy na oświatę i wychowanie na mieszkańca (wzrost wydatków o 1000 złotych obniża osiągnięcia o 4,11 punktu), udział dzieci w wieku 3–5 lat objętych opieką w przedszkolach publicznych (wzrost o 1 punkt procentowy podnosi osiągnięcia o 0,07 punktu; tabela 17, model 2).

Włączenie zmiennych kontrolnych z modeli EWD wytrąca wpływ większości czynników lokalnych. Jedyną istotną zmienną pozostaje odsetek bezrobotnych. Wzrost wartości tego wskaźnika o 1 punkt procentowy wiąże się z obniżeniem osiągnięć na egzaminie o 0,17 punktu. Kontrola wcześniejszych osiągnięć pozwala ograniczyć znaczenie tego czynnika (tabela 17, model 3).

Po włączeniu do modeli zmiennych indywidualnych (IQ, SES, struktura rodziny) istotnymi predyktorami osiągnięć pozostają wielkość gminy i udział bezrobotnych. Wraz ze wzrostem wielkości gminy o 1% osiągnięcia szkolne spadają o 0,01 punktu. Z kolei wzrost udziału bezrobotnych wiąże się z obniżeniem osiągnięć o 0,24. Zestaw zmiennych indywidualnych w mniejszym stopniu ogranicza oddziaływanie czynników lokalnych na osiągnięcia uczniów (tabela 17, model 4).

Tabela 16. Korelacje Pearsona pomiędzy zmiennymi z poziomu lokalnego

zmienna	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
wydatki w gminie ogółem na osobę w 2010 roku (2)	0,554													
poziom ubóstwa (3)	-0,761	-0,549												
udział i sektora w lokalnej gospodarce (4)	-0,631	-0,182	0,556											
udział ii sektora w lokalnej gospodarce (5)	-0,514	-0,329	0,484	0,281										
udział iii sektora w lokalnej gospodarce (6)	0,695	0,333	-0,635	-0,71	-0,876									
udział sektora ht w lokalnej gospodarce (7)	0,888	0,601	-0,838	-0,589	-0,498	0,662								
I, podmiotów gosp. na 100 mieszkańców (8)	0,736	0,563	-0,716	-0,497	-0,538	0,645	0,762							
I, podmiotów ht na 100 mieszkańców (9)	0,853	0,709	-0,826	-0,507	-0,500	0,622	0,93	0,884						
udział bezrobotnych w liczbie osób w wieku produkcyjnym (10)	-0,368	-0,217	0,614	0,302	0,076	-0,208	-0,52	-0,304	-0,438					
wydatki na osobę na oświatę i wychowanie (11)	0,251	0,389	-0,135	0,055	0,033	-0,052	0,226	0,093	0,198	-0,104				
wydatki bieżące na gimnazja na ucznia (12)	0,091	0,039	-0,129	-0,035	-0,071	0,07	0,107	0,082	0,089	-0,15	0,254			
liczba bibliotek na osobę (13)	-0,608	-0,236	0,519	0,559	0,347	-0,536	-0,597	-0,551	-0,520	0,306	-0,075	-0,137		
księgozbiór bibliotek na 1000 ludności (14)	-0,442	-0,036	0,352	0,466	0,129	-0,329	-0,39	-0,273	-0,308	0,182	0,021	0,002	0,623	
odsetek dzieci objętych wychowaniem przedszkolnym w wieku 3–5 lat w 2003 r. (15)	0,747	0,420	-0,655	-0,626	-0,452	0,647	0,718	0,668	0,695	-0,367	0,178	0,12	-0,539	-0,267

(1) wielkość gminy (logarytm)

Przy kontroli zmiennych z modeli EWD oraz zmiennych indywidualnych, jedynym czynnikiem lokalnym, który oddziałuje na poziom osiągnięć w części humanistycznej, jest udział bezrobotnych w liczbie osób w wieku produkcyjnym. Wzrost udziału bezrobotnych o 1 punkt procentowy obniża osiągnięcia o 0,19 punktu na skali o średniej 100 i odchyleniu standardowym 15 (tabela 17, model 5).

Po zakończeniu procedury krokowej eliminacji wstecznej w modelu dla części matematyczno-przyrodniczej pozostały następujące czynniki lokalne: wielkość gminy (wzrost wielkości o 1% obniża osiągnięcia o 0,03 punktu), udział zarejestrowanych bezrobotnych w liczbie mieszkańców w wieku produkcyjnym (wzrost udziału o 1 punkt procentowy obniża osiągnięcia o 0,23 punktu), liczba podmiotów HT na liczbę mieszkańców (wzrost o podmiot na 100 mieszkańców podnosi osiągnięcia o 6,35 punktu), liczba bibliotek na tysiąc mieszkańców (wzrost o jedną obniża osiągnięcia o 3,16 punktu) oraz udział dzieci w wieku 3–5 lat objętych opieką w przedszkolach publicznych (wzrost o 1 punkt procentowy podnosi osiągnięcia o 0,08 punktu; tabela 18, model 2).

Włączenie zmiennych kontrolnych z modeli EWD całkowicie wytrąca wpływ jedynie powszechności opieki przedszkolnej oraz liczby bibliotek, a także ogranicza wpływ pozostałych czynników (tabela 18, model 3). W modelu tym wzrost wielkości gminy o 1% obniża osiągnięcia o 0,01 punktu. Wzrost udziału bezrobotnych o 1 punkt procentowy obniża osiągnięcia na egzaminie o 0,18 punktu. Z kolei wzrost liczby podmiotów HT o jeden na 100 mieszkańców gminy podnosi osiągnięcia o 1,79 punktu.

Po włączeniu do modeli zmiennych indywidualnych (IQ, SES, struktura rodziny) istotne statystycznie pozostają: wielkość gminy, udział bezrobotnych oraz liczbę podmiotów HT (tabela 18, model 4). Wraz ze wzrostem wielkości gminy o 1% spada oszacowanie osiągnięć o 0,02 punktu. Z kolei wzrost udziału bezrobotnych obniża osiągnięcia o 0,21 punktu. Wzrost liczby podmiotów HT o 1 na 100 mieszkańców prowadzi do podniesienia osiągnięć o 2,77 punktu. Zestaw zmiennych indywidualnych w niewiele mniejszym stopniu ogranicza oddziaływanie czynników lokalnych na osiągnięcia uczniów.

Przy kontroli zmiennych z modeli EWD oraz zmiennych indywidualnych nadal istotnie oddziałują wymienione powyżej trzy czynniki (tabela 18, model 5). Wraz ze wzrostem wielkości gminy o 1% spada oszacowanie osiągnięć o 0,01 punktu. Z kolei wzrost udziału bezrobotnych obniża osiągnięcia o 0,18 punktu. Wzrost liczby podmiotów HT o 1 na 100 mieszkańców prowadzi do podniesienia osiągnięć o 1,20 punktu na skali o średniej 100 i odchyleniu standardowym 15. Szczegółowe zestawienie wyników estymacji poszczególnych modeli przedstawiają tabele 17 i 18.

Tabela 17. Czynniki lokalne a wynik egzaminu gimnazjalnego w części humanistycznej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)
oszacowanie efektów stałych										
poziom ucznia										
sprawdzian					11,929	(0,249)			9,955	(0,265)
sprawdzian ²					-0,422	(0,133)			-0,403	(0,128)
sprawdzian ³					-0,373	(0,083)			-0,354	(0,083)
pleć					1,781	(0,331)			1,330	(0,311)
dysleksja na egz. gimnazjalnym					2,402	(0,635)			1,562	(0,676)
dysleksja na spr. szk. podst.					-3,740	(0,923)			-2,882	(0,931)
test matryc Ravena (I pomiar)							7,022	0,258	2,101	(0,195)
wykształcenie rodziców							0,965	0,109	0,445	(0,083)
HTREI							0,043	0,022	0,031	(0,018)
zamożność gosp. domowego							1,879	0,225	1,134	(0,185)
wyposażenie w sprzęt elektro.							-0,489	0,187	-0,360	(0,154)
wiek ucznia							-0,061	0,055	0,078	(0,041)
starszy niż główna kohorta							-8,400	1,862	-3,622	(1,599)
młodszy niż główna kohorta							4,503	1,946	2,381	(2,038)
jedynak (ref.)										
najmłodszy							-1,744	0,480	-1,246	(0,437)
średni							-3,087	0,739	-2,052	(0,544)
najstarszy							-0,061	0,517	-0,141	(0,463)
brak danych							-3,607	1,744	-0,547	(1,254)
stała	100,771	(0,497)	118,816	(6,160)	108,165	(2,629)	115,828	3,202	109,700	(2,461)
poziom oddziału										
poziom szkoły										
wielkość gminy (logarytm)			-1,822	(0,688)	-0,484	(0,333)	-1,059	0,389	-0,504	(0,314)
udział dzieci 3–5 lat w przedszk.			0,073	(0,028)	0,004	(0,017)	0,013	0,022	-0,007	(0,016)
wydat. w tys. na oświatę/osobę			-4,114	(1,982)	-1,644	(1,041)	-1,631	1,184	-1,070	(0,921)
udział bezrobotnych			-0,252	(0,101)	-0,174	(0,061)	-0,236	0,072	-0,187	(0,057)
I. podm. HT/100 mieszk.			4,949	(1,393)	0,815	(0,547)	1,217	0,724	0,014	(0,487)
oszacowanie efektów losowych										
wariancja efektów szkół	19,31		8,35		4,88		4,32		4,30	
wariancja efektów oddziałów	16,87		16,93		2,35		5,12		1,81	
wariancja na poziomie ucznia	174,44		174,64		76,70		113,11		70,03	
poziom oddziału										
współczynniki korelacji wewnątrzgrupowej										
poziom szkół	0,092		0,042		0,058		0,035		0,056	
poziom oddziałów	0,080		0,085		0,028		0,042		0,024	
poziom szkół i oddziałów	0,172		0,126		0,086		0,077		0,080	
łącznie										
podsumowanie										
deviance	37115,95		37061,74		33189,86		34965,60		32758,77	
liczba szacowanych parametr.	4		9		15		21		27	
liczba uczniów	4590		4590		4590		4590		4590	
liczba klas	291		291		291		291		291	
liczba szkół	150		150		150		150		150	

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 18. Czynniki lokalne a wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)
oszacowanie efektów stałych										
poziom ucznia										
sprawdzian					11,949	(0,267)			9,522	(0,288)
sprawdzian ²					1,215	(0,114)			1,227	(0,111)
sprawdzian ³					-0,399	(0,061)			-0,373	(0,062)
pleć					-0,87	(0,288)			-1,489	(0,257)
dysleksja na egz. gimnazjalnym					1,977	(0,838)			1,213	(0,789)
interakcja dys. na spr. i egz.					-4,192	(0,888)			-3,234	(0,851)
test matrycy Ravena (I pomiar)							8,101	(0,255)	3,689	(0,192)
wykształcenie rodziców							0,732	(0,098)	0,231	(0,077)
HTREI							0,063	(0,019)	0,044	(0,018)
zamożność gosp. domowego							0,934	(0,183)	0,337	(0,163)
wiek ucznia							-0,011	(0,052)	0,133	(0,042)
starszy niż główna kohorta							-3,373	(1,269)	-2,218	(1,081)
młodszy niż główna kohorta							6,379	(2,306)	4,849	(2,051)
struktura rodziny (ref. pełna 2p)										
pełna 3 pokoleniowa (3p)							-0,480	(0,423)	0,199	(0,319)
niepełna dwupokoleniowa							-1,386	(0,739)	-1,067	(0,575)
niepełna trzypokoleniowa							-1,728	(0,989)	-1,330	(0,880)
inna							-0,877	(1,677)	-0,020	(1,229)
brak danych							-8,262	(1,910)	-5,046	(1,605)
stała	100,029	(0,513)	124,764	(8,104)	111,219	(3,166)	118,433	(4,692)	109,886	(2,632)
poziom oddziału										
poziom szkoły										
wielkość gminy (logarytm)			-2,951	(0,955)	-1,107	(0,32)	-1,724	(0,554)	-0,910	(0,270)
I. podm. HT/100 mieszk.			6,353	(1,786)	1,794	(0,571)	2,770	(1,014)	1,199	(0,481)
udział bezrobotnych			-0,229	(0,124)	-0,179	(0,082)	-0,210	(0,084)	-0,179	(0,073)
biblioteki/1000 mieszkańców			-3,161	(1,572)	-0,449	(1,081)	-0,999	(1,281)	0,033	(0,952)
udział dzieci 3–5 lat w przedszk.			0,059	(0,029)	-0,006	(0,018)	0,001	(0,023)	-0,016	(0,017)
oszacowanie efektów losowych										
wariancja efektów szkół	21,801		10,331		3,568		6,009		3,390	
wariancja efektów oddziałów	20,978		20,831		5,233		5,379		3,194	
wariancja na poziomie ucznia	159,830		159,936		66,335		97,958		58,016	
poziom oddziału										
współczynniki korelacji wewnątrzgrupowej										
poziom szkół	0,108		0,054		0,047		0,055		0,052	
poziom oddziałów	0,104		0,109		0,070		0,049		0,049	
poziom szkół i oddziałów łącznie	0,212		0,163		0,117		0,104		0,101	
podsumowanie										
deviance	34436,20		34386,37		30526,96		32173,48		29918,52	
liczba szacowanych parametrów	4		8		14		20		26	
liczba uczniów	4591		4591		4591		4591		4591	
liczba klas	291		291		291		291		291	
liczba szkół	150		150		150		150		150	

Przeprowadzone analizy pozwalają na stwierdzenie, że modele EWD efektywnie kontrolują wpływ czynników lokalnych na osiągnięcia uczniów w części humanistycznej egzaminu gimnazjalnego. W części matematyczno-przyrodniczej modele te znoszą wpływ poziomu upowszechnienia opieki przedszkolnej i znacznie ograniczają wpływ pozostałych czynników lokalnych, jednak trzy z nich pozostają istotne. Udział bezrobotnych jest tą zmienną, która ma znaczenie dla szacowania wyniku z obu części egzaminu, przy kontroli zmiennych z modeli EWD. W efekcie czynnik ten oddziałuje na pomiar efektywności. Zrozumienie wpływu czynników lokalnych a szczególnie udziału bezrobotnych oraz wypracowanie metod kontroli tych zmiennych wymaga dodatkowych badań.

7.5. Podsumowanie

Przeprowadzone analizy pozwalają w zdecydowanej większości przyjąć postawione hipotezy badawcze. W każdym z analizowanych obszarów istnieją relacje wiążące przynajmniej jeden ze wskaźników z osiągnięciami, zgodnie z przewidywaniami na podstawie przeglądu literatury. Jedynie w przypadku czasu uczęszczania do przedszkola czynnik ten nie ma znaczenia dla osiągnięć w części humanistycznej egzaminu gimnazjalnego. Zaprezentowane analizy pozwalają również pozytywnie zweryfikować hipotezę o wyłączeniu lub znacznym ograniczeniu siły związku czynników społeczno-lokalnych na osiągnięcia na egzaminie gimnazjalnym, poprzez włączenie kontroli wcześniejszych osiągnięć edukacyjnych.

Przedstawiając wnioski w formie bardziej syntetycznej, można podzielić wskaźniki poszczególnych czynników lokalnych na trzy grupy. Pierwsza to zmienne nieoddziałujące na osiągnięcia edukacyjne. Druga to takie czynniki, których wpływ na osiągnięcia można kontrolować dzięki wykorzystaniu modeli EWD. Trzecia to takie, których oddziaływanie trwa również przy kontroli zmiennych z modeli EWD.

Zmienne, które samodzielnie nie mają wpływu na osiągnięcia na egzaminie gimnazjalnym to:

- wskaźnik zamożności, jakim są wydatki gminy zrelatywizowane liczbą mieszkańców,
- wskaźniki wydatków edukacyjnych, tak wydatki na oświatę i wychowanie, jak i wydatki bieżące na gimnazja w przeliczeniu na ucznia,
- wskaźnik dostępności kulturowej, jakim jest liczba woluminów w bibliotekach publicznych na mieszkańca.

Czynniki, których wpływ można kontrolować, włączając jedynie zmienne z modeli EWD to:

- wskaźnik zamożności, jakim jest poziom ubóstwa,
- wszystkie wskaźniki struktury lokalnej gospodarki, czyli udział każdego z trzech

- sektorów gospodarki, udział podmiotów HT w lokalnej gospodarce, liczba podmiotów gospodarczych, oraz liczba podmiotów HT zrelatywizowana liczbą mieszkańców,
- wskaźnik dostępności kulturowej, jakim jest liczba bibliotek na 1000 mieszkańców,
 - wskaźnik upowszechnienia opieki przedszkolnej, czyli udział dzieci objętych taką wśród dzieci w tym wieku w gminie.

Do ostatniej grupy można zatem zaliczyć jedynie:

- udział zarejestrowanych bezrobotnych w populacji osób w wieku produkcyjnym.

Warto jednak zaznaczyć, że przy kontroli zmiennych z modeli EWD znaczenie tej zmiennej dla szacowania efektywności gimnazjów jest bardzo ograniczone.

Zmienne o dużym znaczeniu dla szacowania efektywności gimnazjów można zidentyfikować, wprowadzając do modelu pustego wszystkie wskaźniki czynników lokalnych i następnie krokowo eliminując te, które przestają być istotne statystycznie. Zmienne, które pozostały w modelach po zastosowaniu takiej procedury, to: wielkość gminy, udział dzieci 3–5-letnich w opiece przedszkolnej, udział zarejestrowanych bezrobotnych w populacji osób w wieku produkcyjnym oraz liczba podmiotów HT zrelatywizowana liczbą mieszkańców. Dla osiągnięć w części humanistycznej egzaminu gimnazjalnego, istotne okazały się dodatkowo wydatki na oświatę i wychowanie na mieszkańca. Z kolei w części matematyczno-przyrodniczej liczba bibliotek na tysiąc mieszkańców. Po wprowadzeniu zmiennych, które uwzględniane są w modelach EWD, istotnymi predyktorami w obu modelach pozostał jedynie udział zarejestrowanych bezrobotnych w populacji osób w wieku produkcyjnym. Dodatkowo w modelu dla części matematyczno-przyrodniczej istotna pozostała wielkość gminy oraz liczba podmiotów HT zrelatywizowana liczbą mieszkańców.

Zmienne te oddziałują na oszacowania osiągnięć szkolnych również przy kontroli zmiennych kontekstowych. Z tego powodu należy rozważyć uwzględnianie tych zmiennych w modelach EWD w celu odciążenia oszacowań efektywności szkół z lokalnych czynników o charakterze pozaszkolnym. Dane te są dostępne i mogą być wykorzystane w modelowaniu.

Dodatkowym ciekawym wnioskiem płynącym z przeprowadzonych analiz jest fakt, że dostęp do zasobów kulturowych inaczej oddziałuje na wyniki w testach na kolejnych poziomach edukacji. Hipoteza taka pozwoliłaby wyjaśnić ujawniony pozytywny wpływ tego czynnika przy kontroli zmiennych z modeli EWD oraz zmiennych kontekstowych. Większe znaczenie dostępności zasobów kulturowych dla uzyskiwanych wyników ma miejsce w gimnazjum niż w szkole podstawowej. Z tego względu nie da się w pełni kontrolować tego czynnika, uwzględniając wyżej wymienione zmienne.

Innym ciekawym spostrzeżeniem jest negatywny związek wielkości gminy i osiągnięć na obu częściach egzaminu, przy kontroli innych istotnych zmiennych lokalnych i/lub zmiennych z modeli EWD i zmiennych kontekstowych.

Ważnym ograniczeniem przeprowadzonych analiz jest jakość danych gromadzonych w Banku Danych Lokalnych GUS. Większość z nich jest dobrej jakości i jest powszechnie wykorzystywana w analizach. Niektóre jednak (np. wskaźniki struktury gospodarki oparte na rejestrze REGON) mogą być obciążone błędem systematycznym. Należy jednak zauważyć, że w powszechnym dostępie nie ma lepszych danych w tym zakresie.

Literatura

- Ananat, E. O., Gassman-Pines, A., Francis, D. V., i Gibson-Davis, C. M. (2011). *Children Left Behind: The Effects of Statewide Job Loss on Student Achievement*. National Bureau of Economic Research, Inc. Pobrano z <http://ideas.repec.org/p/nbr/nberwo/17104.html>.
- Broomhall, D. E. i Johnson, T. G. (1994). Economic Factors That Influence Educational Performance in Rural Schools. *American Journal of Agricultural Economics*, 76(3), s. 557–567.
- Camilli, G., Vargas, S., Ryan, S. i Barnett, W. S. (2010). Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development. *Teachers College Record*, 112(3), s. 579–620.
- Chan, C. (2008). The impact of school library services on student achievement and the implications for advocacy: A review of the literature. *Access*, 22(4), 15–20.
- Costrell, R., Hanushek, E. i Loeb, S. (2008). What Do Cost Functions Tell Us About the Cost of an Adequate Education? *Peabody Journal of Education*, 83(2), s. 198–223. doi:10.1080/01619560801996988
- Domański, H. i Pokropek, A. (2011). *Podziały terytorialne, globalizacja a nierówności społeczne: wprowadzenie do modeli wielopoziomowych*. Warszawa: Wydawnictwo IFiS PAN.
- Feinstein, L. i Symons, J. (1999). Attainment in secondary school. *Oxford Economic Papers*, 51(2), s. 300–321. doi:10.1093/oep/51.2.300
- Geoffroy, M.-C., Côté, S. M., Giguère, C.-É., Dionne, G., Zelazo, P. D., Tremblay, R. E. i Séguin, J. R. (2010). Closing the gap in academic readiness and achievement: the role of early childcare. *Journal of Child Psychology and Psychiatry*, 51(12), s. 1359–1367. doi:10.1111/j.1469-7610.2010.02316.x

Giza-Poleszczuk, A., Trutkowski, C., Krzyżanowska, Ł., Olesińska, O., Olesińska, H. i Zając, T. (2012). *Po co Polakom Biblioteki?* (s. 62). Warszawa: Fundacja Rozwoju Społeczeństwa Informacyjnego. Pobrano z http://www.biblioteki.org/repository/PLIKI/DOKUMENTY/RAPORTY/RAPORT_PO_CO_POLAKOM_BIBLIOTEKI_www.pdf

GUS. (2013). BDL: Bank Danych Lokalnych. *O Banku*. Pobrano 30 kwiecień 2013, z http://www.stat.gov.pl/bdl/app/strona.html?p_name=obanku

Hanushek, E. A. (2001). Spending on Schools. [w:] Moe, T. (red.), *A Primer on American Education* (ss. 69–88). Stanford, CA: Hoover Institution Press. Pobrano z <http://hanushek.stanford.edu/publications/spending-schools>

Hanushek, E. A. (2013). Financing Schools. [w:] Hattie, J. i Anderman, E. M. (red.), *International Guide to Student Achievement* (ss. 134–136). New York: Routledge. Pobrano z <http://hanushek.stanford.edu/publications/financing-schools>

Hanushek, E. A. i Woessmann, F. (2011). The Economics of International Differences in Educational Achievement. [w:] Hanushek, E. A., Machin, S. i Woessmann, L. (red.), *Handbook of the economics of education* (T. 3). Amsterdam; London: North-Holland. Pobrano z <http://www.sciencedirect.com/science/book/9780444534293>

Herbst, M. (2004). Zróżnicowanie jakości kapitału ludzkiego w Polsce. Od czego zależą wyniki edukacyjne? *Studia Regionalne i Lokalne*, 3(17), 89–104.

Herczynski, J. i Herbst, M. (2002). *Pierwsza odsłona: społeczne i terytorialne zróżnicowanie wyników sprawdzianu ósmoklasistów i egzaminu gimnazjalnego przeprowadzonych wiosną 2002 roku*. Warszawa: Fundacja Klub Obywatelski.

James, R. (2001). Participation disadvantage in Australian higher education: An analysis of some effects of geographical location and socioeconomic status. *Higher Education*, 42(4), 455–472. doi:Article

Konarski, R. (2009). *Modele równań strukturalnych: teoria i praktyka*. Wydawnictwo Naukowe PWN.

Konarzewski, K. (2012). TIMSS i PIRLS 2011 *Osiągnięcia szkolne polskich trzecioklasistów w perspektywie międzynarodowej*. Warszawa: Centralna Komisja Egzaminacyjna. Pobrano z http://www.cke.edu.pl/images/stories/000000000000000000_pirls_2011/timss_pirls_2011_raport_koncowy.pdf

Leak, J., Duncan, G. J., Li, W., Magnuson, K., Schindler, H. i Yoshikawa, H. (2010). Is Timing Everything? How Early Childhood Education Program Impacts Vary by Starting Age, Program Duration and Time Since the End of the Program. Zaprezentowano na Association for Policy

Analysis and Management meeting, Boston, MA. Pobrano z http://www.gse.uci.edu/docs/Leak_Duncan_Li_Timing_Paper_APPAM_102810.pdf

Levine, P. B. (2011). How Does Parental Unemployment Affect Children's Educational Performance. [w:] Duncan, G. J. i Murnane, R. J. (red.), *Whither opportunity?: rising inequality, schools, and children's life chances* (ss. 315–335). New York : Chicago: Russell Sage Foundation ; Spencer Foundation.

OECD. (2010). *The High Cost of Low Educational Performance The Long-run Economic Impact of Improving PISA Outcomes*. Paris: OECD Publishing.

Schütz, G., Ursprung, H. i Woessmann, L. (2005). Education Policy and Equality of Opportunity. CESifo Group Munich. Pobrano z http://ideas.repec.org/p/ces/ceswps/_1518.html

Scientific Software International. (2012). Pobrano 19 marzec 2012, z <http://www.ssicentral.com/hlm/example6-2.html>

Stevens, A. H. i Schaller, J. (2009). Short-run Effects of Parental Job Loss on Children's Academic Achievement. *National Bureau of Economic Research Working Paper Series*, No. 15480. Pobrano z <http://www.nber.org/papers/w15480>

Świdurska, A. (2009). Kształtowanie się przemysłu wysokiej techniki w Polsce. *Prace Komisji Geografii Przemysłu*, 13, 56–67.

Ustawa z 27 czerwca 1997 r. o bibliotekach; Dz.U. nr 85, poz. 539 z późn zm.

Williams, D., Wavell, C. i Coles, L. (2001). Impact of school library services on achievement and learning. *Aberdeen: Robert Gordon University*.

Yu, C. H. (2012). Beyond Gross National Product: An Exploratory Study of the Relationship between Program for International Student Assessment Scores and Well-being Indices. *Review of European Studies*, 4(5). doi:10.5539/res.v4n5p119

Rozdział 8

Selekcja uczniów na progu gimnazjum

Prezentowane w tym rozdziale wyniki dotyczą wpływu poziomu selekcji na progu gimnazjum na wyniki uzyskiwane przez uczniów na egzaminie gimnazjalnym. Rozpatrzenie tego problemu jest istotne dla trafności interpretacji przyczynowo-skutkowej wskaźników EWD dla szkół. Tak jak opisane to było w koncepcji badania trafności metody EWD, kluczowym problemem dla zasadności przyczynowo-skutkowej interpretacji wskaźników EWD jest niespełnienie założenia, że przydział uczniów do szkół jest losowy. Szkoły silnie selekcyjne, które dobierają większość uczniów na podstawie innych kryteriów niż rejon zamieszkania, zaburzają to założenie, nie wiemy jednak w jakim stopniu, i jakie to ma znaczenie dla szacowania EWD. Sam fakt segregacji na progu gimnazjów ze względu na uprzednie osiągnięcia można by uwzględnić w modelu EWD poprzez wprowadzenie średnich dla szkół na wejściu. Jednak rodzi to dwa ważne problemy. Po pierwsze, siła segregacji międzyszkolnej jest bardzo zróżnicowana – słaba na terenach wiejskich, dość silna w dużych miastach (por. Dolata, Jasińska i Modzelewski, 2012). Po drugie, zasadne jest podejrzenie, że średnia na wejściu może być skorelowana z konstruktami, który chcemy wskaźnikować, czyli efektywnością nauczania. Z tego względu warto najpierw przeanalizować ten problem w inny sposób. W tym rozdziale zostaną przedstawione analizy odnoszące się do wpływu, inaczej niż przez średnią osiągnięć na wejściu wskaźnikowanej selekcji do gimnazjum, na osiągnięcia uczniów kończących gimnazjum i wskaźniki EWD.

8.1. Selekcja i segregacja uczniów jako problem badawczy

Selekcja uczniów do szkoły jest procesem, podczas którego szkoła wybiera kandydatów do niej na podstawie danego kryterium. Może ono być zależne od różnych cech osobowych lub społecznych jednostki bądź jej rodziny. Selekcję należy rozpatrywać w kontekście szerszego zjawiska, jakim jest segregacja szkolna. (Clark i Maas, 2012; Dolata, 2008). W tym znaczeniu oba procesy (selekcji i segregacji) można powiązać z szerszą kategorią problemów badawczych – nierównościami społecznymi i wynikającymi z nich nierównościami edukacyjnymi. Zagadnienia te są niezwykle istotne i wielokrotnie podejmowane w badaniach nad edukacją prowadzonych przez socjologów i przedstawicieli innych nauk społecznych, głównie z powodu kluczowej roli, jaką przypisuje się systemowi szkolnictwa w kształtowaniu struktury społecznej (Domański, 2004).

Mechanizmem, za pomocą którego nierówności edukacyjne mogą narastać, jest segregacja szkolna (Dolata, 2008; Waslander, Pater, i van der Weide, 2010). Roman Dolata, autor książki

Szkola–segregacje–nierówności (2008), bardzo wnikliwie analizującej zagadnienia nierówności edukacyjnych w kontekście polskich i międzynarodowych badań edukacyjnych, wskazał dwa główne aspekty, które należy wyróżnić, by móc definiować segregację na poziomie operacyjnym: układ segregacyjny i kryterium segregacji. Ze względu na układ można wyróżnić trzy główne rodzaje segregacji: segregację przestrzenną (miasto–wieś), segregację międzyszkolną i segregację wewnątrzszkolną. W przypadku kryterium segregacji badacze odwołują się do różnych cech społecznych, psychologicznych i ekonomicznych uczniów, takich jak płeć, status społeczny, uprzednie osiągnięcia szkolne czy też dochód rodziny. W ten sposób otrzymujemy układ, w którym możemy analizować segregację ze względu na poziom, na jakim występuje, oraz cechy, ze względu na które jest przeprowadzana. Natężenie segregacji na danym poziomie układu można mierzyć. Wskaźnik, jaki zostanie przyjęty w tego rodzaju badaniach zależy od tego, na jakie kryterium segregacji będziemy zwracać uwagę (Allen, 2006; Goldstein i Noden, 2003; Gorard i Cheng, 2011). Natężenie segregacji na danym poziomie szacuje się sprawdzając, w jakim stopniu wskaźnik osiągnięć uczniowskich zależy od danej cechy segregacyjnej. Przykładem takich analiz są modele hierarchiczne, pozwalające na szacowanie zróżnicowania międzyszkolnego ze względu na daną cechę, która stanowi przyjęte kryterium segregacji (Dolata, Jasińska i Modzelewski, 2012).

Należy w tym miejscu zauważyć różnice w postrzeganiu praktyk segregacyjnych w edukacji. W społeczeństwie demokratycznym segregacje między- czy wewnątrzszkolne oparte na cechach społecznych ucznia (np. statusie rodziny, pochodzeniu) są nie do zaakceptowania, natomiast stosowane są praktyki wykorzystujące jako kryterium uprzednie osiągnięcia szkolne (np. tworzenie klas w oparciu o stopień zaawansowania języka obcego), które są czasami uważane za rozwiązania wartościowe ze względu na efektywność kształcenia (Dolata, 2008). Jednak analiza tego rodzaju działań z punktu widzenia paradygmatu konfliktu zwraca uwagę na niebezpieczeństwo funkcji ukrytej takich rozwiązań, które mogą prowadzić do reprodukcji nierówności społecznych (Bourdieu i Passeron, 1990).

Procesy segregacji w edukacji, na każdym z trzech poziomów wyróżnionych powyżej, mogą być (świadomie bądź nie) kierowane polityką edukacyjną instytucji odpowiedzialnych za funkcjonowanie systemu oświaty. Przykładem może być reforma edukacyjna z 1999 roku. Na poziomie całego systemu szkolnictwa w Polsce, decyzja o wprowadzeniu gimnazjów była równocześnie decyzją o przedłużeniu procesu kształcenia młodzieży w ogólnym toku kształcenia o dwa lata (Wosik, 2002). Gimnazja jako szkoły rejonowe miały znacznie utrudnić funkcjonowanie mechanizmów selekcji międzyszkolnej na progu szkolnym, jednak przynajmniej w dużych miastach proces ten postępuje (Dolata, 2011). Na poziomie lokalnym i regionalnym to samorządy decydują bezpośrednio o różnego rodzaju aspektach formalnego funkcjonowania szkół. Niektóre rozwiązania przez nie proponowane mogą prowadzić

do zwiększenia się segregacji międzyszkolnej. Z tego powodu zagadnienia wpływu polityki edukacyjnej na procesy segregacji rozpatrywane są często z punktu widzenia funkcjonowania szkoły jako systemu quasi-rynkowego w oświacie, lub inaczej, mechanizmów rynkowych w edukacji (Bradley i Taylor, 2002; Dolata, 2008; Elacqua, 2012; Exley, 2009; Gorard, Taylor i Fitz, 2001; Waslander i in., 2010). W tym nurcie analizie podlega wpływ zmian organizacyjnych w oświacie, przeprowadzonych w nurcie liberalnym, na poziom nierówności edukacyjnych, wyniki uczniów, innowacyjność kształcenia czy sposób funkcjonowania szkół. Autorzy przeglądu badań z tego nurtu, przygotowanego na zlecenie OECD (Waslander i in., 2010), wskazują że „[...] *wpływ mechanizmów rynkowych na sektor edukacji jest niewielki lub zgoła żaden*”, choć w niektórych aspektach uzyskano wyniki wskazujące na pozytywne powiązania, zwłaszcza z wynikami testów sprawdzających umiejętności czytania i kompetencji matematycznych. W konkluzji, autorzy opracowania doszli do wniosku, że procesy rynkowe w oświacie nie są w stanie zadziałać zgodnie z założeniami. Powodem takiego stanu rzeczy jest istnienie dużej liczby założeń dotyczących sytuacji, w jakiej muszą znajdować się wszyscy aktorzy procesu edukacyjnego (zarządzający oświatą, dyrektorzy szkół, rodzice i uczniowie), by podejmować racjonalne decyzje. Dodatkową barierą jest brak uwzględnienia kontekstu, w jakim funkcjonują konkretne szkoły. Jako przykład można tutaj podać problem związany z uprzywilejowaną sytuacją szkół o najlepszej opinii w danym środowisku edukacyjnym. W przypadku otworzenia rynku edukacyjnego, szkoła ta będzie przyjmować najlepszych uczniów, co da jej gwarancję wysokich wyników w przyszłości i zwiększy prawdopodobieństwo naboru najlepszych uczniów w następnych latach. Proste wprowadzenie koncepcji rynkowych w edukacji (np. bonu edukacyjnego), bez uwzględnienia jej kontekstu, nie prowadzi do osiągnięcia zakładanych rezultatów w postaci polepszenia efektów kształcenia w lokalnym układzie szkolnym (Chaber, 2000; Dolata, 2008; Ladd, 2002). Nie istnieje także proste przełożenie między wielkością nakładów szkolnych a poziomem osiągnięć szkolnych, co wykazano zarówno w realiach USA (Hanushek, 1986), jak i w analizach opartych o polskie dane edukacyjne (Jakubowski, 2007; Śleszyński, 2004).

8.2. Pytania badawcze i hipotezy

Problematykę prezentowaną w rozdziale można streścić w następujących pytaniach badawczych:

- Czy szkoły różniące się selekcją uczniów rozpoczynających w nich naukę różnią się wynikami egzaminacyjnymi?
- Jaki jest wpływ selekcji uczniów rozpoczynających naukę w szkole na wyniki egzaminacyjne?

- Czy szkoły różniące się selekcją uczniów rozpoczynających w nich naukę różnią się wskaźnikiem EWD?
- Jaki jest wpływ selekcji uczniów rozpoczynających naukę w szkole na wskaźniki EWD?

W tym miejscu należy rozważyć wzajemne stosunki między terminami „selekcja” i „segregacja”. Niewątpliwie są to pojęcia bardzo pokrewne. Oba wskazują na fakt podziału uczniów ze względu na przyjęte kryterium. Należy jednak zauważyć, że w powszechnej opinii segregacja jest słowem nacechowanym negatywnie, selekcja natomiast ma konotacje neutralne. Z drugiej strony, mówienie o procesie segregacji odnosi się raczej do całego systemu oświaty. Selekcja opisuje proces, którego dokonuje szkoła, przyjmując lub nie przyjmując uczniów, którzy do niej aplikują (Clark i Maas, 2012). Oczywiście, samo zagadnienie selekcji uwzględnia także procesy autoselekcji, decyzji podejmowanych przez samych uczniów i ich rodziców, związanych z wyborem szkoły, a także selekcji negatywnej, czyli opuszczania szkół rejonowych przez uczniów o niskich wynikach. Jednak zagadnienia autoselekcji nie będą rozważane w tym rozdziale. Będą natomiast analizowane te procesy, które w dużym stopniu są zależne od szkoły, bardziej zasadne wydaje się zatem używanie pojęcia selekcja.

Przywołana we wcześniejszej części literatura przedmiotu wskazuje, że prawdopodobny jest pozytywny wpływ selekcji uczniów do szkół na osiągnięcia uczniowskie, mierzone wynikami testów egzaminacyjnych (Waslander, i in., 2010). Oznacza to, że szkoły o większym natężeniu selekcji osiągają nieco lepsze wyniki – z tego względu w przedstawionych w rozdziale analizach założono taki właśnie związek między selekcją uczniów na wejściu do szkoły a surowym wynikiem egzaminacyjnym.

Brakuje opracowań wskazujących na podobny efekt wpływu selekcyjności na wskaźniki edukacyjnej wartości dodanej. Istniejące analizy wskazują na coraz silniejsze zjawisko różnicowania się osiągnięć egzaminacyjnych szkół, czego wytłumaczeniem może być właśnie proces selekcji uczniów do szkół (Dolata i in., 2012; Żółtak, 2011). Jednak ponieważ wskaźniki EWD są miarami względnych postępów uczniów w zakresie wiedzy i umiejętności, prawdopodobne jest, że w ich przypadku wpływ selekcji uczniów nie będzie widoczny.

8.3. Używane narzędzia badawcze

W badaniach EWD zebrano dane, pozwalające na przeprowadzenie analiz dotyczących zarówno segregacji szkolnej ze względu na cechy społeczne na poziomie całego systemu szkolnego, ale także informacje dotyczące praktyk selekcyjnych prowadzonych przez szkoły.

Dane dotyczące selekcji uczniów do szkoły zbierano dwukrotnie, podczas I i III etapu badania. Zapytano dyrektorów o liczebności trzech grup uczniów: wszystkich przyjętych do szkoły, przyjętych do szkoły spoza rejonu jej działalności oraz takich, którzy nieskutecznie starali się o przyjęcie. Zbierano informacje dla klas pierwszych, na przestrzeni lat 2007–2012. Pytania o rok 2009 powtórzyły się na pierwszym i trzecim etapie badania. Zabieg ten miał na celu weryfikację informacji dotyczących selekcji uczniów do szkoły.

Po zakończeniu III etapu badania, zebrane dane zostały zweryfikowane. Proces ten składał się z kilku kroków. Pierwszym z nich było sprawdzenie spójności wewnętrznej danych, z uwzględnieniem logiki powiązania ze sobą pytań. Następnie sprawdzono spójność danych w dwóch kolejnych etapach badania, wykorzystując pytanie o selekcję do szkoły w 2009 roku, które powtórzyło się na I i III etapie badania. Ostatnim kryterium poprawności zebranych danych było porównanie liczby przyjętych do szkoły uczniów w 2009 roku z liczbą uczniów zdających w szkole egzamin gimnazjalny w 2012 roku.

Proces weryfikacji danych doprowadził do zakwestionowania wyników dla 44 szkół (ze względu na braki lub niespójność danych uzyskanych podczas obu etapów badania). W rezultacie weryfikacji, zdecydowano o pozyskaniu brakujących danych w dodatkowym kontakcie z badanymi szkołami. Dzięki temu uzyskano dane umożliwiające policzenie wykorzystywanych w analizach wskaźników selekcji dla 142 szkół (ze 150 objętych badaniem).

8.4. Wskaźniki selekcji

Dane zebrane podczas badania podłużnego w gimnazjach pozwoliły na ujęcie problemów selekcji uczniów z uwzględnieniem dwóch aspektów tego zjawiska (Jasińska, 2010). Z jednej strony, badając ten problem należy brać pod uwagę odsetek uczniów spoza rejonu, który uczy się w szkole. Jednocześnie trzeba także brać pod uwagę liczbę uczniów, która chciała dostać się do danej szkoły spoza rejonu jej funkcjonowania. Jeśli szkoła przyjmuje wszystkich uczniów, którzy starali się o przyjęcie do niej, nie należy myśleć o niej jako o szkole selekcyjnej.

Na podstawie zebranych danych policzono oba wskaźniki selekcji, opisane we wcześniejszym akapicie. Wskaźniki opracowano łącznie dla wszystkich lat, z których posiadano dane. Oznacza to, że opisują one proces selekcji w szkole na przestrzeni 5 lat (od 2007 do 2012 roku). Decyzję tę podjęto, by uchwycić stałą tendencję szkoły w tym zakresie. Dodatkowo, tak policzone wskaźniki są bardziej odporne na losowe wahania.

Pierwszym z policzonych wskaźników był odsetek uczniów spoza rejonu, który został przyjęty do klas na przestrzeni lat 2007–2012 (roboczo nazwiemy go wskaźnikiem S1). Obliczono go według następującej formuły:

$$\frac{\text{suma liczby uczniów spoza rejonu z 5 lat}}{\text{suma liczby wszystkich uczniów przyjętych do szkoły z 5 lat}} \cdot 100\%$$

Drugim z używanych wskaźników selekcji jest wskaźnik biorący pod uwagę liczbę uczniów chcących dostać się do danej szkoły (roboczo nazwiemy go wskaźnikiem S2). Został on obliczony w następujący sposób:

$$\frac{\text{suma liczby uczniów nieprzyjętych z 5 lat}}{\text{suma liczby uczniów nieprzyjętych z 5 lat} + \text{suma liczby uczniów przyjętych spoza rejonu z 5 lat}} \cdot 100\%$$

Poniżej w tabeli znajdują się podstawowe statystyki opisowe obu policzonych wskaźników selekcji. Wynika z nich przede wszystkim silna asymetryczność obu wskaźników, która uniemożliwia wykorzystanie ich w takiej postaci w modelowaniu hierarchicznym. Równocześnie należy zauważyć, że wśród przebadanych szkół znajdują się także szkoły silnie selekcyjne, przyjmujące do klas pierwszych na przestrzeni lat 2007–2012 jedynie uczniów spoza rejonu. Są to szkoły pozbawione własnego obwodu szkolnego. Z drugiej strony, merytoryczna interpretacja wskaźników prowadzi do wniosku, że analiza obliczonych wskaźników selekcji pojedynczo jest niewystarczająca. Szkoły posiadające wysokie wyniki we wskaźniku S1 mogą przyjmować wszystkich, którzy się do nich zgłaszają. Wskaźnik ten nie bierze pod uwagę informacji o liczbie chętnych, starających dostać się do szkoły. Zasadne zatem wydaje się połączenie informacji z obu wskaźników selekcji w jedną zmienną.

Tabela 1. Statystyki opisowe dla wskaźników selekcji S1 i S2

wskaźnik selekcji	N ważnych	braki danych	średnia	odch. st.	min	max	skośność	kurtosis	kwartył 1	mediana	kwartył 3
S1	148	2	22,75	23,72	0	100	1,512	2,034	4,98	14,34	32,22
S2	142	8	6,59	13,02	0	66,4	2,476	5,996	0	0	7,31

Na podstawie obu wskaźników selekcji zbudowano typologię, która miała określać rodzaj szkoły pod względem procesu przyjmowania uczniów spoza rejonu do szkoły. Przyjęto, że typologia będzie składała się z trzech kategorii (założenie to oparto na analizie obu wskaźników selekcji):

1. Szkoła nieselekcyjna – jest to szkoła, która nie selekcionuje uczniów przyjmowanych do klas pierwszych według specjalnego kryterium. Odzwierciedleniem tego faktu w szkole jest odsetek przyjmowanych uczniów spoza rejonu, mniejszy bądź równy 4%. Do tej

kategorii zakwalifikowano także szkoły, w których przyjęto wszystkich uczniów, którzy chcieli zostać przyjęci.

2. Szkoła umiarkowanie selekcyjna – przyjęto, że będą to szkoły, w których odsetek uczniów spoza rejonu jest większy od 4%, jednak jednocześnie odsetek uczniów spoza rejonu nieprzyjętych do niej nie przekracza 25%.
3. Szkoła selekcyjna – ten typ reprezentują szkoły, w których następuje silny proces selekcji, a uczniowie są w niej dobierani na podstawie dodatkowych kryteriów. Odsetek uczniów przyjętych spoza rejonu przekracza 4%, a udział tego rodzaju uczniów nieprzyjętych do klas pierwszych w przeciągu lat 2007–2012 przekracza 25%.

Przyjęcie używanych podczas tworzenia typologii progów zostało podyktowane względami praktycznymi. W przypadku pierwszego progu, założono, że jeśli w klasie przeciętnie jeden uczeń został przyjęty spoza rejonu, nie świadczy to o istnieniu w szkole procesów selekcyjnych. Ponieważ średnia liczba uczniów w klasie w badanych szkołach wynosiła około 23 uczniów w 2012 roku, przyjęto jako pierwszy próg 4%, co odpowiada około 1 uczniowi spoza rejonu w klasie.

Drugi próg został wyznaczony na podstawie analizy rozkładu wyników szkół pod względem wskaźnika S2, przy wyłączeniu szkół, w których uznano, że nie zaszły procesy selekcji (czyli wartość wskaźnika S1 była mniejsza bądź równa 45). W tym wypadku zaobserwowano na histogramie podział na dwie grupy szkół, dla których wartość graniczną można było wyznaczyć między 25% a 30% wartości wskaźnika. Z powodów analitycznych (zwiększenia ilości szkół w kategorii selekcyjne) zdecydowano się na próg 25%. Oba wskaźniki zostały zatem oparte na eksploracji danych, jednak należy zwrócić uwagę na ich pewną arbitralność. Jednak dla celów empirycznych wydają się one wystarczające.

Podsumowanie informacji dotyczących typologii selekcji (roboczo nazwanej TS1) znajduje się w tabeli 2. Należy zauważyć, że zdecydowanie dominują szkoły umiarkowanie selekcyjne – w analizowanych danych stanowią one blisko 71% placówek, dla których możliwe jest przeprowadzenie analiz. Najmniej liczną grupą są szkoły selekcyjne, które stanowią 7,1%. Istotna jest także stosunkowo niewielka liczba braków danych – tylko dla 9 szkół nie udało się wskazać odpowiadającego im typu (wynika to z braków danych dla wskaźników selekcji S1 i S2).

Tabela 2. Typologia selekcji - podsumowanie

	częstość	procent	procent ważnych
szkoła nieselekcyjna	31	20,7	22,0
szkoła umiarkowanie selekcyjna	100	66,7	70,9
szkoła selekcyjna	10	6,7	7,1
ogółem	141	94,0	100,0
braki danych	9	6	

Ostatnim etapem przygotowania danych do analiz hierarchicznych było przekodowanie typologii selekcji na zmienne dychotomiczne (o formacie zero-jedynkowym), które tylko w takiej postaci mogą zostać umieszczone w tego rodzaju analizach. W naszych analizach zostały wykorzystane dwie zmienne określające typ selekcji w szkole:

- SL-A – szkoła umiarkowanie selekcyjna,
- SL-B – szkoła silnie selekcyjna.

Zdecydowano, by kategorią odniesienia były szkoły nieselekcyjne. Dzięki temu interpretacja otrzymanych wyników będzie łatwiejsza – otrzymane efekty będzie można tłumaczyć w kategoriach zmiany typu szkoły z nieselekcyjnej na umiarkowanie selekcyjną i selekcyjną.

8.5. Schemat analizy danych

W celu sprawdzenia, w jakim stopniu selekcja uczniów przyjmowanych do pierwszych klas wpływa na ich osiągnięcia szkolne, policzono szereg hierarchicznych modeli liniowych. Zastosowano trzypoziomowe modele z efektem losowym dla stałej regresji, związanym z podziałem na szkoły i klasy. Wykorzystywano dwie zmienne zależne: łączny wynik egzaminu gimnazjalnego z matematyki i przedmiotów przyrodniczych oraz łączny wynik egzaminu gimnazjalnego z języka polskiego i przedmiotów humanistycznych. Obie zmienne zostały przekształcone na skalę o średniej 100 i odchyleniu standardowym 15 w populacji uczniów. Zmienne zostały dokładnie opisane w rozdziale 1.

Analizy pokazujące związek między typologią selekcyjności szkoły a osiągnięciami szkolnymi i EWD, przeprowadzono w dwóch zasadniczych etapach. Pierwszy z nich miał na celu pokazanie relacji dwóch zmiennych dychotomicznych, reprezentujących typ selekcji występujący w szkole, z wynikami egzaminu. Drugi etap analiz miał na celu uchwycenie relacji między wskaźnikami selekcyjności szkoły a osiągnięciami uczniów, przy kontroli wyników na sprawdzianie, płci i informacji o dysleksji. Jest to zatem model, który został liczony analogicznie do modelu EWD. Dodatkowo na drugim etapie przygotowano także modele, w których brano pod uwagę zmienne z poziomu ucznia, które można uznać za wskaźniki inteligencji uczniów (mierzonej testem matrycy Ravena na początku gimnazjum), wykształcenia (liczba lat nauki lepiej wykształconego rodzica) i pozycji społecznej (wyższa pozycja rodzica na międzynarodowej skali pozycji społecznej HTREI).

Analizy wykonano za pomocą oprogramowania HLM 6.06, wykorzystując metodę estymacji *full maximum likelihood*. Do oszacowania istotności efektów stałych wykorzystano odporne (robust) błędy standardowe. Wartości istotne na poziomie istotności $p < 0,05$ zaznaczono

pogrubionym drukiem. Analizy wykonano z uwzględnieniem wag warunkowych na wszystkich trzech poziomach, odpowiednio przeskalowanych za pomocą programu HLM (Scientific Software International, 2012).

8.6. Wyniki dla części matematyczno-przyrodniczej

Pierwszym z modeli, który wyliczono dla opisu zależności pomiędzy selekcyjnością szkół a wynikami egzaminacyjnymi, był model pusty, w którym znajdowała się jedynie zmienna zależna – wynik z egzaminu gimnazjalnego w części matematyczno-przyrodniczej. Wartość stałej regresji w tym modelu (1) wynosi 99,8, co oznacza, że wyłączenie z analizy 9 szkół, dla których niemożliwe było przypisanie kategorii w ramach typologii selekcji nie zmieniło znacząco średnich wyników w rozpatrywanej próbie wyników (wynoszą one 100). Model 2 opisuje zależność między selekcją w szkole a wynikami egzaminu. Należy zauważyć, że istotny efekt wykazuje tutaj tylko jeden typ szkoły – szkoła selekcyjna. Ponieważ zmienne opisujące selekcję uczniów w szkole są dychotomiczne, oznacza to, że średnie wyniki szkół zaliczonych do grupy wysoce selekcyjnej są o prawie 13 punktów wyższe od średnich wyników szkół nie-selekcyjnych. Należy uznać to za bardzo silny efekt, na poziomie prawie jednego odchylenia standardowego wyniku egzaminu gimnazjalnego w części matematyczno-przyrodniczej.

Model 3 został przygotowany analogicznie do sposobu obliczania wskaźnika EWD. Na poziomie ucznia, jako zmienne zależne zostały wprowadzone wyniki uzyskane na sprawdzianie, informacje o płci i dysleksji na zdawanych egzaminach. Model 4 jest analogiczny do modelu 3, z tym że na poziomie szkoły zostały włączone do analizy wskaźniki selekcji. Zmienne SL-A i SL-B okazały się nie mieć istotnego efektu w modelach 3 i 4. Podobne rezultaty możemy zauważyć w modelu 5, w którym dodatkowo włączono do analizy wskaźniki opisujące inteligencję ucznia oraz status społeczno-ekonomiczny jego rodziny. Można zatem stwierdzić, że nie zaobserwowano istotnej statystycznie zależności między wskaźnikami selekcji szkolnej a osiągnięciami uczniów kończących gimnazjum, przy kontroli ich wcześniejszych wyników egzaminacyjnych, płci i dysleksji.

8.7. Wyniki dla części humanistycznej

W przypadku części humanistycznej możemy zaobserwować podobny schemat wyników, jak ten otrzymany we wcześniejszej analizie. Najistotniejsze wydaje się być porównanie modeli 2 i 4. Pierwszy z nich (model 2) prezentuje wyniki analizy, w której uwzględniono wynik na egzaminie w części humanistycznej jako zmienną zależną oraz dwa wskaźniki typów selekcji w szkołach (SL-A i SL-B). Należy zauważyć, że istnieje istotny, silny i pozytywny efekty wskaźnika

Tabela 3. Selekcja w szkole a wyniki z egzaminu w części matematyczno-przyrodniczej. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

oszacowanie efektów stałych						
poziom uczeń	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)
sprawdzian						
sprawdzian ²					11,94 (0,268)	11,92 (0,269)
sprawdzian ³					1,22 (0,121)	1,21 (0,119)
pleć					-0,40 (0,063)	-0,40 (0,063)
dysleksja na egz. gimnazjalnym					-0,96 (0,295)	-0,97 (0,296)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym					1,98 (0,844)	1,97 (0,845)
test matryc Ravena (I pomiar)					-4,20 (0,895)	-4,20 (0,897)
liczba lat nauki lepiej wykształconego z rodziców						3,84 (0,20)
wyższa z pozycji rodziców na międzynarodowej skali pozycji społecznej (HTREI)						0,28 (0,07)
stała	99,81 (0,53)		98,81 (0,69)		100,00 (0,351)	99,80 (0,603)
poziom szkoły						
szkola umiarkowanie selekcyjna (SL-A)			0,607 (0,890)		0,128 (0,651)	-0,254 (0,601)
szkola selekcyjna (SL-B)			12,771 (4,883)		2,623 (1,571)	1,784 (0,965)
oszacowanie efektów losowych						
wariancja efektów szkół	20,46		13,13		4,31	
wariancja efektów oddziałów klasowych	20,93		20,92		5,41	
wariancja na poziomie ucznia	158,26		158,36		65,42	
współczynniki korelacji wewnątrzgrupowej						
poziom szkół	0,102		0,068		0,057	
poziom klas	0,105		0,109		0,072	
poziom szkół i klas łącznie	0,207		0,177		0,129	
podsumowanie						
deviance	35507,79		35478,40		31479,06	
liczba szacowanych parametrów	4		6		10	
AIC	35516		35490		31499	
BIC	35541		35529		31563	
liczba uczniów	4439		4439		4439	
liczba klas	273		273		273	
liczba szkół	141		141		141	

Tabela 4. Selekcja w szkole a wyniki z egzaminu w części humanistycznej. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)
oszacowanie efektów stałych										
poziom ucznia										
sprawdzian					11,95	(0,260)	11,93	(0,257)	10,10	(0,290)
sprawdzian ²					-0,46	(0,137)	-0,47	(0,141)	-0,49	(0,133)
sprawdzian ³					-0,39	(0,087)	-0,40	(0,088)	-0,36	(0,087)
pleć					1,76	(0,341)	1,75	(0,340)	1,53	(0,321)
dysleksja na egz. gimnazjalnym					1,87	(0,676)	1,87	(0,674)	1,17	(0,664)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym					-3,02	(0,811)	-3,02	(0,811)	-2,34	(0,788)
test matryc Ravena (I pomiar)									2,20	(0,20)
liczba lat nauki lepiej wykształconego z rodziców									0,69	(0,08)
wyższa z pozycji rodziców na międzynarodowej skali pozycji społecznej (HTREI)									0,05	(0,020)
stała	100,65	(0,51)	99,52	(0,766)	101,07	(0,372)	100,81	(0,779)	90,40	(1,115)
poziom szkoły										
szkoła umiarkowanie selekcyjna (SL-A)			0,873	(0,943)			0,263	(0,800)	-0,158	(0,791)
szkoła selekcyjna (SL-B)			11,416	(4,286)			1,89	(1,702)	0,27	(1,216)
oszacowanie efektów losowych										
wariancja efektów szkół	18,29		12,51		5,37		5,17		4,83	
wariancja efektów oddziałów klasowych	17,21		17,17		2,00		2,02		1,50	
wariancja na poziomie ucznia	174,87		175,00		78,03		78,04		72,86	
współczynniki korelacji wewnątrzgrupowej										
poziom szkół	0,087		0,061		0,063		0,061		0,061	
poziom klas	0,082		0,084		0,023		0,024		0,019	
poziom szkół i klas łącznie	0,169		0,145		0,086		0,084		0,080	
podsumowanie										
deviance	35897,56		35872,62		32167,26		32165,03		31848,58	
liczba szacowanych parametrów	4		6		10		12		15	
AIC	35906		35885		32187		32189		31879	
BIC	35931		35923		32251		32266		31975	
liczba uczniów	4439		4439		4439		4439		4439	
liczba klas	273		273		273		273		273	
liczba szkół	141		141		141		141		141	

SL-B. Szkoły selekcyjne otrzymują średnio o nieco ponad 11 punktów więcej od szkół nieselekcyjnych. Jest to bardzo silny efekt, na poziomie 76% odchylenia standardowego. Włączenie do analiz wskaźnika wcześniejszych osiągnięć szkolnych (wynik sprawdzianu), informacji o płci i dysleksji skutkuje zmniejszeniem tego wpływu (model 4 – analogiczny do modelu EWD). Wskaźnik określający szkołę selekcyjną staje się nieistotny. Podobne rezultaty możemy zauważyć w modelu 5, w którym dodatkowo uwzględnione zostały informacje dotyczące inteligencji ucznia i jego statusu społeczno-ekonomicznego.

8.8. Podsumowanie

W prezentowanym rozdziale przedstawiono wyniki analiz mających na celu zbadanie relacji między selekcją, przeprowadzaną przez gimnazja w momencie naboru do klas pierwszych, a osiągnięciami uczniów mierzonymi wynikami egzaminacyjnymi oraz wskaźnikami EWD. Jako wskaźniki selekcji przyjęto dwie dychotomiczne zmienne, określające szkołę jako umiarkowanie selekcyjną lub selekcyjną – punktem odniesienia dla tych dwóch kategorii był typ szkoły nieselekcyjnej. Typologia selekcji w szkołach została opracowana na podstawie wskaźników określających odsetek uczniów spoza rejonu w szkole oraz wskaźnik biorący pod uwagę liczbę osób spoza rejonu, które chciały dostać się do szkoły. By wyeliminować losowy wpływ rekrutacji z danego roku, oraz biorąc pod uwagę to, że analizowane zjawisko ma być odniesione do względnie trwałej cechy szkoły, wskaźniki zostały opracowane na podstawie danych o rekrutacji do klas pierwszych badanych szkół z lat 2007 – 2012.

Otrzymane wyniki wskazują, że szkoły zaklasyfikowane do grupy silnie selekcyjnych mają wyraźnie wyższe średnie wyniki egzaminów końcowych. Przewaga ta okazała się bardzo podobna w obu częściach egzaminu gimnazjalnego. W przypadku części matematyczno-przyrodniczej jest to około 13 punktów, natomiast w części humanistycznej około 11 punktów.

Uwzględnienie w analizach zmiennych wykorzystywanych do modelowania EWD, czyli wyników na sprawdzianie, informacji o dysleksji na egzaminach oraz płci, zmienia znacznie rezultaty analiz. W modelach EWD wskaźniki określające typ selekcji okazały się nie mieć istotnego statystycznie znaczenia. Zatem możemy uznać, że w przypadku modeli biorących pod uwagę wcześniejsze osiągnięcia szkolne, szkoły różniące się od siebie nasileniem selekcji nie różnią się wartościami wskaźników EWD. Nie zaobserwowano wpływu selekcji uczniów rozpoczynających naukę w szkole na tego rodzaju miary. Należy w tym miejscu zaznaczyć, że przyjęty sposób tworzenia typologii selekcji jest jednym z możliwych, a jego wartość należy rozpatrywać głównie użytkowo. Potrzebne są dalsze analizy, które mogą prowadzić do stworzenia wskaźników selekcji, opartych na danych zebranych w badaniach podłużnych.

Podsumowując znaczenie otrzymanych wyników dla trafności modeli EWD należy uznać, że stanowią one pośredni dowód na trafność tej metody. Selekcja szkolna ma wpływ na końcowe wyniki egzaminacyjne. Jednak, przy kontroli wyników sprawdzianu (czyli na wejściu do szkoły), wpływ praktyk selekcyjnych staje się nieistotny. Możemy zatem dojść do wniosku, że modele EWD liczone dla gimnazjów w dostateczny sposób kontrolują wpływ praktyk selekcyjnych odbywających się w szkołach.

Literatura

- Allen, R. (2006). *What should an index of school segregation measure?* London: Centre for the Economics of Education, London School of Economics and Political Science.
- Bourdieu, P., Passeron, J. (1990). *Reprodukcja. Elementy teorii systemu nauczania*. Warszawa: PWN.
- Bradley, S., Taylor, J. (2002). The Effect of the Quasi-market on the Efficiency-equity Trade-off in the Secondary School Sector. *Bulletin of Economic Research*, 54(3), 295–314.
- Chaber, A. (2000). *Bon Edukacyjny: Kontrowersje i Realia* (praca magisterska). Instytut Socjologii Uniwersytetu Warszawskiego, Warszawa.
- Clark, W. A. V., Maas, R. (2012). Schools, Neighborhoods and Selection: Outcomes Across Metropolitan Los Angeles. *Population Research and Policy Review*, 31(3), 339–360.
- Dolata, R. (2008). *Szkoła-segregacja-nierówności*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Dolata, R. (2011). *Analiza różnicowania się systemu oświaty w Polsce na poziomie szkół podstawowych i gimnazjów, edycja 2011*. Warszawa: CKE.
- 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*, (1).
- Domański, H. (2004). *Struktura społeczna*. Warszawa: Scholar.
- Elacqua, G. (2012). The impact of school choice and public policy on segregation: Evidence from Chile. *International Journal of Educational Development*, 32(3), 444–453.
- Exley, S. R. (2009). Exploring pupil segregation between specialist and non-specialist schools. *Oxford Review of Education*, 35(4), 451–470.
- Goldstein, H., Noden, P. (2003). Modelling Social Segregation. *Oxford Review of Education*, 29(2), 225–237.

- Gorard, S., Cheng, S. C. (2011). Pupil clustering in English secondary schools: one pattern or several? *International Journal of Research in Method in Education*, 34(3), 327–339.
- Gorard, S., Taylor, C., Fitz, J. (2001). Social exclusion and public policy: the relationship between local school admission arrangements and segregation by poverty. *International Journal of Sociology and Social Policy*, 21(4/5/6), 10–36.
- Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, 24(4).
- Jakubowski, M. (2007). Efektywność wydatków na gimnazja. *Gospodarka Narodowa*, (11–12).
- Jasińska, A. (2010). *Selekcje na progu gimnazjum a jego wyniki egzaminacyjne. Niepublikowany raport*. Warszawa: Centralna Komisja Egzaminacyjna.
- Ladd, H. F. (2002). School Vouchers: A Critical View. *Journal of Economic Perspectives*, 16(4), 3–24.
- Śleszyński, P. (2004). *Ekonomiczne uwarunkowania wyników sprawdzianów szóstoklasistów i egzaminu gimnazjalnego przeprowadzonych w latach 2002–2004*. MENiS.
- Waslander, S., Pater, C., van der Weide, M. (2010). *Markets in Education: An Analytical Review of Empirical Research on Market Mechanisms in Education* (OECD Education Working Papers). OECD.
- Wosik, E. (2002). *Zmiany w systemie oświaty. Wyniki badań empirycznych*. Warszawa.
- Żółtak, T. (2011). Znaczenie informacji o średnim wyniku uczniów na wejściu dla własności jednorocznych wskaźników EWD gimnazjów. [w:] *Ewaluacja w edukacji: koncepcje, metody, perspektywy*. Kraków: Grupa Tomami.

Rozdział 9

Style zarządzania szkołą

Jednym z badanych zagadnień był związek stylów zarządzania szkołą z miarami EWD. Jeśli EWD jest cechą zależną od sposobu funkcjonowania szkoły, istniejący w szkole styl zarządzania powinien być związany z poziomem tego wskaźnika. Zgodnie z przyjętym planem analizy trafności metody EWD, analizy przedstawione w tym rozdziale pozwolą ocenić ważny aspekt trafności kryterialnej. Czy szkoły uzyskujące wysokie wskaźniki EWD są zarządzane w sposób, który budzi nasz niepokój, czy też odwrotnie, odnajdujemy w nich przejawy stylów zarządzania adekwatnych do edukacyjnej misji szkoły? Podjęte analizy są ważne również wtedy, gdy szukamy powiązania EWD z takimi cechami szkoły, które są od niej zależne – czyli można je zmieniać w ramach funkcjonowania szkół. Należy zauważyć, że celem analiz rozpatrywanych w tym rozdziale nie było szukanie potwierdzenia funkcjonowania założonego w teorii mechanizmu przyczynowo-skutkowego. Skupiono się na analizach prowadzących do wykrycia związków między praktykami zarządzania a wynikami egzaminacyjnymi rozpatrywanymi w modelu EWD.

W prezentowanych poniżej analizach skupiono się na danych zebranych od nauczycieli. Badanymi aspektami w ramach stylów zarządzania były partycypacja oraz kultura organizacyjna. Sposób ich rozumienia, konceptualizacji, a także pomiar może być bardzo odmienny, w zależności od przyjętego podejścia (Bush, 2003). Za wybór koncepcji teoretycznej oraz opracowanie wstępnych wersji narzędzi badawczych odpowiadało Interdyscyplinarne Centrum Badań i Rozwoju Organizacji przy Zakładzie Psychologii Społecznej Instytutu Psychologii UJ (ICBiRO).

Ze względu na użycie w badaniach nowatorskich narzędzi badawczych przeprowadzone analizy miały charakter eksploracyjny. Były w nich wykorzystywane wskaźniki dotyczące postrzegania przez nauczycieli różnych wymiarów funkcjonowania szkoły jako organizacji. Założono, że skale zbudowane na podstawie opinii nauczycieli są dobrym odzwierciedleniem rzeczywistego funkcjonowania szkoły w analizowanych obszarach. Należy zauważyć, że istnienie wpływu zarządzania na osiągnięcia uczniów jest raczej pośrednie – pewne praktyki w tym zakresie wpływają na pracę nauczycieli, którzy następnie wpływają bezpośrednio na osiągnięcia uczniów. Dokładny opis tego rodzaju powiązań postulowanych teoretycznie, oraz wyniki badań empirycznych nad nimi, zostaną przedstawione bliżej w części rozdziału poświęconej koncepcjom teoretycznym leżącym u podstaw wybranych sposobów pomiaru interesujących nas czynników stylu zarządzania szkołą.

9.1. Podstawy teoretyczne

9.1.1. Partycypacja w zarządzaniu

Definicje pojęcia partycypacji różnią się w zależności od przyjętego kontekstu badawczego, często ograniczają się do wskazania tych aspektów funkcjonowania szkoły, na które partycypacja ma wpływ (Kołodziejczyk, 2011). W najbardziej ogólnym ujęciu, z perspektywy nauczyciela możemy partycypację rozumieć jako stopień, w jakim osoby będące pracownikami szkoły, mogą wpływać na funkcjonowanie szkoły (Conley, 1991). W badaniach wykorzystujących metody ilościowe, partycypację definiuje się często jako postrzeganie przez pracowników ich wpływu na podejmowanie decyzji w szkole (Miller i Rowan, 2006). Charakter tych decyzji oraz kwestie, których one dotyczą, zależą od ograniczeń prawnych, ale należy zauważyć, że nawet w systemach określających w dużym stopniu hierarchicznie obowiązki nauczyciela i dyrektora, istnieją możliwości włączenia tych pierwszych w podejmowanie decyzji dotyczących ważnych aspektów funkcjonowania szkoły, zwłaszcza na poziomie klasy (Tołwińska, 2011a).

Partycypacja nauczycieli w zarządzaniu szkołą pozwala myśleć o niej jako o instytucji, w której istnieje o wiele większy potencjał rozwoju niż w biurokratycznie zarządzanej organizacji. Dzieje się tak dzięki demokratyzacji procesu podejmowania decyzji i przyjęcia przez nauczycieli odpowiedzialności za funkcjonowanie szkoły (Bauch i Goldring, 1998). Oczywiście należy zauważyć, że odmiennie niż organizacje czysto biznesowe, szkoły są instytucjami, które działają w przestrzeni ściśle określonej przez prawo oświatowe – mają one zadania, cele i sposoby ich realizacji, które są z góry określone i w dużej części niemożliwe do modyfikacji (Bush, 2003). Obecnie jednak obserwujemy w Polsce proces demokratyzacji szkoły. Partycypacja nauczycieli w zarządzaniu szkołą jest jednym z elementów koncepcji nowego nadzoru pedagogicznego (Mazurkiewicz, 2012). W tym kontekście badanie sposobów zarządzania szkołą w nurcie partycypacyjnym wydaje się być szczególnie interesujące – jest zgodne z kierunkiem zmian, następujących w postrzeganiu szkoły przez nadzór pedagogiczny.

Zagadnienie partycypacji nauczycieli w zarządzaniu szkołą ma długą tradycję badawczą w krajach zachodnich (Somech, 2010). Istnieje wiele badań analizujących wpływ partycypacji nauczycieli w podejmowaniu decyzji w szkole na różne aspekty jej funkcjonowania. Można zauważyć dwa główne nurty. Pierwszy z nich skupia się na analizach jej wpływu na różne charakterystyki nauczycieli jako pracowników szkoły. Sharon Conley (1991), autorka artykułu podsumowującego badania dotyczące partycypacji nauczycieli, wskazuje w tym kontekście między innymi na badania dotyczące satysfakcji z pracy, postrzeganego poziomu stresu, lojalności w stosunku do miejsca pracy, postrzeganej efektywności pracy szkoły, współpracy oraz alienacji. Drugi nurt badawczy wiąże się

z wpływem partycypacji na osiągnięcia szkolne uczniów. Rezultaty badań w tym nurcie są niejednoznaczne – niektóre badania wykazują słabe pozytywne efekty postrzeganego przez nauczycieli wpływu na decyzje podejmowane w szkole na osiągnięcia (Rowan, Chiang i Miller, 1997), w innych wyniki nie są rozstrzygające (Miller i Rowan, 2006; Raudenbush, Rowan i Kang, 1991). Związki partycypacji z osiągnięciami uczniów są raczej pośrednie – uczestnictwo nauczycieli w podejmowaniu decyzji wpływa na inne zmienne, które są kluczowe dla osiągnięć uczniów – w tym kontekście najczęściej wskazuje się na sposób prowadzenia zajęć przez nauczycieli (Marks i Louis, 1997; Smylie, Lazarus i Brownlee-Conyers, 1996). Istnieją zatem przesłanki empiryczne, by uznać, że partycypacja może być czynnikiem, leżącym po stronie szkoły, który może wpływać na efektywność nauczania.

W Polsce badania partycypacji w zarządzaniu w szkole są prowadzone, chociaż w większości sprowadzają się do rozważań teoretycznych dotyczących przebiegu samego procesu oraz możliwych implikacji rozszerzenia partycypacji w szkole dla jej funkcjonowania (Prokopowicz i Żmuda, 2009; Tołwińska, 2011a). Prowadzone są także badania empiryczne w tym zakresie, jednak skupiają się one zwykle tylko na pomiarze opinii osób zaangażowanych w życie szkoły dotyczących ich partycypacji w jej zarządzaniu (Kołodziejczyk, 2011; Szymański, 2001; Tołwińska, 2011b).

Oprócz zagadnień bezpośrednio związanych z partycypacją w naszym badaniu zmierzono o zmienne, które wpływają na możliwość pełnego wykorzystania w szkole partycypacyjnego stylu zarządzania (Heller, Pusic, Strauss i Wilpert, 1998). Są to:

- zaufanie do współpracowników,
- zaufanie do przełożonego,
- możliwość uzyskania informacji zwrotnej o swojej pracy,
- możliwość samorozwoju,
- identyfikacja z instytucją.

Wymienione powyżej cechy szkoły jako organizacji, są istotne z punktu widzenia rozwoju w niej zarządzania partycypacyjnego. Zaufanie, zarówno do współpracowników, jak i do przełożonego, jest powszechnie uważane za warunek konieczny partycypacji w zarządzaniu. Bez niego, ten styl zarządzania może stać się bezproduktywny, głównie z powodu konfliktów, które są naturalnym czynnikiem składowym dopuszczenia pracowników do procesu podejmowania decyzji. Brak zaufania w dobre intencje wszystkich zaangażowanych w proces zarządzania, może doprowadzić do załamania partycypacyjnego stylu zarządzania (Heller i in., 1998).

Następnym czynnikiem kluczowym dla poprawnego funkcjonowania partycypacji jest możliwość uzyskania informacji zwrotnej o swojej pracy. Brak informacji o efektach pracy osób może stanowić problem, ponieważ osoby zaangażowane w partycypację powinny mieć możliwość rozwoju osobistego, także w kontekście wiązania podejmowanych działań z ich skutkami. Bez tego

mechanizmu partycypacja może nie dać pozytywnych rezultatów w konsekwencji niewystarczającej wiedzy osób zaangażowanych o skuteczności podjętych przez nich działań (Heller i in., 1998).

Dwie ostatnie badane cechy systemu zarządzania szkołą to możliwość samorozwoju i identyfikacji z instytucją. Można je traktować jak konsekwencje istnienia w organizacji partycypacji oraz równocześnie jako składowe mechanizmu zwiększające jego skuteczność (Glew, O’Leary-Kelly, Griffin i Van Fleet, 1995). Obserwowana przez nauczycieli możliwość samorozwoju jest związana ze stopniem, w jakim postrzegają oni szkołę jako miejsce, które umożliwia im rozwój. Natomiast identyfikacja z instytucją zwiększa stopień utożsamiania osoby z celami funkcjonowania szkoły. Nie są one kluczowymi elementami warunkującymi wystąpienie partycypacji, jednak powinny one pojawić się w miarę rozwoju tego rodzaju zarządzania. Równocześnie mogą one istnieć w organizacjach niezależnie od partycypacji w zarządzaniu, stanowiąc okoliczności ułatwiające wprowadzenie tego rodzaju rozwiązań.

9.1.2. Kultura organizacyjna szkoły

W przypadku kultury organizacyjnej szkoły mamy do czynienia z pojęciem o bardzo szerokim polu znaczeniowym. Właściwie wszystkie konstrukty teoretyczne opisane powyżej są w niektórych przypadkach wskazywane jako aspekty kultury organizacyjnej szkoły (Augustyniak, 2009). Przywództwo, sposób podejmowania decyzji, zaufanie, identyfikacja z instytucją są rozpatrywane w kontekście kultury organizacyjnej jako jej wymiary lub aspekty będące z nią w bliskiej relacji (Hargreaves, 1995; Masłowski, 2001). Jednak podstawowa różnica między wyróżnionymi wcześniej czynnikami a kulturą organizacyjną szkoły to istnienie pewnych nieformalnych zestawów przekonań, wartości, podstawowych założeń dotyczących funkcjonowania szkoły i postrzegania tych zagadnień przez jej personel (Masłowski, 2001; Schein, 2010). Trudno wskazać na badania, które specyficznie odnoszą się do tak postrzeganej koncepcji kultury organizacyjnej szkoły, głównie ze względu na fakt problemów definicyjnych i pokrywanie się tego pojęcia z innymi aspektami funkcjonowania szkoły jako organizacji, wspomnianymi powyżej. Badania wskazujące w tytule na odniesienie się do kultury organizacyjnej po głębszej lekturze okazują się być zwykle poświęcone elementom zarządzania szkołą, które łącznie określa się mianem kultury organizacyjnej (Tsereteli, Martskvishvili i Aptarashvili, 2011; Zhu, Devos i Li, 2011).

W podłużnych badaniach przeprowadzanych w ramach projektu EWD zdecydowano się na wybór narzędzia do badania kultury organizacyjnej, opartego na koncepcji Kima Camerona i Roberta Quinna (2006). Wychodząc od definicji, według której kultura organizacyjna to zespół przekonań, norm i wartości powszechnie akceptowanych w organizacji i stanowiących spójny system, wyróżnili oni dwa główne wymiary tak pojmowanego zjawiska: orientacja na otoczenie instytucji vs orientacja na wnętrze instytucji oraz elastyczność vs stabilność. Złożenie tych dwóch wymiarów pozwala wyróżnić cztery podstawowe typy kultury:

- Kultura klanu – jest to organizacja charakteryzująca się naciskiem na stabilność wewnętrzną oraz elastyczność. Jest ona przyjazna pracownikom, którzy chętnie ze sobą współpracują – więzy między pracownikami mogą nabrać charakteru osobistego, prawie rodzinnego. Lider może przyjmować rolę mentora. Istotnym elementem jest także rozwój instytucji przez długoterminowe korzyści płynące z rozwoju osobistego pracowników (nauczycieli). Widoczny jest nacisk na spójność grupową w ramach personelu. Ważnymi cechami organizacji, które stają się wyznacznikiem jej sukcesu, są lojalność, zaangażowanie, praca zespołowa oraz konsensus.
- Kultura adhokracji – organizacja zorientowana na otoczenie oraz elastyczna w działaniu. Istnieje w niej nacisk na kreatywność, dynamikę, przedsiębiorczość. Liderzy pełnią rolę innowatorów, wytyczają długoterminowe cele. Spójność grupowa jest zapewniana przez ciągłe dążenie do wprowadzania nowych sposobów postępowania, wprowadzania innowacji. Istotnym staje się fakt bycia liderem na tym polu. Organizacja jest zorientowana na ciągły rozwój, także przez podejmowanie wyzwań i pozyskiwanie nowych zasobów. Pracownicy są zachęceni do wykazywania się inicjatywą, mają duże możliwości samodzielnego działania. Unikalna oferta, niestandardowe sposoby postępowania stają się istotne dla postrzegania sukcesu w tego rodzaju organizacji.
- Kultura rynku – istotne wartości to stabilność i kontrola, przy jednoczesnym nastawieniu na otoczenie. Organizacja tego typu jest nastawiona na realizację zadań i ich rezultaty. Pracowników można scharakteryzować jako ambitnych i zorientowanych na realizację celów. Przywódca jest nastawiony na rezultaty, konkurowanie z innymi. Organizacja jest spójna dzięki dążeniu do osiągnięcia sukcesu w kategorii bycia lepszą od innych, podobnych. Istotne są tutaj: reputacja, rezultaty i sukces w osiąganiu celów. Wskaźnikiem realizacji celów organizacji jest jej pozycja względem innych.
- Kultura hierarchii – organizacja, którą można scharakteryzować jako nastawioną na stabilność, kontrolę oraz integrację wewnętrzną. Wysoce sformalizowana instytucja, w której obowiązują ściśle procedury oraz sztywna hierarchia. Przywódca to przede wszystkim sprawny organizator i koordynator działań pracowników, który nastawiony jest na efektywność działania. Celem organizacji jest sprawne funkcjonowanie, które ma być zapewnione przez formalne przepisy i procedury postępowania. One także zapewniają instytucji spójność. Organizacja ma być miejscem stabilnym, funkcjonującym wydajnie oraz bez zakłóceń. Pracownicy skupiają się na bezpieczeństwie zatrudnienia, stałości i przewidywalności działań. Sukces postrzegany jest w tej organizacji jako przewidywalność rezultatów, dotrzymywanie formalnych zobowiązań, brak generowania dodatkowych kosztów dla jej funkcjonowania.

9.2. Adaptacja narzędzi badawczych kierowanych do nauczycieli

W pomiarze różnych aspektów stylów zarządzania szkołą zastosowano zróżnicowane narzędzia badawcze. Pierwotnie większość kwestionariuszy była używana w badaniach przedsiębiorstw, należało zatem przeprowadzić ich adaptację nie tylko pod względem językowym (tłumaczenie na język polski), ale także kulturową (dopasowanie do kontekstu funkcjonowania szkoły). Wstępne wersje narzędzi badawczych zostały przygotowane przez Interdyscyplinarne Centrum Badań i Rozwoju Organizacji (ICBiRO) przy Zakładzie Psychologii Społecznej Instytutu Psychologii UJ. W przypadku partycypacji i zmiennych jej towarzyszących, zostały one oparte o indeks partycypacji, przygotowany przez ICBiRO do badania tego aspektu zarządzania w przedsiębiorstwach. Natomiast narzędzie do badania kultury organizacyjnej to adaptacja kwestionariusza do oceny kultury organizacji *Organizational Culture Assessment Instrument* (w skrócie OCAI), często używanego do badań organizacji (Hartnell, Ou i Kinicki, 2011). Narzędzie badawcze dla badania wskaźników partycypacji (oraz innych czynników z nią związanych) przyjęły formę baterii pytań, z odpowiedziami w formie skali Likerta. Kwestionariusz do oceny kultury organizacji (OCAI) miał inną strukturę, która wymagała od badanych przyznania punktów w zależności od stopnia zgodności ich przekonań z podanymi stwierdzeniami.

Praca nad językową i kontekstową adaptacją została zorganizowana według następujących zasad. Pracownicy ICBiRO przesłali do zespołu badawczego EWD przetłumaczone na język polski pytania, wraz z opisem konstruktywów teoretycznych, do których się odnosiły. W ramach zespołu została przeprowadzona ich wstępna adaptacja kulturowa, dopasowująca pytania do kontekstu pracy szkoły i specyfiki jej funkcjonowania. Następnie poprawione pytania wraz z komentarzem były odsyłane z powrotem do przedstawicieli ICIBRO, którzy odnosili się do uwag zespołu z punktu widzenia spójności z konstruktem teoretycznym, który leżał u podstawy danego narzędzia. W rezultacie tego procesu wypracowano wersje narzędzi badawczych, które przedstawiono wybranym nauczycielom. Przeprowadzono z nimi wywiady jakościowe, które miały na celu sprawdzenie sposobów rozumienia pytań. Na podstawie wywiadów ostatecznie zmodyfikowano pytania.

9.3. Schemat analizy danych

W celu sprawdzenia, w jakim stopniu partycypacja i kultura organizacyjna związane są z osiągnięciami uczniów, także w przypadku, gdy kontrolujemy ich wcześniejsze osiągnięcia (czyli w przypadku modelu analogicznego do EWD), policzono szereg hierarchicznych modeli liniowych, które odzwierciedliły złożony schemat doboru próby. Były to modele trzypoziomowe z losową stałą, uwzględniające poziom ucznia, klasy i szkoły. Wykorzystywano

dwie zmienne zależne: łączny wynik egzaminu gimnazjalnego z matematyki i przedmiotów przyrodniczych oraz łączny wynik egzaminu gimnazjalnego z języka polskiego i przedmiotów humanistycznych. Obie zmienne zostały przekształcone na skalę o średniej 100 i odchyleniu standardowym 15 w populacji uczniów. Zmienne zostały dokładnie opisane w rozdziale 1. Używane w analizach średnie oceny nauczycieli na każdej ze skal możemy traktować jako miary postrzegania przez nauczycieli szkoły na danym wymiarze jej funkcjonowania.

Analizy, pokazujące związki różnych aspektów stylów zarządzania z osiągnięciami szkolnymi i EWD, przeprowadzono w dwóch etapach. Pierwszy z nich miał na celu pokazanie związku danej wiązki zmiennych, dotyczącej jednego z wymiarów funkcjonowania szkoły, z wynikami egzaminu. Zmienne niezależne były dodawane na poziomie szkoły, jako zagregowane do średniej opinii nauczycieli na danej skali. Drugi etap to trzypoziomowy model EWD, do którego także dodano analizowane zmienne niezależne na poziomie szkoły. Dzięki temu można wykazać, jakie znaczenie, dla wyników egzaminacyjnych i modelu EWD, ma uwzględnianie danej grupy wskaźników związanych z postrzeganiem przez nauczycieli różnych wymiarów funkcjonowania szkoły jako organizacji. Dodatkowo w drugim etapie przygotowano także model, w którym brano pod uwagę zmienne z poziomu ucznia, które można uznać za zmienne kontrolne: wskaźniki inteligencji uczniów (mierzonej testem matryc Ravena na początku gimnazjum), wykształcenia rodziców (liczba lat nauki lepiej wykształconego rodzica) i pozycji społecznej rodziny (wyższa pozycja rodzica na międzynarodowej skali pozycji społecznej HTREI). Należy zauważyć, że każdy z prezentowanych wymiarów stylów zarządzania szkołą (partycypacja, kultura organizacyjna) był analizowany osobno. Ten sposób postępowania był podyktowany istnieniem teorii leżących u ich podstaw oraz tłumaczących związki między poszczególnymi czynnikami w ich ramach.

Opisane powyżej analizy można sprowadzić do analizy pięciu modeli. Pierwszy z nich, model pusty, stanowi punkt odniesienia do pozostałych modeli. Pozwala na oszacowania stopnia, w jakim na poziom danej zmiennej zależnej (wynik egzaminu w części humanistycznej lub w części matematyczno-przyrodniczej) wpływa fakt zmniejszenia liczebności próby, spowodowanego brakami danych. Drugi model zawiera zmienne na poziomie szkoły. Trzeci model jest modelem analogicznym do EWD – uwzględnił wyniki sprawdzianu, płeć i dysleksję na egzaminach. W następnym z modeli (4) dodano zmienne z poziomu szkoły. Ostatni z modeli (5) zawiera wszystkie zmienne z modelu (4) oraz dodatkowo zmienne będące wskaźnikami inteligencji ucznia i jego pozycji społecznej.

Analizy wykonano za pomocą oprogramowania HLM 6.06, wykorzystując metodę estymacji *full maximum likelihood*. Do oszacowania istotności efektów stałych wykorzystano odporne (*robust*) oszacowania błędów standardowych. Wartości istotne na poziomie istotności $p < 0,05$ zaznaczono pogrubionym drukiem. Analizy wykonano z uwzględnieniem wag warunkowych na wszystkich trzech poziomach, odpowiednio przeskalowanych za pomocą programu.

9.4. Przygotowanie danych do analizy

9.4.1. Partycypacja w zarządzaniu

Dane uzyskane z 43 pytań kierowanej do nauczycieli ankiety, zostały poddane konfirmacyjnej analizie czynnikowej, w celu zweryfikowania zakładanej struktury czynników. Dodatkowym celem było stworzenie ocen czynnikowych dla skali partycypacji i pozostałych czynników rozpatrywanych w jej kontekście. Przed modelowaniem dane nauczycielskie zostały przygotowane do analizy: usunięto z nich dane od nauczycieli, dla których nie notowano wariancji w odpowiedziach na pytania lub w bazie widniały dla nich jedynie braki danych. Obniżyło to liczbę analizowanych przypadków do 2514, czyli do około 96% wszystkich nauczycieli badanych w III etapie. Analizę przeprowadzono w programie Mplus, przy założeniu zagnieżdżenia nauczycieli w szkołach i kategoryjnej skali odpowiedzi. Założono w modelu istnienie siedmiu czynników, jednak część pytań nie była dobrze reprezentowana w zakładanych czynnikach. Część pytań została po wstępnej analizie przyporządkowana innym czynnikom niż zakładane (stało się tak w przypadku 3 pytań), natomiast część została usunięta (6 pytań). Otrzymane wyniki wskazują na dobre dopasowanie modelu do otrzymanych danych: RMSEA jest na poziomie 0,031, CFI/TLI są powyżej 0,97.

Ostatecznie przyjęto, że odwzorowano zakładaną strukturę teoretyczną. Na analizowany model w przypadku partycypacji składają się zatem następujące czynniki (w nawiasie podano liczbę pytań, składających się ostatecznie na dany wskaźnik):

- partycypacja decyzyjna (5 pytań) – im wyższy wskaźnik, tym nauczyciel ocenia sytuację w swojej szkole jako bardziej sprzyjającą partycypacji,
- partycypacyjny styl przywództwa (6 pytań) – im wyższy wskaźnik, tym nauczyciel postrzega dyrektora swojej szkoły jako bardziej partycypacyjnego, umożliwiającego uczestnictwo innych osób w podejmowaniu decyzji w szkole, delegującego uprawnienia,
- zaufanie do przełożonego (5 pytań) – im wyższy wskaźnik, tym nauczyciel wykazuje większe zaufanie do dyrektora jako lidera,
- zaufanie do współpracowników (5 pytań) – im wyższy wskaźnik, tym nauczyciel wykazuje większe zaufanie do współpracowników,
- identyfikacja z instytucją (6 pytań) – im wyższy wskaźnik, tym nauczyciel w większym stopniu identyfikuje się ze szkołą, w której pracuje,
- możliwość samorozwoju (6 pytań) – im wyższy wskaźnik, tym nauczyciel w większym stopniu postrzega szkołę jako miejsce, w którym może się sam rozwijać pod względem zawodowym,
- informacja zwrotna (4 pytania) – im wyższy wskaźnik, tym nauczyciel w większym stopniu postrzega szkołę jako miejsce, gdzie otrzymuje dużo informacji dotyczących swojej pracy i postępowania.

Dla wszystkich nauczycieli zapisano wartości czynnikowe dla każdej ze skal. Następnie uzyskane wartości zostały przekształcone tak, by skala była skierowana zgodnie ze swoją nazwą (im wyższa wartość, tym większe nasilenie danej cechy). Wyniki dla skal uzyskane przez nauczycieli zostały zagregowane do poziomu szkoły. Średnie wyniki szkół zostały następnie wystandaryzowane tak, by ich średnia wynosiła 0, a odchylenie standardowe 1. Na poziomie szkoły sprawdzono, czy nie ma przypadków, w których w danej placówce został wyznaczony wynik na podstawie odpowiedzi mniej niż 50% nauczycieli. W przypadku opisywanych skal, nie było takich przypadków, zatem w analizach uwzględniono 4769 uczniów, 291 klas i 150 szkół.

9.4.2. Kultura organizacyjna

Kwestionariusz do oceny kultury organizacji OCAI (*Organizational Culture Assessment Instrument*), skonstruowany na podstawie wyżej opisanej teorii typów kultury, jest narzędziem stosowanym szeroko w badaniach organizacji różnego typu, zwłaszcza biznesowych (Hartnell i in., 2011). Należy jednak zauważyć, że w polskich badaniach edukacyjnych nie stosowano go wcześniej na próbach ogólnopolskich. Kwestionariusz wykorzystany do badania postrzegania kultury organizacyjnej przez nauczycieli, składał się z 24 pytań, pogrupowanych w 6 bloków. Każdy blok dotyczył jednej z kluczowych cech dla pomiaru kultury organizacyjnej szkoły. Były to:

- dominująca cecha szkoły,
- przywództwo w szkole,
- zarządzanie pracownikami,
- spoiwo szkoły,
- wymiar strategiczny,
- kryteria sukcesu.

W ramach bloków pytań znajdowały się cztery stwierdzenia, z których każde było charakterystyczne tylko dla jednego z przedstawianych wcześniej typów kultury organizacyjnej szkoły. Respondenci mieli za zadanie rozdzielić 100 punktów między te cztery stwierdzenia, w zależności od stopnia, w jakim każde z nich pasuje do ich szkoły. Badani byli proszeni o wykonanie tej czynności, skupiając się na perspektywie obecnie postrzeganej w szkole kultury organizacyjnej, jak i preferowanego przez nich stanu rzeczy za trzy lata. Należy w tym miejscu zauważyć, że w tym rozdziale zostaną przedstawione wyniki analiz opierających się na wskaźnikach opisujących kulturę organizacyjną, jaka zdaniem badanych nauczycieli obecnie panuje w szkole.

Przed wyliczeniem końcowych wyników dla respondentów, dane zostały wyczyszczone w celu wyeliminowania błędów polegających na przypisaniu w ramach jednego bloku pytań

złej liczby punktów (ponad 100). Przygotowane w ten sposób wyniki zostały zsumowane zgodnie z założeniami twórców kwestionariusza i podzielone przez liczbę pytań, tworzących dany typ kultury (dla każdego typu było to 6 pytań). W rezultacie dla każdego nauczyciela otrzymano cztery zmienne, z których każda miała zakres od 0 do 100. Wzrost wyników na danej skali oznacza zatem, że dany nauczyciel w większym stopniu postrzega swoją szkołę w kategoriach danej kultury organizacyjnej. W następnym kroku zagregowano wyniki uzyskane w ramach szkoły, licząc średnie wartości dla każdej ze skal. Tak otrzymane wyniki potraktowano jako wskaźniki kultury organizacyjnej szkoły.

Ostatecznie w analizach uwzględniono 4736 uczniów, 289 klas i 149 szkół. Jedną szkołę wyeliminowano z powodu zbyt małego odsetka nauczycieli, którzy odpowiedzieli poprawnie na pytania dotyczące tego zagadnienia (progiem było tutaj, podobnie jak we wcześniejszych analizach, 50% nauczycieli).

9.5. Wyniki

9.5.1. Partycypacja w zarządzaniu

9.5.1.1. Wyniki dla części matematyczno-przyrodniczej

Poniżej prezentowane analizy związku partycypacji i czynników z nią powiązanych z osiągnięciami uczniów i EWD szkoły dotyczą wyników z egzaminu gimnazjalnego w części matematyczno-przyrodniczej. Przypomnijmy, że ta zmienna została przygotowana w taki sposób, jak postępuje się przy okazji analiz EWD, czyli wyniki z obu zakresów egzaminu gimnazjalnego zostały przeskalowane, a następnie przekształcone na skalę o średniej 100 i odchyleniu standardowym 15. Wyniki analiz zostały przedstawione w tabeli 1.

Analizy z wykorzystaniem zmiennych charakteryzujących styl organizacyjny szkoły nie wykazały istotnych statystycznie efektów ani w wypadku osiągnięć uczniów, ani w przypadku kontroli uprzednich osiągnięć (model EWD). Także uwzględnienie innych zmiennych kontrolnych z poziomu ucznia mierzących pozycję społeczną rodzica ucznia (lata nauki dłużej kształcącego się rodzica oraz pozycja na skali HTREI) oraz inteligencję ucznia (wynik na teście matryc Ravena, który odbył się w drugim semestrze pierwszej klasy w gimnazjum) nie zmieniło obrazu zależności. Innymi słowy dodanie zmiennych związanych z partycypacją na poziomie szkoły, nie zmienia znacząco parametrów dopasowania modelu oraz siły efektów losowych (wariancji efektów) na poziomie szkoły. Z tego powodu musimy uznać, że cechy szkoły związane z postrzeganą przez nauczycieli partycypacją i czynnikami z nią związanymi, nie wpływają znacząco na osiągnięcia uczniów w części matematyczno-przyrodniczej rozpatrywane bez kontroli uprzednich osiągnięć, jak i w modelu EWD.

Tabela 1. Partycypacja i czynniki jej towarzyszące a wyniki z egzaminu w części matematyczno-przyrodniczej. Wyniki trzyposziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej									
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)
oszacowanie efektów stałych									
poziom uczenia									
sprawdzian				11,99	(0,267)		11,99	(0,270)	9,53
sprawdzian ²				1,26	(0,117)		1,26	(0,117)	1,22
sprawdzian ³				-0,41	(0,065)		-0,41	(0,065)	-0,383
pleć				-0,98	(0,284)		-0,98	(0,284)	-1,485
dysleksja na egz. gimnazjalnym				1,88	(0,804)		1,89	(0,805)	1,00
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym				-4,044	(0,852)		-4,03	(0,859)	-2,892
test matryc Ravena (I pomiar)									0,05
liczba lat nauki lepiej wykształconego z rodziców									0,32
wyższa z pozycji rodziców na międzynarodowej skali pozycji społecznej (HTREI)									3,808
stała	99,3	(0,528)	100,05	(0,540)	100,00	(0,338)	100,00	(0,332)	93,83
poziom szkoły									
partycypacja decyzyjna			-1,752	(1,006)			-0,158	(0,549)	-0,244
partycypacyjny styl przywództwa			2,158	(1,629)			0,636	(1,060)	0,645
zaufanie do przełożonego			-0,839	(1,473)			-0,080	(0,876)	0,028
zaufanie do współpracowników			1,010	(0,866)			-0,40	(0,544)	-0,35
identyfikacja z instytucją			-0,655	(1,067)			-0,043	(0,893)	-0,376
możliwość samorozwoju			1,685	(1,324)			-0,101	(0,825)	0,069
informacja zwrotna			-1,538	(1,218)			0,326	(0,681)	0,361
oszacowanie efektów losowych									
wariancja efektów szkół	21,29		20,37		4,46		4,36		4,03
wariancja efektów oddziałów	21,16		20,94		5,17		5,14		3,43
wariancja na poziomie ucznia	159,58		159,58		66,41		66,41		58,29
współczynniki korelacji wewnątrzgrupowej									
poziom szkół	0,105		0,101		0,059		0,057		0,061
poziom klas	0,105		0,104		0,068		0,068		0,052
poziom szkół i klas łącznie	0,210		0,206		0,127		0,125		0,114
podsumowanie									
deviance	3818942		3818422		3388448		3388216		3323372
liczba szacowanych parametrów	4		11		10		17		20
AIC	38197		38206		33904		33916		33274
BIC	38223		38277		33969		34026		33403
liczba uczniów	4769		4769		4769		4769		4769
liczba klas	291		291		291		291		291
liczba szkół	150		150		150		150		150

Tabela 2. Partycypacja i czynniki jej towarzyszące a wyniki z egzaminu w części humanistycznej. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części humanistycznej						
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)
oszacowanie efektów stałych						
poziom ucznia						
sprawdzian					11,92 (0,249)	11,93 (0,252)
sprawdzian ²					-0,41 (0,129)	-0,41 (0,128)
sprawdzian ³					-0,37 (0,081)	-0,37 (0,082)
pleć					1,65 (0,330)	1,411 (0,312)
dysleksja na egz. gimnazjalnym					1,73 (0,657)	1,74 (0,657)
dysleksja na sprawdzianie					-2,80 (0,790)	-2,80 (0,791)
test matryc Ravena (I pomiar)						2,21 (0,20)
liczba lat nauki lepiej wykształconego z rodziców						0,71 (0,08)
wyższa z pozycji rodziców na międzynarodowej skali prestiżu zawodów (htrel)						0,043 (0,018)
stała	100,66 (0,50)		100,80 (0,508)		101,01 (0,362)	90,12 (0,898)
poziom szkoły						
partycypacja decyzyjna			-1,747 (0,931)		-0,217 (0,502)	-0,089 (0,450)
partycypacyjny styl przywództwa			1,779 (1,360)		0,639 (0,854)	0,517 (0,834)
zaufanie do przełożonego			-0,224 (1,320)		0,199 (0,743)	0,236 (0,743)
zaufanie do współpracowników			1,574 (0,771)		0,14 (0,462)	0,20 (0,469)
identyfikacja z instytucją			-0,280 (0,976)		0,275 (0,642)	0,280 (0,649)
możliwość samorozwoju			1,562 (1,170)		-0,037 (0,586)	-0,273 (0,576)
informacja zwrotna			-2,507 (1,229)		-0,723 (0,738)	-0,576 (0,764)
oszacowanie efektów losowych						
wariancja efektów szkół	18,85		17,35		5,16	4,86
wariancja efektów oddziałów	17,82		17,60		2,33	1,67
wariancja na poziomie ucznia	173,97		173,98		77,21	72,00
współczynniki korelacji wewnątrzgrupowej						
poziom szkół	0,089		0,083		0,063	0,062
poziom klas	0,085		0,084		0,028	0,021
poziom szkół i klas łącznie	0,174		0,167		0,088	0,083
podsumowanie						
deviance	3854893		3854067		3451941	3416748
liczba szacowanych parametrów	4		11		10	20
AIC	38557		38563		34539	34549
BIC	38583		38634		34604	34659
liczba uczniów	4769		4769		4769	4769
liczba klas	291		291		291	291
liczba szkół	150		150		150	150

9.5.1.2. Wyniki dla części humanistycznej

W modelu, w którym brano pod uwagę wpływ zmiennych związanych z partycypacją i czynnikami jej towarzyszącymi, istotne okazały się efekty na skali zaufania do współpracowników oraz informacji zwrotnej. Pierwsza ze zmiennych niezależnych jest pozytywnie związana (efekt na poziomie 1,574), natomiast pozycja szkoły na skali „informacja zwrotna” jest ujemnie związana z wynikami egzaminacyjnymi (efekt na poziomie -2,507). W związku z tym, że skale opisujące style organizacyjne są wystandaryzowane, wartości uzyskanych efektów należy interpretować następująco: o ile punktów zmienia się wynik na egzaminie w części humanistycznej, jeśli wartości zmiennej niezależnej wzrosną o jedno odchylenie standardowe. Biorąc pod uwagę tę interpretację, musimy uznać efekty uzyskane dla dwóch istotnych zmiennych na poziomie szkoły za słabe.

W przypadku uwzględnienia w modelu zmiennych używanych do wyznaczania EWD (czyli wcześniejsze osiągnięcia szkolne, płeć i dysleksję), zmienne związane z zarządzaniem przestają być znacząco powiązane ze zmienną zależną. Nic nie zmienia dodanie do wcześniej omawianego modelu kontrolnych zmiennych z poziomu ucznia, mierzących pozycję społeczną oraz inteligencję ucznia.

9.5.2. Kultura organizacyjna

9.5.2.1. Wyniki dla części matematyczno-przyrodniczej

Uwzględnienie w analizach typów kultury organizacyjnej dało ciekawe rezultaty. W przypadku analiz bez uwzględnienia uprzednich osiągnięć, dwa typy kultury okazały się być znacząco statystycznie powiązane z wynikami egzaminacyjnymi. Są to klanowy typ kultury (efekt na poziomie 1,259) oraz rynkowy typ kultury (efekt na poziomie 3,26). Uwzględnienie w modelu zmiennych, których używa się przy wyliczaniu EWD, spowodowało zmniejszenie się siły efektów, ale cały czas są one istotne oraz pozytywne. Dla klanowego typu kultury wartość ta wynosi 0,8, natomiast dla rynkowego typu kultury jest to 1,34. Przy uwzględnieniu dodatkowego zestawu zmiennych z poziomu ucznia, mierzącego jego pozycję społeczną i inteligencję, typ klanowy oraz rynkowy nadal jest istotnie powiązany z osiągnięciami. Wielkość efektu w przypadku dwóch istotnych zmiennych była bardzo zbliżona do uzyskanych wcześniej przy modelu EWD.

Miary dopasowania wskazują na niewielką pozytywną zmianę ich wielkości w przypadku dodania do analiz zmiennych opisujących kulturę organizacyjną na poziomie szkoły. Natomiast w przypadku efektów losowych, widzimy zmniejszenie się wariancji na poziomie szkół po dodaniu tych zmiennych, co jest dla nas sygnałem, że rzeczywiście są one powiązane na poziomie szkoły ze zmienną zależną.

Tabela 3. Kultura organizacyjna szkoły a wyniki z egzaminu w części matematyczno-przyrodniczej. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części matematyczno-przyrodniczej		(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)	(s.e.)
oszacowanie efektów stałych											
poziom uczni											
sprawdzian						12,03	(0,265)	11,99	(0,268)	9,54	(0,291)
sprawdzian ²						1,26	(0,118)	1,24	(0,115)	1,21	(0,110)
sprawdzian ³						-0,42	(0,065)	-0,41	(0,064)	-0,388	(0,068)
płeć						-0,95	(0,284)	-0,93	(0,286)	-1,443	(0,260)
dysleksja na egz. gimnazjalnym						1,85	(0,809)	1,91	(0,809)	0,97	(0,753)
interakcja: dysleksja na sprawdzian i dysleksja na egz. gimnazjalnym						-4,07	(0,856)	-4,10	(0,854)	-2,91	(0,816)
test matryc Ravena (I pomiar)										3,79	(0,20)
liczba lat nauki lepiej wykształconego z rodziców										0,32	(0,08)
wyższa z pozycji rodziców na międzynarodowej skali prestiżu zawodów (HTREI)										0,049	(0,019)
stała		99,99	(0,52)	100,07	(0,487)	100,04	(0,338)	100,03	(0,334)	93,81	(0,922)
poziom szkoły											
klanowy typ kultury				1,259	(0,631)			0,808	(0,381)	0,897	(0,338)
adhokacyjny typ kultury				-0,035	(0,490)			-0,037	(0,273)	-0,121	(0,265)
rynkowy typ kultury				3,260	(0,903)			1,399	(0,453)	1,313	(0,417)
hierarchiczny typ kultury				-0,634	(0,416)			-0,39	(0,235)	-0,25	(0,229)
oszacowanie efektów losowych											
wariancja efektów szkół		20,86		13,50		4,20		2,92		3,06	
wariancja efektów oddziałów		21,35		21,34		5,22		5,26		3,53	
wariancja na poziomie ucznia		160,15		160,15		66,43		66,42		58,26	
współczynniki korelacji wewnątrzgrupowej											
poziom szkół		0,103		0,069		0,055		0,039		0,047	
poziom klas		0,106		0,109		0,069		0,071		0,054	
poziom szkół i klas łącznie		0,209		0,179		0,124		0,110		0,102	
podsumowanie											
deviance		37940,75		37907,24		33648,43		33625,08		32983,52	
liczba szacowanych parametrów		4		8		10		14		17	
AIC		37949		37923		33668		33653		33018	
BIC		37975		37975		33733		33744		33127	
liczba uczniów		4736		4736		4736		4736		4736	
liczba klas		289		289		289		289		289	
liczba szkół		149		149		149		149		149	

Tabela 4. Kultura organizacyjna szkoły a wyniki z egzaminu w części humanistycznej. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmieniana zależna: wynik egzaminu gimnazjalnego w części humanistycznej									
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)	(4)	(s.e.)	(5)
oszacowanie efektów stałych									
poziom uczni									
sprawdzian					11,92	(0,251)	11,90	(0,248)	10,05
sprawdzian ²					-0,40	(0,129)	-0,41	(0,130)	-0,45
sprawdzian ³					-0,37	(0,081)	-0,37	(0,082)	-0,346
pieć					1,60	(0,329)	1,61	(0,331)	1,367
dysleksja na egz. gimnazjalnym					1,71	(0,659)	1,76	(0,671)	1,01
dysleksja na sprawdzianie					-2,86	(0,791)	-2,89	(0,791)	-2,212
test matrycy Ravena (I pomiar)									2,21
liczba lat nauki lepiej wykształconego z rodziców									0,72
wyższa z pozycji rodziców na międzynarodowej skali prestiżu zawodów (HTREI)									0,041
stała	100,70	(0,50)	100,71	(0,488)	101,05	(0,362)	101,00	(0,355)	90,02
poziom szkoły									
klanowy typ kultury			1,296	(0,674)			0,895	(0,385)	1,036
adwokacyjny typ kultury			-0,263	(0,470)			-0,203	(0,260)	-0,332
rynkowy typ kultury			2,690	(0,884)			0,987	(0,400)	0,885
hierarchiczny typ kultury			-0,436	(0,447)			-0,34	(0,257)	-0,24
oszacowanie efektów losowych									
wariancja efektów szkół	18,86		14,49		5,40		4,62		4,25
wariancja efektów oddziałów	17,88		17,82		2,28		2,30		1,67
wariancja na poziomie ucznia	173,86		173,88		77,25		77,25		72,01
współczynniki korelacji wewnątrzgrupowej									
poziom szkół	0,090		0,070		0,064		0,055		0,054
poziom klas	0,085		0,086		0,027		0,027		0,021
poziom szkół i klas łącznie	0,174		0,157		0,090		0,082		0,076
podsumowanie									
deviance	38279,99		38258,87		34281,73		34268,81		33920,07
liczba szacowanych parametrów	4		8		10		14		17
AIC	38288		38275		34302		34297		33954
BIC	38314		38327		34366		34387		34064
liczba uczniów	4736		4736		4736		4736		4736
liczba klas	289		289		289		289		289
liczba szkół	149		149		149		149		149

Podsumowując, istnieje istotny efekt dla dwóch typów kultury organizacyjnej szkoły na wyniki egzaminacyjne. Są to klanowy i rynkowy typ kultury. Efekt dla obu tych zmiennych jest pozytywny, silniejszy dla rynkowego typu kultury, ale w skali bezwzględnej niezbyt mocny. Natomiast należy zauważyć, że efekt ten nieco słabnie w momencie dodania zmiennych używanych do modelowania EWD. Co ciekawe, spadek efektu jest wyraźnie obserwowany w przypadku rynkowego typu kultury (prawie o połowę w porównaniu do modelu bez zmiennych wykorzystywanych w modelowaniu EWD). W przypadku klanowego typu kultury, spadek efektu jest minimalny między modelami EWD a modelami zawierającymi tylko wynik na wyjściu.

9.5.2.2. Wyniki dla części humanistycznej

Analiza związków typów kultury organizacyjnej szkoły z wynikami uczniów w części humanistycznej daje rezultaty zbliżone do rezultatów dla egzaminu matematyczno-przyrodniczego. W przypadku modelu bez kontroli uprzednich osiągnięć, istotny pozytywny efekt obserwujemy tylko przy rynkowym typie kultury (efekt na poziomie 2,69). Jeśli natomiast do analizy dołączone zostaną zmienne potrzebne do szacowania EWD, pozytywne efekty widzimy w przypadku klanowego typ kultury (na poziomie 0,90) oraz rynkowego typ kultury (efekt na poziomie 0,99). Obie z tych zmiennych uzyskują istotne statystycznie efekty dla modelu, w którym brano pod uwagę zmienne określające pozycję społeczną ucznia i jego inteligencję. W przypadku typu hierarchicznego wynosi on 0,89, natomiast dla klanowego typu kultury jest to 1,03.

Miary dopasowania wskazują na pozytywne rezultaty dołączenia zmiennych opisujących kulturę organizacyjną do modeli wyjaśniających wyniki egzaminacyjne. Dodanie ich powoduje niewielki spadek miary *deviance*. Jednak należy zauważyć, że różnice są bardzo niewielkie. Z drugiej strony, oszacowanie efektów losowych na poziomie szkoły zmniejsza się na poziomie szkoły przy dodaniu analizowanych zmiennych. Świadczy to o tym, że istotnie tłumaczą one wariancję wyników uczniów, którą można powiązać z działaniem szkoły.

Podsumowując, możemy zauważyć istotny pozytywny efekt dwóch typów kultury organizacyjnej szkoły: hierarchicznej i rynkowej na wyniki uczniów w części humanistycznej egzaminu gimnazjalnego. Efekty dla typu rynkowego są istotne w przypadku wszystkich modeli. Hierarchiczny typ kultury jest istotny jedynie przy modelach analogicznych do EWD. Dodanie dodatkowych zmiennych z poziomu ucznia określających jego pozycję społeczną i inteligencję, nie zmienia istotności tych dwóch typów kultury i w bardzo niewielkim stopniu modyfikuje wielkość efektów.

9.6. Podsumowanie

Uzyskane rezultaty są niejednoznaczne i świadczą o zróżnicowanym wpływie różnych wymiarów stylów zarządzania szkołą na wyniki egzaminacyjne.

Partycypacja i zmienne jej towarzyszące (zaufanie do współpracowników, zaufanie do przełożonego, możliwość uzyskania informacji zwrotnej o swojej pracy, możliwość samorozwoju, identyfikacja z instytucją) nie mają wpływu na wyniki egzaminacyjne w przypadku części matematyczno-przyrodniczej. W przypadku części humanistycznej egzaminu w modelu, w którym analizowano tylko wynik egzaminu na zakończenie gimnazjum, uzyskano negatywny efekt na skali „informacja zwrotna” oraz pozytywny na skali „zaufanie do współpracowników”. Szczególnie interesujący w kontekście organizacji pracy szkoły jest negatywny wynik dla informacji zwrotnej. Należy zauważyć, że pytania na tej skali dotyczyły skali i sposobów przekazywania informacji zwrotnej. W żaden sposób natomiast nie odnoszą się one do jej jakości, która jeśli jest zła, może prowadzić do obserwowanego układu wyników. Należy jednak zauważyć, że zaproponowane tutaj wyjaśnienie tego wyniku jest czysto hipotetyczne – jego potwierdzenia należy szukać w następnych badaniach. Należy również zauważyć, że efekty na tych dwóch skalach stają się nieistotne w przypadku modelu analogicznego do EWD.

Drugim z analizowanych w rozdziale wymiarów funkcjonowania szkoły była kultura organizacyjna. W obu częściach egzaminu istotne wyniki uzyskano dla rynkowego typu kultury. Efekt dla tej skali w obu częściach egzaminu jest pozytywny. Widoczny jest jego spadek po dodaniu do analiz zmiennych używanych w analizie EWD (w obu częściach mniej więcej o połowę), ale cały czas efekt jest pozytywny. Przypomnijmy w tym miejscu charakterystykę tego typu kultury. Jest to nastawienie na rezultaty, na rywalizację z innymi szkołami i motywowanie pracowników do realizacji celów. Istotne jest także zorientowanie w działaniu na otoczenie zewnętrzne. Można zatem postawić hipotezę, że tego rodzaju działanie sprzyja wysokim wynikom egzaminacyjnym, jednak ten efekt staje się mniej znaczący w przypadku analiz uwzględniających wpływ wcześniejszych wyników uczniów (choć cały czas jest istotny).

Pozytywne efekty daje także klanowy typ kultury. W części matematyczno-przyrodniczej jego wpływ można zauważyć już na poziomie analiz w modelu bez uwzględnienia uprzednich osiągnięć, jednak w modelu analogicznym do EWD staje się on istotny w przypadku obu części egzaminu gimnazjalnego. Dla klanowego typu kultury ważne jest nastawienie na pracowników. Organizacja wykazująca ten typ kultury jest nastawiana na współpracę między pracownikami. Więzy między nimi są bardzo silne, mogą nabrać charakteru osobistego, prawie rodzinnego, a dyrektor często przyjmuje rolę mentora.

Rzeczywistość tego rodzaju instytucji odbywa się przez rozwój osobisty pracowników (nauczycieli). Wyznacznikiem sukcesu działania takiej organizacji są: lojalność, zaangażowanie, praca zespołowa oraz konsensus. Wydaje się zatem zasadne przyjąć, że tego rodzaju działanie daje pozytywne rezultaty, jeśli chodzi o efektywność nauczania w szkole mierzonej wskaźnikami EWD.

Przyglądając się wynikom otrzymanym w analizie różnych zmiennych związanych z zarządzaniem szkołą, można zauważyć interesujące tendencje. W przypadku egzaminu w części matematyczno-przyrodniczej wynikiom sprzyja nastawienie na współpracę z otoczeniem zewnętrznym i wykorzystywanie zasobów leżących poza szkołą. Świadczą o tym wyniki otrzymane dla rynkowego typu kultury. W przypadku części humanistycznej obserwujemy pozytywne efekty tych aspektów stylów zarządzania, które zakładają współpracę między nauczycielami, rozwój ich kompetencji. Negatywne efekty zdają się mieć style zarządzania, nastawione na przekazywanie dokładnych instrukcji działania i ciągły monitoring prowadzonych działań. Mogą o tym świadczyć efekty na skalach informacji zwrotnej – należy jednak zauważyć, że efekty te stają się nieistotne w modelu EWD. Następną z interesujących tendencji zaobserwowanych w wynikach jest pozytywny wpływ tych sposobów zarządzania, które w centrum mają nauczycieli jako pewną wspólnotę. Widoczne jest to zwłaszcza w spójnych wynikach otrzymanych w przypadku kultury organizacyjnej. W obu częściach egzaminu uzyskano pozytywne efekty dla klanowej kultury organizacyjnej. Zarówno w przypadku modeli bez kontroli uprzednich osiągnięć, jak i w modelach analogicznych do EWD, uzyskano podobne, pozytywne efekty. Charakterystyczne dla tego typu kultury są wartości pracy zespołowej, konsensusu, zaangażowania w funkcjonowanie organizacji.

Podsumowując, przedstawione w rozdziale analizy wskazują na związki stylów zarządzania ze wskaźnikami EWD. Niektóre z nich należy uznać za pożądane z punktu widzenia traktowania EWD jako miary efektywności pracy szkoły. Wykazano, że istnieje pozytywny efekt tych elementów stylów zarządzania, które nastawione są na pracowników, ich rozwój oraz utrzymywanie w szkole poczucia wspólnoty. Wyniki te są spójne dla części matematyczno-przyrodniczej oraz humanistycznej (choć w tym drugim przypadku tendencja jest bardziej wyraźna). Efekt ten jest widoczny w analizach uwzględniających jedynie surowe wyniki, jak i analogicznych do EWD.

Z drugiej strony, interesujący z punktu widzenia trafności samego wskaźnika EWD jest wpływ rynkowego typu kultury na wyniki i EWD. Ten typ kultury, charakteryzujący się dużą konkurencyjnością, jest relatywnie mocno związany z surowymi wynikami egzaminu, natomiast w przypadku analiz z modelem analogicznym do EWD, efekt staje się słabszy o połowę. Tego rodzaju rezultat można traktować jako pośredni dowód na trafność

wskaźnika EWD. Ponieważ rynkowy styl kultury jest nastawiony na konkurencyjność i osiągnięcie celów oraz ocenę na tle innych, jego związek z surowymi wynikami jest zgodny z przewidywaniami. Natomiast fakt, że kontrola wcześniejszych osiągnięć zmniejsza wpływ tego rodzaju kultury może oznaczać, że miara EWD bardziej pokazuje te aspekty pracy szkoły, które odnoszą się do relatywnych efektów pracy z uczniami.

Przedstawione wyniki analiz wskazują, że istnieją aspekty pracy szkoły związane ze stylami zarządzania, które mają wpływ na wyniki szkoły. Są one interesujące z dwóch powodów. Po pierwsze, wykazanie tego rodzaju związków jest istotne dla wskazania tych aspektów pracy szkoły, na które do pewnego stopnia szkoła jako instytucja ma wpływ. W przypadku analizowanych w rozdziale aspektów zarządzania trzeba pamiętać, że zmiany w tym zakresie wiążą się nie tylko ze zmianami w funkcjonowaniu placówek, ale także w sądach dotyczących rzeczywistości, które mają na temat danej organizacji jej członkowie. Z drugiej strony, analizy wskazują istnienie związków między danymi aspektami stylów zarządzania, a wskaźnikami EWD. Większość tych związków ma pożądany charakter z punktu widzenia trafności wskaźnika – te aspekty stylów zarządzania, które nastawione są na rozwój pracowników i premiąją nastawienie na więzi, między nimi dają pozytywne rezultaty w części humanistycznej. Natomiast te elementy stylów zarządzania, które są nastawione na rywalizację ze szkołami i osiągnięcie przewagi względem innych, mają o wiele słabsze efekty w przypadku modeli EWD niż w modelach z surowymi wynikami egzaminacyjnymi. Można zatem założyć, że zaprezentowane w rozdziale analizy wykazują na pożądane z punktu widzenia trafności metody EWD rezultaty.

Literatura

Augustyniak, E. (2009). *Kultura organizacyjna szkoły*. Kraków: Wyd. Naukowo-Dydaktyczne AGH.

Bauch, P. A. i Goldring, E. B. (1998). Parent – teacher participation in the context of school governance. *Peabody Journal of Education*, 73(1), 15–35.

Bush, T. (2003). *Theories of educational leadership and management* (3rd ed.). London; Thousand Oaks: Sage Publications.

Cameron, K. S. i Quinn, R. E. (2006). *Diagnosing and changing organizational structure: Based on the competing values framework*. San Francisco: Jossey-Bass.

Conley, S. (1991). Review of Research on Teacher Participation in School Decision Making. *Review of Research in Education*, 17(1), 225–266.

- Glew, D. J., O'Leary-Kelly, A. M., Griffin, R. W. i Van Fleet, D. D. (1995). Participation in Organizations: A Preview of the Issues and Proposed Framework for Future Analysis. *Journal of Management*, 21(3), 395–421.
- Hargreaves, D. H. (1995). School Culture, School Effectiveness and School Improvement. *School Effectiveness and School Improvement*, 6(1), 23–46
- Hartnell, C. A., Ou, A. Y. i Kinicki, A. (2011). Organizational culture and organizational effectiveness: a meta-analytic investigation of the competing values framework's theoretical suppositions. *The Journal of applied psychology*, 96(4), 677–694.
- Heller, F. A., Pusic, E., Strauss, G. i Wilpert, B. (1998). *Organizational participation: myth and reality*. Oxford; New York: Oxford University Press.
- Kołodziejczyk, J. (2011). *Partycypacja uczniów, rodziców i nauczycieli w zarządzaniu szkołą. [w:] Ewaluacja w nadzorze pedagogicznym: refleksje*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Marks, H. M. i Louis, K. S. (1997). Does Teacher Empowerment Affect the Classroom? The Implications of Teacher Empowerment for Instructional Practice and Student Academic Performance. *Educational Evaluation and Policy Analysis*, 19(3), 245–275.
- Maslowski, R. (2001). *School culture and school performance: an explorative study into the organizational culture of secondary schools and their effects*. Twente: Twente University Press.
- Mazurkiewicz, G. (2012). Zaangażowani przywódcy warunkiem istnienia przydatnej ewaluacji. [w:] *Przywódstwo i zmiana w edukacji. Ewaluacja jako mechanizm doskonalenia*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Miller, R. J. i Rowan, B. (2006). Effects of Organic Management on Student Achievement. *American Educational Research Journal*, 43(2), 219–253.
- Prokopowicz, P. i Żmuda, G. (2009). *Pełna partycypacja w zarządzaniu – nowy paradygmat w kierowaniu organizacją*. Warszawa: Centralny Ośrodek Doskonalenia Nauczycieli.
- Raudenbush, S. W., Rowan, B. i Kang, S. J. (1991). A Multilevel, Multivariate Model for Studying School Climate With Estimation Via the EM Algorithm and Application to U.S. High-School Data. *Journal of Educational and Behavioral Statistics*, 16(4), 295–330.
- 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.
- Schein, E. H. (2010). *Organizational culture and leadership*, fourth edition.

Smylie, M. A., Lazarus, V. i Brownlee-Conyers, J. (1996). Instructional Outcomes of School-Based Participative Decision Making. *Educational Evaluation and Policy Analysis*, 18(3), 181–198.

Somech, A. (2010). Participative Decision Making in Schools: A Mediating-Moderating Analytical Framework for Understanding School and Teacher Outcomes. *Educational Administration Quarterly*, 46(2), 174–209

Szymański, M. (2001). *Kryzys i zmiana. Studia nad przemianami edukacyjnymi w Polsce w latach dziewięćdziesiątych*. Kraków: Wydawnictwo Naukowe Akademii Pedagogicznej.

Tołwińska, B. (2011a). Kierowanie szkołą: rola dyrektora – partycypacja nauczycieli. [w:] *Przywództwo edukacyjne w szkole i jej otoczeniu*. Warszawa: Difin.

Tołwińska, B. (2011b). *Kompetencje społeczne dyrektorów szkół*. Kraków: Oficyna Wydawnicza Impuls.

Tsereteli, M., Martskvishvili, K. i Aptarashvili, I. (2011). The impact of public high school management system on academic achievement. *Problems of Education in the 21st Century*, 32, 121–128.

Zhu, C., Devos, G. i Li, Y. (2011). Teacher perceptions of school culture and their organizational commitment and well-being in a Chinese school. *Asia Pacific Education Review*, 12(2), 319–328.

Rozdział 10

Cechy nauczyciela

Celem rozdziału jest weryfikacja, czy wskaźniki edukacyjnej wartości dodanej dla gimnazjów korelują z takimi charakterystykami nauczycieli (np. wiek, staż pracy, wykształcenie, odbyte kursy doskonalenia zawodowego) i ich pracy (np. style nauczania), które potwierdzają trafność interpretacji EWD jako miary efektywności nauczania. Odpowiedzi na tak postawione pytanie badawcze, udzielono na podstawie wyników analizy danych, dotyczących nauczycieli matematyki i języka polskiego, którzy uczyli w klasach trzecich gimnazjum oraz danych dotyczących uczniów, w tym ich wyników z części matematycznej i polonistycznej egzaminu gimnazjalnego.

Cel rozdziału wpisuje się w obszar badań edukacyjnych, których współczesna formuła polega na testowaniu wpływu różnych czynników na osiągnięcia edukacyjne uczniów. Czynniki podlegające testowaniu rozważa się zazwyczaj (1) na poziomie indywidualnym (np. motywacja ucznia do nauki, status społeczno-ekonomiczny rodziny ucznia, płeć ucznia), (2) na poziomie oddziału klasowego (np. charakter grupy rówieśniczej, liczebność oddziału, styl pracy nauczyciela) i (3) na poziomie szkoły (np. lokalizacja szkoły, nakłady finansowe, infrastruktura). Osiągnięcia uczniów mierzone są najczęściej za pomocą ogólnokrajowych egzaminów zewnętrznych, realizowanych w formie standaryzowanych testów.

Weryfikacja trafności wskaźników EWD, uwzględniająca charakterystyki oddziału klasowego i szkoły, kieruje się inną logiką, niż ocena trafności uwzględniająca charakterystyki ucznia. Dodając do modelu EWD kolejne charakterystyki ucznia, można ocenić, czy wskaźnik EWD otrzymany z takiego modelu jest lepszy od stosowanego obecnie. Oceny dokonuje się, korelując ze sobą obecnie stosowany wskaźnik EWD z alternatywnym, uwzględniającym więcej czynników na poziomie ucznia.

Oczywiście dodawanie do modelu EWD zmiennych na poziomie oddziału i szkoły nie ma sensu, ponieważ wtedy model wyjaśniałby część wariancji na poziomie oddziału i szkoły, która jest miarą efektywności pracy nauczycieli i szkoły. Natomiast czynniki na poziomie oddziału i szkoły mogą służyć za kryteria trafności zewnętrznej wskaźnika EWD. Jeśli charakter związku tych zmiennych z wynikami uczniów w modelu EWD jest kierunkiem oczekiwanym, tzn. zgodnym z teorią, to można na tej podstawie orzekać o trafności EWD.

Wskaźniki EWD w wersji polskiej tworzone są z modeli dwupoziomowych, które uwzględniają charakterystyki ucznia i szkoły. Należy więc interpretować polskie wskaźniki EWD jako wskaźniki efektywności pracy szkół. Przez dodanie poziomu oddziału, tj. uwzględnienie faktu grupowania uczniów w oddziałach, można mówić o wskaźnikach efektywności pracy nauczycieli i szkół.

10.1. Wpływ nauczyciela na osiągnięcia szkolne

Przegląd literatury na temat wpływu nauczyciela na osiągnięcia szkolne uczniów, należy rozpocząć od przywołania pierwszego dużego badania nad rolą szkoły w formowaniu osiągnięć uczniów. Raport Jamesa Colemana (Coleman i in., 1966) kwestionował wpływ czynników szkolnych, w tym wpływ nauczyciela na procesy uczenia się. Wskazywał natomiast na dużą rolę środowiska rodzinnego i rówieśniczego. Obecnie wiadomo, że hipotezy zespołu Colemana są nieprawdziwe. Zostały bowiem postawione przy wykorzystaniu problematycznej metodologii, która nie uwzględniała hierarchicznej natury danych edukacyjnych. Jednak raport sprowokował szeroką debatę wśród badaczy i polityków. Zaowocowała ona licznymi badaniami przynoszącymi nowe dowody w kwestii udziału czynników szkolnych w wyjaśnianiu zróżnicowania wyników uczniów. Badacze ogniskowali swoje wysiłki na odpowiedzi na trzy pytania: (1) Czy szkoły i nauczyciele istotnie różnią się w swoich możliwościach podnoszenia osiągnięć uczniów? (2) Jeśli tak, to w jakim stopniu zróżnicowanie efektywności nauczycieli (jakości stylu pracy) determinuje wyniki uczniów? (3) Czy zróżnicowanie efektywności nauczycieli da się wyjaśnić obserwowalnymi charakterystykami nauczycieli (np. odbyte szkolenia zawodowe, staż pracy, znajomość nauczanego przedmiotu, itp.) lub charakterystykami ich pracy (np. style nauczania, stymulowanie pracy w grupach, stosowanie oceniania kształtującego itp.)?

Wyróżnić można dwa podejścia analityczne do szacowania efektu nauczyciela. Pierwsze polega na analizie wariancji międzyoddziałowej wyników uczniów. Przyjmuje się w nim założenie, że cała wariancja międzyoddziałowa w wynikach uczniów może zostać przypisana efektywności nauczyciela. Nie chodzi natomiast o ustalenie, która konkretna cecha nauczyciela czy jego stylu pracy, wyjaśnia wariancję międzyoddziałową. Drugie podejście polega na modelowaniu funkcji produkcyjnej szkół. Analiza ma na celu ustalenie związku między konkretną cechą nauczyciela a wynikami uczniów. Za pomocą równań regresji szacuje się efekt konkretnej cechy nauczyciela lub jego pracy na zmienną zależną, którą są wyniki uczniów.

10.2. Dowody z analiz wariancji międzyoddziałowej

Barbara Nye, Spyros Konstantopoulos i Larry Hedges (2004) podsumowali wyniki 15 oszacowań wariancji międzyoddziałowej. Były one raportowane w pięciu badaniach (Armour, 1976; Goldhaber i Brewer, 1997; Hanushek, 1971; 1992; Murnane i Phillips, 1981). Wyniki sugerują, że od 7% do 21% wariancji osiągnięć szkolnych uczniów wyjaśnia różna efektywność nauczycieli.

Brian Rowan, Richard Correnti i Robert Miller (2002), raportowali od 18% do 28% (matematyka) oraz od 12% do 23% (czytanie) wyjaśnionej wariancji międzyoddziałowej w modelu

pustym z losową stałą oraz odpowiednio między 8% i 18% oraz między 4% i 16% dla modelu uwzględniającego wcześniejsze osiągnięcia ucznia, status społeczno-ekonomiczny, skład społeczny szkoły.

Nye, Konstantopoulos i Hedges (2004) analizowali dane z badania z losowym przypisaniem uczniów szkół podstawowych i nauczycieli do porównywanych grup. Dane pochodziły z rozpoczętego w 1985 roku 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, przynależność do grupy etnicznej; na poziomie oddziału: liczebność oddziału, 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 z losową stałą odpowiednio: 14% i 9%.

Jeszcze silniejszych dowodów co do rozmiaru efektu nauczyciela dostarczają badania uwzględniające czynnik genetyczny. Są to badania na bliźniętach. Czynnik genetyczny wyjaśnia do 65% osiągnięć edukacyjnych (Oliver i in., 2004). Uwzględniając czynnik genetyczny, efekt nauczyciela szacuje się na 8% (Byrne i in., 2010). Jak wykazała Jeanette Taylor i in. (2010), jakość pracy nauczyciela moderuje efekt genetycznego potencjału poznawczego ucznia.

Powyższe dowody są spójne i pozwalają założyć, że zróżnicowanie efektywności pracy nauczycieli może wyjaśnić około 8% do 12% zróżnicowania osiągnięć szkolnych uczniów. Przy czym efekt ten jest silniejszy w nauczaniu matematyki niż czytania. Należy zaznaczyć, że efektywny nauczyciel ma większy wpływ na proces uczenia się, niż jakiegokolwiek inne czynniki kontrolowane przez system szkolny (Rivkin, Hanushek i Kain, 2005; Wright, Horn i Sanders, 1997). Stąd też nauczyciel powinien być centralnym punktem zainteresowania decydentów polityki oświatowej. Jeśli bowiem chcemy naprawiać coś w edukacji, to najważniejszym tematem powinny stać się umiejętności nauczycielskie (Akiba, LeTendre i Scribner, 2007; Hanushek, 1997).

10.3. Dowody z modelowania funkcji produkcyjnej szkół

Nie ma wątpliwości co do tego, że nauczyciele, różniąc się efektywnością (jakością pracy), wpływają na wyniki uczniów. Nie jest natomiast jasne, jakie konkretne cechy nauczyciela i jego stylu pracy, pozwalają wyjaśnić zróżnicowanie w efektywności nauczycieli 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) stwierdzają, że mała liczba empirycznych dowodów 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 jakości [efektywności] nauczycieli.

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 Hanush'em (1986; 1989; 1996; 1997) a Robem Greenwaldem, Larrym Hedgesem i Richardem Laine'em (1994a; 1994b; 1996a). Hanushek stoi na stanowisku, że metaanaliza dostępnych badań wskazuje na brak silnego i pozytywnego związku charakterystyk nauczycieli z wynikami uczniów, przynajmniej gdy kontrolowany jest status społeczno-ekonomiczny rodziny ucznia. Hedges, Laine i Greenwald uważają, że analiza tych samych danych, prowadzi do wniosku o istnieniu pozytywnego związku między charakterystykami nauczycieli a wynikami uczniów, a jego siła jest na tyle duża, że ma praktyczne znaczenie dla twórców polityk oświatowych. Debata ta nie znalazła swojego ostatecznego rozstrzygnięcia. Dowody przywoływane przez każdą ze stron są przekonujące. Wydaje się, że praktycznym wnioskiem z tej debaty, jest stwierdzenie, że dominująca w badaniach tego typu metodologia (głównie ekonometryczne analizy danych populacyjnych lub danych z reprezentatywnych badań ilościowych) i narzędzia pomiaru charakterystyk nauczycieli (deklaracje zbierane w kwestionariuszach) nie są wystarczająco dokładne. Wiele bowiem indywidualnych, rodzinnych, rówieśniczych, sąsiedzkich, szkolnych czynników pozostaje niemierzalnych lub uwikłanych jest w skomplikowane zależności. Mimo braku zgody co do znaczenia poszczególnych charakterystyk nauczyciela i jego pracy w wyjaśnianiu zmienności osiągnięć szkolnych, należy przytoczyć dowody empiryczne na rzecz różnych stanowisk.

Brian Byrne (1983) podsumował wyniki 30 badań na temat związku poziomu wiedzy przedmiotowej nauczyciela z wynikami uczniów. Siedemnaście badań wykazało związek pozytywny a trzynaście brak związku. Jednakże w badaniach relacjonujących brak związku, Byrne odnotował 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 pięć, które orzekały o pozytywnym związku wiedzy przedmiotowej z efektywnością nauczycieli, mierzoną jako średnie wyniki uczniów. Wyniki nauczycieli na testach językowych (*verbal ability*), oceniających umiejętności czytania, wiedzę gramatyczną, zakres słownictwa, okazują się słabo korelować z wynikami testów językowych (Bowles i Levin, 1968; Hanushek, 1971; 1992; Ehrenberg i Brewer, 1995; Ferguson i Ladd, 1996). Silniejsze związki prezentują wyniki egzaminów przedmiotowych, którym poddawani są nauczyciele, a wynikami uczniów w zakresie matematyki (Rowan, Chiang i Miller 1997; Ferguson i Ladd 1996). 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 uczniów. Na podstawie objętych analizą publikacji, autorzy wykazali wpływ posiadania przez nauczyciela wyższego wykształcenia na wyniki testów kompetencji matematycznych. Nie ma pewnych dowodów tego związku w przedmiotach przyrodniczych, humanistycznych i językach obcych. Badanie Lindy Darling-Hammond (2000) dowodzi 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, takie jak: stosunek liczby uczniów do nauczycieli (*student-teacher ratio*), liczebność oddziału, proporcja nauczycieli w personelu szkoły, w analizach Darling-Hammond pokazują pozytywny, jednak bardzo mały lub nieistotny związek z wynikami uczniów. Daniel Goldhaber i Dominic Brewer (1997; 2000), a także Rowan, Chiang i Miller (1997) wykazali brak związku wyników uczniów i poziomu wykształcenia nauczycieli. Autorzy raportowali jednak 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.

Carolyn Evertson, Willis Hawley i Marilyn Zlotnik (1985) oraz Byrne (1983) wykazali stały pozytywny efekt formalnego wykształcenia pedagogicznego nauczycieli na wyniki uczniów. Podobnie i tu występuje problem z małą wariancją, co uniemożliwia zaobserwowanie istotnego efektu. Spośród trzynastu badań, jedenaście raportowało wyższą efektywność nauczycieli z pełnymi uprawnieniami zawodowymi, nad nauczycielami bez dyplomu lub posiadającymi niepełne uprawnienia zawodowe. Umiejętności dydaktyczne są powiązane z wiedzą przedmiotową i łącznie korelują z wynikami uczniów. Wielu autorów wykazało także, że lepsze wyniki osiągają uczniowie nauczycieli, którzy uczestniczą w programach doskonalenia zawodowego z zakresu dydaktyki przedmiotowej (Brown, Smith i Stein, 1996, Wiley i Yoon, 1995).

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 trzech 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 spada u nauczycieli zbliżających się do wieku emerytalnego. Nye, Konstantopoulos i Hedges (2004) oszacowali liniowy związek stażu nauczania z wynikami uczniów na 0,085 (matematyka) oraz 0,094 (kompetencje językowe) odchylenia standardowego wyników uczniów.

Dowiedziano istotnego związku klimatu nauczania z wynikami uczniów (Anderson, 1982; Freiberg, 1999). Klimat nauczania rozważany jest jako charakterystyka stylu pracy nauczycieli. Valerie Lee i Anthony Bryk (1989) wykazali, że klimat nauczania ukierunkowany na osiągnięcia, jest istotnym czynnikiem wyjaśniającym zróżnicowanie wyników uczniów. Im przywództwo nauczycieli bardziej nastawione jest na korygowanie zachowań niepożądanych oraz wskazywanie rozwiązań poprawnych, efektywność nauczania okazuje się być wyższa (Brown, 2001). Z kolei Copper i Mullen (1994) zwrócili uwagę na spójność grupy nauczycieli, jako na czynnik pozytywnie oddziałujący na osiągnięcia szkolne uczniów. Ronald Edmonds (1979) obok bezpiecznego i stabilnego klimatu szkoły wskazał na pozytywny związek z wynikami uczniów takich czynników jak silne przywództwo dyrektora, zgodność co do celów i metod nauczania wśród kadry nauczycielskiej, wysokie wymagania wobec uczniów, używanie 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 nauczanie grupowe, efektywne strategie nauczania, efektywne zarządzanie klasą, stosowanie wzmocnień pozytywnych uczniów, i zaangażowanie się rodziców w życie szkoły.

W metaanalizie ponad 100 studiów eksperymentalnych nad pracą nauczyciela, Robert Marzano, Jane Marzano, i Debra Pickering (2003) wskazują, że istotne dla osiągnięcia sukcesu edukacyjnego mogą być częste podsumowania i nakłanianie uczniów do robienia notatek, stosowanie wzmocnień pozytywnych, wyrażanie oznak uznania uczniom, gdy na to zasługują, zadawanie prac domowych i praca indywidualna na lekcjach, wspólne rozwiązywanie problemów, stawianie dobrze zdefiniowanych celów i zapewnienie wsparcia w ich realizacji, generowanie hipotez i testowanie ich, aktywowanie u uczniów posiadanej już wiedzy.

Hattie (2009) podsumował wyniki 31 metaanaliz dotyczących wpływu nauczyciela na wyniki uczniów. Tabela 1 prezentuje wyniki jego pracy. Oprócz metaanalizy traktującej o efekcie nauczyciela w ogóle, zaprezentowano zbiorcze wyniki metaanaliz, poświęconych konkretnym charakterystykom nauczycieli i ich pracy. Hattie do raportowania efektów wykorzystał miarę d (Cohen, 1988). W badaniach edukacyjnych efekt równy 0,2 należy interpretować jako słaby, równy 0,4 jako umiarkowany, a równy 0,6 jako silny. Do największych przyrostów wyników uczniów przyczyniają się nauczyciele, którzy brali udział w szkoleniach z zakresu prowadzenia lekcji, wykorzystujących dokładne wideoanalizy. Wyniki uczniów podnoszą nauczyciele, którzy formułują zrozumiałe i precyzyjne komunikaty. Zdolności językowe nauczycieli pozwalają im na komunikowanie idei w przejrzysty i przekonujący sposób (Murnane, 1985). Efektywni są także ci nauczyciele, którzy mają dobre relacje ze swoimi uczniami. Spośród zmiennych,

które mierzą relacje nauczycieli i uczniów, największy wpływ ($d > 0,4$) na wyniki uczniów mają: promowanie samodzielności, empatia, przyjazność, zachęcanie do myślenia analitycznego, krytycznej oceny i myślenia kreatywnego, zachęcanie do nauki, indywidualizacja nauczania.

Tabela 1. Podsumowanie informacji z metaanaliz na temat efektu nauczyciela na osiągnięcia edukacyjne uczniów (Hattie, 2009, s. 109)

efekt	liczba metaanaliz	I. badań	I. oszacowań efektów	d	(s.e.)
pogłębione szkolenia nauczycieli (wideoanalizy)	4	402	439	0,88	b.d.
klarowność komunikacji	1	b.d.	b.d.	0,75	b.d.
relacje nauczyciela z uczniami	1	229	1450	0,72	(0,011)
doskonalenie zawodowe	5	537	1884	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)
łącznie	31	2225	5559	0,49	0,049

Podsumowując przytoczone wyniki badań, wskazać można na pozytywny wpływ na wyniki uczniów szkoleń z zakresu prowadzenia lekcji, formułowania przez nauczyciela zrozumiałych i precyzyjnych komunikatów do uczniów, wiedzy przedmiotowej nauczyciela z nauczanego przedmiotu – zwłaszcza w przypadku matematyki, formalnego wykształcenia pedagogicznego, stażu – najbardziej efektywni nauczyciele, to ci z ponad pięcioletnim stażem, jednak efektywność spada wraz ze zbliżaniem się do wieku emerytalnego, spójności grupy nauczycieli, zgodności co do celów i metod nauczania wśród kadry nauczycielskiej, wysokich wymagań wobec uczniów, wykorzystywania wyników uczniów do ewaluacji pracy.

10.4. Przygotowanie danych do analizy

Przed przejściem do właściwych analiz, dane zostały odpowiednio przygotowane. Usunięto dane dotyczące nauczycieli, którzy wzięli udział tylko w I etapie badania i nie powtórzyli swojego udziału w III. etapie ($n = 949$). W próbie pozostało $n = 2630$ nauczycieli, w tym $n = 1890$ nauczycieli, którzy brali udział zarówno w I i III etapie. W przypadku braku

danych metryczkowych zebranych w III etapie, tam gdzie było to możliwe, uzupełniono część brakujących danych za pomocą informacji zebranych podczas I etapu.

Następnym krokiem przygotowania danych do analiz było połączenie nauczycieli z oddziałami, które uczyli. Przypisanie takie zrealizowano dla nauczycieli języka polskiego i matematyki. Analizy prowadzono na dwóch zbiorach danych. Jeden z nich dotyczył kompetencji matematycznych, gdzie zmienną zależną były wyniki uczniów z części matematycznej egzaminu gimnazjalnego. Drugi dotyczył kompetencji językowych, gdzie zmienną zależną były wyniki uczniów z części polonistycznej egzaminu. Należy to podkreślić, ponieważ w pozostałych rozdziałach tego raportu analizy prowadzone były na zmiennych zależnych, dotyczących wyników z części matematyczno-przyrodniczej i części humanistycznej. W tym rozdziale zrezygnowano z tego typu analiz, ponieważ niemożliwe jest uchwycenie wpływu na zmienną zależną charakterystyk konkretnych nauczycieli przypisanych do oddziałów. W przypadku przedmiotów matematyczno-przyrodniczych i humanistycznych konkretne oddziały były bowiem w zdecydowanej większości nauczane przez kilku nauczycieli.

Do uczniów z oddziałów biorących udział w III etapie badania, przypisano nauczycieli, którzy uczyli te oddziały w ciągu ostatniego roku nauki w gimnazjum. W przypadku języka polskiego 260 (96%) oddziałów miało przez ten okres jednego nauczyciela. W przypadku matematyki 202 (94%) oddziałów. W tabeli 2 zaprezentowano dla porównania także dane dla przedmiotów z bloku przyrodniczego i humanistycznego. Tak więc analizy prowadzone z uwzględnieniem jako zmiennej zależnej wyników egzaminu tylko z części matematycznej i tylko polonistycznej, są najbardziej trafne w szacowaniu efektu nauczyciela.

Tabela 2. Liczba nauczycieli przypadająca na jeden oddział

liczba nauczycieli	jęż. pol.	mat.	przyrod.	human.
1	260	202	7	63
2	11	12	17	150
3		1	67	70
4			109	4
5			80	1
6			11	
7			0	
8			1	

Gdy nauczycieli danego przedmiotu było więcej niż jeden do jednego oddziału, zastosowano agregowanie do średnich charakterystyk nauczycieli. Stosowano zaokrąglenia do jedności w przypadku agregacji zmiennych kategorialnych, tj. płci, poziomu

wykształcenia, stopnia awansu zawodowego. Dla wartości 0,5 stosowano zaokrąglenia w górę. Ze względu na braki danych ostatecznie w analizach dotyczących nauczania języka polskiego uwzględniono 223 oddziałów, a w analizach dotyczących nauczania matematyki 214 oddziałów.

10.5. Style nauczania i przebiegu lekcji

Do pomiaru stylów nauczania i przebiegu lekcji zastosowano skale zbudowane na podstawie baterii pytań zarówno z kwestionariuszy kierowanych do uczniów, jak i nauczycieli. W ten sposób można było porównać pomiar tych samych konstruktów z różnych perspektyw. Dokumentacja techniczna skal znajduje się w aneksie.

Opracowanie ostatecznych wersji skal poprzedzono eksploracyjną analizą czynnikową (EFA). Celem analizy była weryfikacja, czy na podstawie zebranych danych zostały odtworzone modele czynnikowe, które skonceptualizowano na etapie tworzenia kwestionariuszy. W większości przypadków przygotowane baterie pytań do mierzenia poszczególnych konstruktów tworzyły jednowymiarowe skale o wysokiej rzetelności. Ostateczne wartości (*score*) zostały obliczone jako wynik confirmacyjnej analizy czynnikowej (CFA). Miary dopasowania skal zostały zaprezentowane w aneksie.

10.6. Selekcja zmiennych niezależnych

Model pusty z losową stałą, w którym zmienną zależną jest wynik egzaminu gimnazjalnego, uwzględnia jedynie fakt grupowania uczniów w oddziałach i szkołach. Fakt grupowania uczniów w oddziałach wyjaśnia 14% wariancji wyników uczniów z egzaminu z części matematycznej i 12% z części polonistycznej. Fakt grupowania uczniów w szkołach wyjaśnia odpowiednio 7% i 10%.

Do modelu pustego dodawano jedna po drugiej zmienne charakteryzujące nauczycieli i ich pracę. W ten sposób oceniono, które zmienne są istotne w wyjaśnianiu różnicowania wyników egzaminów, a które nie. Etap ten służył wyselekcjonowaniu zmiennych do dalszych analiz, które rokowały jako istotne predyktory osiągnięć uczniów. Testowano 26 zmiennych niezależnych, w tym 14 charakterystyk demograficznych, związanych ze stażem pracy, stopniem awansu zawodowego, doskonaleniem zawodowym nauczycieli. Wśród testowanych zmiennych znalazło się siedem skal, skonstruowanych na podstawie pytań z kwestionariuszy nauczycielskich oraz pięć skal, skonstruowanych na podstawie pytań z kwestionariuszy uczniowskich.

Wykorzystywane w badaniu narzędzia zostały opracowane 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.

Lista zmiennych prezentowana jest w tabelach 3, 4 i 5. Podsumowano nich wyniki eksploracyjnych analiz, mających na celu wyłonienie do dalszych analiz tych zmiennych, które mają potencjał w wyjaśnianiu zmienności osiągnięć szkolnych uczniów. Podano współczynniki regresji i błędy standardowe, gdy do modelu pustego dodano konkretną zmienną. Zawsze w modelu obecny był tylko jeden predyktor, wprowadzony na poziom oddziałów klasowych. Istotne predyktory, które dodatnio korelują ze zmienną zależną oznaczono „+”, a ujemnie „-”. Nieistotne predyktory oznaczono kropką. Przyjęto poziom istotności $p \sim 0,05$. Nie są raportowane wartości statystyki p . Zainteresowany czytelnik może je obliczyć samodzielnie, dzieląc współczynnik regresji przez błąd standardowy, a następnie odczytując wartość krytyczną z rozkładu normalnego -Z lub wykorzystać dostępne kalkulatory. Podano także właściwe interpretacje skal wraz z oczekiwanym kierunkiem korelacji. W celu ułatwienia interpretacji skale aktywizujący styl nauczania, pomocny nauczyciel, wymagający nauczyciel, wymagający nauczyciel (dla mat. i pol.), pomocny nauczyciel (dla mat. i pol.), koncentracja na wymaganiach (dla mat. i pol.), aktywizujący styl nauczania (dla mat. i pol.) zostały odwrócone, tzn. wartości skal zostały pomnożone przez -1. Dokumentacja techniczna tworzenia skal zamieszczona jest w aneksie. Dokumentacja raportuje skale oryginalne, czyli takie które zachowują kierunek kodowania wartości zmiennych w kwestionariuszach.

Płeć jako samodzielny czynnik okazała się dodatnio korelować z wynikami uczniów z części polonistycznej egzaminu. W nauczaniu języka polskiego nauczycielki przyczyniają się bardziej do podniesienia wyników uczniów niż nauczyciele. Nie należy przypisywać jednak dużej wagi do tego wniosku, ponieważ odsetek mężczyzn nauczycieli w próbie jest bardzo niski (2,5%), co wpływa na niestabilność tego efektu.

Korelacja roku urodzenia nauczyciela i stażu pracy jest wysoka i wynosi 0,938. Zgodnie z oczekiwaniami (przynajmniej jeśli chodzi o nauczanie matematyki) zaobserwowano nieistotny, jednak dodatni liniowy związek stażu pracy z wynikami uczniów (Murnane i Phillips, 1981; Klitgaard i Hall, 1975). Nie testowano związku nieliniowego. Uzyskane oszacowania efektu stażu (0,005 odchylenia standardowego. w przypadku matematyki i -0,001 w przypadku języka polskiego), są jednak znacznie niższe, niż uzyskane przez Nye, Konstantopoulou i Hedgesa (2004), odpowiednio: 0,085 (matematyka) oraz 0,094 (kompetencje językowe) odchylenia standardowego wyników uczniów.

Odnotowano pozytywny związek wykształcenia z wynikami uczniów w nauczaniu matematyki. Jest to zgodne z dowodami przedstawionymi przez Nye, Konstantopoulou i Hedgessa (2004) oraz Wayne'a i Youngsa (2003).

Tabela 3. Zestawienie zmiennych istotnych i nieistotnych w wyjaśnianiu wyników uczniów – charakterystyki nauczycieli. Zmienne testowane były pojedynczo w trzypoziomowym modelu regresji z losową stałą

zmienna	matematyka	współczynniki; (s.e.)	jęz. polski	współczynniki; (s.e.)
pleć nauczyciela	.	1,76 (1,25)	+	3,13 (1,59)
rok urodzenia nauczyciela	.	-0,09 (0,06)	.	-0,05 (0,07)
staż pracy	.	0,07 (0,06)	.	-0,01 (0,06)
wykształcenie	+	6,13 (3,05)	.	-2,04 (3,13)
stopień awansu zawodowego	.	1,43 (0,80)	.	0,48 (0,74)
brał(a) udział w szkoleniu z analizy wyników egzaminów zewnętrznych	.	0,21 (1,90)	.	-2,11 (1,92)
(...) metody edukacyjnej wartości dodanej	.	-0,80 (1,53)	.	-2,05 (1,46)
(...) szkoleniu z ewaluacji wewnętrznej	.	-0,83 (1,88)	.	0,20 (1,36)
brał(a) udział w ciągu trzech ostatnich lat w studiach podyplomowych	.	-3,89 (1,63)	.	-0,40 (1,11)
(...) w kursach kwalifikacyjnych	.	-1,23 (1,27)	.	-0,65 (1,16)
(...) w kursach doskonalących	.	-0,02 (1,40)	.	0,59 (1,15)
(...) w szkoleniowych radach pedagogicznych	-	-2,11 (0,62)	.	-0,67 (0,47)
(...) w konferencjach metodycznych	.	0,47 (1,36)	-	-2,95 (1,53)
indeks doskonalenia zawodowego	.	-0,29 (0,99)	.	-0,90 (0,76)

., nieistotny; +, pozytywny; -, negatywny

Specyficzne dla kontekstu polskiego zmienne, takie jak stopień awansu zawodowego oraz różne formy doskonalenia zawodowego okazały się być nieistotnymi predyktorami wyników uczniów. Odnotowano jedynie niewielką ujemną korelację uczestnictwa w szkoleniowych radach pedagogicznych i w konferencjach metodycznych z wynikami uczniów

Kolejne dwie tabele (4 i 5) prezentują wyniki wstępnych analiz związków z wynikami uczniów skal opisujących style nauczania. Skale konstruowane były osobno na podstawie odpowiedzi uczniów na pytania tworzące skale i osobno na podstawie informacji uzyskanych od nauczycieli.

Tabela 4. Zestawienie zmiennych istotnych i nieistotnych w wyjaśnianiu wyników uczniów – skale na bazie informacji od nauczycieli. Zmienne testowane były pojedynczo w trzypoziomowym modelu regresji z losową stałą

skala	matematyka	współczynniki; (s.e.)	j. polski	współczynniki; (s.e.)	interpretacja skali i oczekiwany kierunek związku ze zmianą zależną
wsparcie w ramach szkoły	.	-0,70 (0,51)	.	-0,74 (0,49)	Im wyższa pozycja na skali, tym bardziej nauczyciel ma rozbudowane sieci wsparcia. Oczek. zw. dodatni.
wsparcie spoza szkoły	.	-0,09 (0,50)	-	-1,34 (0,51)	Im wyższa pozycja na skali, tym bardziej nauczyciel ma rozbudowane sieci wsparcia. Oczek. zw. dodatni.
otwarty styl prowadzenia zajęć	.	2,43 (1,38)	+	3,84 (1,38)	Im wyższa wartość na skali, tym bardziej otwarty styl nauczania. Oczek. zw. dodatni.
utrzymanie dyscypliny	+	0,98 (0,52)	.	0,26 (0,52)	Im wyższa wartość na skali, tym większa dyscyplina panuje na lekcji. Oczek. zw. dodatni.
aktywizujący styl nauczania	.	-1,00 (0,62)	.	0,75 (0,55)	Im wyższa wartość na skali, tym bardziej aktywizujący styl nauczania. Oczek. zw. dodatni.
pomocny nauczyciel	+	0,98 (0,50)	.	0,48 (0,44)	Im wyższa wartość na skali, tym bardziej otwarty i nastawiony na aktywizowanie uczniów styl nauczania. Oczek. zw. dodatni.
wymagający nauczyciel	.	-0,10 (0,61)	.	0,69 (0,57)	Im wyższe wartości na skali, tym bardziej wymagający nauczyciel. Oczek. zw. dodatni.

„.” nieistotny; „+” pozytywny; „-” negatywny

Skale, które okazały się istotne i zachowują się zgodnie z oczekiwanym kierunkiem związku to skale otwarty styl prowadzenia zajęć, utrzymanie dyscypliny, pomocny nauczyciel, autorytet nauczyciela/utrzymanie dyscypliny (dla mat. i pol.), pomocny nauczyciel (dla mat. i pol.), koncentracja na wymaganiach (dla mat. i pol.).

Z przeprowadzonych analiz warto sformułować praktyczny wniosek, do wykorzystania w przyszłych badaniach szkolnych uwarunkowań efektywności kształcenia. Mianowicie, mierzenie stylów pracy nauczyciela, powinno odbywać się przez pytanie uczniów, a nie samych nauczycieli. Skale opracowane na podstawie odpowiedzi uczniów są bardziej rzetelne, niż te konstruowane na podstawie odpowiedzi nauczycieli. Mimo zapewnień o anonimowości badania, nauczyciele mogą czuć się zagrożeni rozliczaniem efektów swojej pracy na podstawie wyników badania lub mają problem z krytyczną oceną swojej pracy. Stąd uzyskiwane przez nich odpowiedzi, nie generują stabilnych wskaźników stylów pracy i atmosfery na lekcjach.

Tabela 5. Zestawienie zmiennych istotnych i nieistotnych w wyjaśnianiu wyników uczniów – skale na bazie informacji od uczniów. Zmienne testowane były pojedynczo w trzypoziomowym modelu regresji z losową stałą

skala	matematyka	współczynniki: (s.e.)	j. polski	współczynniki: (s.e.)	interpretacja skali i oczekiwany kierunek związku ze zmianą zależną
autorytet nauczyciela / utrzymanie dyscypliny (dla mat. i pol.)	+	2,81 (0,82)	+	4,30 (0,77)	Im wyższa wartość na skali, tym większa dyscyplina panuje na lekcji. Oczek. zw. dodatni.
wymagający nauczyciel (dla mat. i pol.)	.	-3,39 (2,33)	-	-2,78 (1,50)	Im wyższe wartości na skali, tym bardziej wymagający nauczyciel. Oczek. zw. dodatni.
pomocny nauczyciel (dla mat. i pol.)	+	2,01 (0,99)	+	2,95 (0,97)	Im wyższe wartości na skali, tym bardziej pomocny nauczyciel. Oczek. zw. dodatni.
koncentracja na wymaganiach (dla mat. i pol.)	+	1,82 (1,19)	+	3,15 (1,12)	Im wyższe wartości na skali, tym bardziej intensywne uczenie pod kątem egzaminu. Oczek. zw. dodatni.
aktywizujący styl nauczania (dla mat. i pol.)	-	-2,54 (1,32)	.	-1,58 (1,33)	Im wyższa wartość skali, tym bardziej aktywizujący styl nauczania. Oczek. zw. dodatni.

., nieistotny; +, pozytywny; -, negatywny

10.7. Wyniki dla części matematycznej

Tabela 6 prezentuje współczynniki i błędy standardowe dla modelu pustego (1M) z uwzględnieniem losowej stałej, oraz modeli do których dokładano kolejno predyktory na poziomie uczniów i oddziałów. Na podstawie stałej w modelu pustym (1M) możemy przewidzieć wynik ucznia z części matematycznej egzaminu. Uczeń średnio otrzymał 101,49 punktów. Wyniki surowe ze sprawdzianu i części matematycznej egzaminu gimnazjalnego zostały wyskalowane za pomocą modelu 3PL. Wyniki uczniów raportowane są na skali o średniej w populacji 100 i odchyleniu standardowym 15. W analizowanej próbie odpowiednio 103 i 15,17. Średnia w próbie różni się od stałej w modelu ze względu na braki danych.

Tabela 6. Uwarunkowania osiągnięć matematycznych oraz charakterystyki nauczycieli i ich pracy. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych.

zmieniana zależna: wynik egzaminu gimnazjalnego w części matematycznej														
	(1M)	(s.e.)	(1MN)	(s.e.)	(2M)	(s.e.)	(2MN)	(s.e.)	(3M)	(s.e.)	(3MN)	(s.e.)	(3MKN)	(s.e.)
oszacowanie efektów stałych														
poziom uczenia														
sprawdzian					12,47	(0,341)	12,49	(0,334)	9,95	(0,387)	9,99	(0,381)	9,92	(0,388)
sprawdzian ²					2,10	(0,289)	2,16	(0,292)	2,20	(0,277)	2,27	(0,279)	2,24	(0,279)
sprawdzian ³					-0,58	(0,075)	-0,58	(0,074)	-0,55	(0,076)	-0,55	(0,076)	-0,56	(0,077)
sprawdzian ⁴					-0,21	(0,047)	-0,22	(0,047)	-0,23	(0,046)	-0,24	(0,046)	-0,24	(0,046)
pleć ^a					-0,63	(0,375)	-0,71	(0,375)	-1,32	(0,327)	-1,38	(0,329)	-1,38	(0,328)
dysleksja na egz. gimnazjalnym ^b					3,02	(1,311)	2,89	(1,295)	2,02	(1,172)	1,93	(1,165)	1,87	(1,169)
dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym ^b					-5,74	(1,370)	-5,62	(1,365)	-4,38	(1,275)	-4,31	(1,281)	-4,17	(1,284)
wyższa z pozycji rodziców na międzynarodowej skali prestiżu zawodów (SIOPS)									0,09	(0,021)	0,09	(0,021)	0,08	(0,022)
zamożność gospodarstwa domowego									0,39	(0,216)	0,42	(0,216)	0,43	(0,218)
inteligencja (test matryc Ravena)									4,49	(0,241)	4,40	(0,235)	4,35	(0,237)
stała	101,49	(0,584)	93,36	(5,033)	100,71	(0,469)	97,49	(2,725)	97,19	(1,009)	96,96	(0,995)	88,95	(3,684)
poziom oddziały														
wykształcenie			5,97	(2,539)			2,72	(1,389)						
utrzymanie dyscypliny (DYSCNAU)														
potoczny nauczyciel (POMOCNAU)			1,12	(0,462)			0,52	(0,224)						
udział w szkoleniowych radach pedagogicznych			-3,22	(1,153)			-2,20	(0,612)						
aktywizujący styl nauczania (ACTIVEM5)			-6,01	(1,686)										
autorytet nauczyciela / utrzymanie dyscypliny - mat (DYSM)			2,51	(0,933)			3,09	(0,456)			2,94	(0,429)	3,05	(0,429)
potoczny nauczyciel - mat (POMOCNM)			3,34	(1,453)			-0,99	(0,537)			-1,23	(0,498)	-1,39	(0,458)
koncentracja na wymaganiach - mat (PODEGZM)														

zmienna zależna: wynik egzaminu gimnazjalnego w części matematycznej										
	(1M)	(s.e.)	(1MN)	(s.e.)	(2M)	(s.e.)	(2MN)	(s.e.)	(3M)	(s.e.)
									(3MN)	(s.e.)
średnia pozycja rodziców uczniów w klasie na międzynarodowej skali prestiżu zawodów (SIOPS)									0,19	(0,085)
średnia zamożność gosp. dom. uczniów w oddziale									-1,26	(1,288)
średni wynik sprawdzianu uczniów w oddziale									0,69	(0,738)
poziom szkoły										
oszacowanie efektów losowych										
wariancja efektów szkół	14,44		23,18		4,81		7,85		4,68	
wariancja efektów oddziałów (% wyj. war. po dodaniu predyktorów poz. oddziałów)	29,50		12,63	(57%)	6,35		0,61	(98%)	4,34	
wariancja na poziomie ucznia (% wyj. war. po dodaniu predyktorów poz. ucznia)	173,51		173,46		80,27	(54%)	80,30		71,24	(59%)
współczynniki korelacji wewnątrzgrupowej										
poziom szkół	0,066		0,111		0,053		0,088		0,058	
poziom oddziałów	0,136		0,060		0,069		0,007		0,054	
poziom szkół i oddziałów łącznie	0,202		0,171		0,122		0,095		0,112	
podsumowanie										
jakość dopasowania (deviance)	27296,68		27234,02		24587,49		24531,00		24170,11	
liczba szacowanych parametrów	4		12		11		16		14	
porównanie dopasowania modelu do modelu pustego	--		0,23%		9,92%		10,13%		11,45%	
poprawa dopasowania przed dodaniem predyktorów	--		0,23%		9,92%		0,23%		1,70%	
									0,18%	
										0,06%

liczba uczniów: 3370; liczba oddziałów: 214; liczba szkół: 143
pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe
a – grupa odniesienia: chłopcy, b – grupa odniesienia: uczniowie bez dysleksji

Istotność stałej w modelu pustym (1M) potwierdza występowanie efektu grupowania uczniów zarówno w oddziałach, jak i szkołach. W modelu pustym (1M) współczynnik korelacji wewnątrzoddziałowej, który można interpretować w procentach wyjaśnionej wariancji wyników egzaminu z części matematycznej, wynosi 0,136. Zgodnie z przyjętą praktyką, można więc mówić o efekcie nauczyciela w nauczaniu matematyki równym 14%.

Po dodaniu do modelu pustego (1M), wyłonionych uprzednio zmiennych, opisujących nauczycieli i ich pracę, model wyjaśniający zróżnicowanie wyników uczniów z części matematycznej egzaminu poprawił swe dopasowanie o 0,23% (1MN). Wprowadzone na poziomie drugim zmienne wyjaśniają 57% wariancji na poziomie oddziałów.

W kolejnym kroku wyłonione zmienne charakteryzujące nauczycieli i ich pracę, zostały przetestowane wspólnie, w modelu uwzględniającym baterię zmiennych, konstruujących wskaźnik EWD, adekwatny dla wyniku egzaminu gimnazjalnego z części matematycznej. Dzięki temu możliwa jest ocena trafności stosowanego w Polsce wskaźnika EWD dla gimnazjum dla części matematycznej egzaminu. Jeśli charakterystyki nauczycieli i ich pracy korelują z wynikami uczniów w modelu EWD zgodnie z oczekiwaniami, formułowanymi w oparciu o wyniki wcześniejszych badań, raportowanych w literaturze światowej, to tym samym potwierdzić można trafność interpretacji EWD jako miary efektywności nauczania.

Model EWD dla wyników z części matematycznej skonstruowany jest z wyniku sprawdzianu, wyniku sprawdzianu podniesionego do drugiej, trzeciej i czwartej potęgi, płci ucznia, dysleksji na egzaminie, dysleksji trwałej, mierzonej jako stwierdzenie dysleksji zarówno na sprawdzianie i egzaminie. Oryginalny model EWD dla wyniku egzaminu gimnazjalnego z części matematycznej zawierał także wynik sprawdzianu podniesiony do piątej potęgi. Jednak ze względu na redukcję liczebności próby, zmienna ta okazała się być nieistotna w modelu. Dlatego usunięto ją z dalszych analiz. Jej usunięcie nie wpłynęło na zmianę poziomów istotności pozostałych w modelu zmiennych, jak i ich współczynników regresji. Nieistotna w modelu EWD w analizowanej próbie okazała się płeć ucznia. Ponieważ jednak model EWD był szacowany na populacji uczniów, a także dlatego, że zmienna płeć jest ważna ze względów teoretycznych, zdecydowano się na pozostawienie jej w modelu.

Zgodnie z oczekiwaniami skala pomocny nauczyciel dodatnio koreluje z wynikami egzaminu z części matematycznej. Skalę należy interpretować w taki sposób, że im wyższa wartość na skali, tym bardziej otwarty i nastawiony na aktywizowanie uczniów styl nauczania. Zgodnie z interpretacją współczynnika regresji, wraz ze wzrostem o jedno odchylenie standardowe na skali Pomocny nauczyciel, uczeń średnio uzyskał o 1,12 punktu (0,075 odch. std. wyniku) więcej na egzaminie z części matematycznej. Czyli im bardziej otwarty i nastawiony na aktywizowanie uczniów styl nauczania, tym wyższe wyniki uczniów.

Udział nauczyciela w ciągu ostatnich trzech lat przynajmniej raz w szkolnych radach pedagogicznych, średnio o 3,22 punktu (0,215 odch. std.) obniża wyniki uczniów.

Skalę aktywizujący styl nauczania należy interpretować tak, że im wyższa wartość skali, tym bardziej aktywizujący styl nauczania. Niezgodnie z oczekiwaniami skala koreluje ujemnie z wynikami uczniów. Ponieważ składa się ona z merytorycznie różnych podskal, takich jak: indywidualizacja nauczania, multimedia, podsumowania, praca w grupach, efekt ten wymaga bardziej dokładnej analizy. Zrezygnowano z niej w tym miejscu, ponieważ skala przestaje być istotnym predyktorem wyników w momencie dodania do modelu baterii pytań konstruujących wskaźnik EWD.

Zgodnie z oczekiwaniami istotna okazała się skala autorytet nauczyciela/utrzymanie dyscypliny. Skala składa się z następujących pytań uporządkowanych według siły skorelowania z czynnikiem: Uczniowie nie słuchają tego, co mówi nauczyciel, Uczniowie podczas lekcji zajmują się swoimi sprawami, Nauczyciel musi długo czekać, zanim uczniowie się uspokoją i w klasie będzie cisza. Im wyższa wartość na skali, tym większa zdolność nauczyciela do utrzymywania dyscypliny na lekcjach i wzbudzania autorytetu. Wyższa dyscyplina na lekcjach o jedno odchylenie standardowe, przekłada się na średnio o 2,51 punktu (0,167 odch. std.) wyższy wynik ucznia.

Skalę pomocny nauczyciel należy interpretować tak, że im wyższe wartości na skali, tym bardziej pomocny nauczyciel. Zgodnie z oczekiwaniami skala dodatnio koreluje z wynikami uczniów. Czyli im bardziej pomocny jest nauczyciel, tym wyższe wyniki osiągają uczniowie. Jednak po dodaniu baterii zmiennych wskaźnika EWD, zmiennych dotyczących statusu społeczno-ekonomicznego i miary inteligencji, występuje odwrotny efekt. Fakt ten można tłumaczyć w następujący sposób. Skala pomocny nauczyciel składa się z następujących pytań podanych w kolejności od najmocniej skorelowanych z czynnikiem: Nauczyciel chętnie pomaga uczniom w nauce, Nauczyciel tłumaczy tak długo, aż wszyscy uczniowie zrozumieją, Nauczyciel wykazuje zainteresowanie postępami w nauce każdego z uczniów, Zajęcia są starannie zaplanowane, Nauczyciel podczas lekcji zachęca uczniów do zadawania pytań, dotyczących przerabianego materiału. Po dodaniu do modelu zmiennych informujących o wcześniejszych osiągnięciach edukacyjnych, statusie społeczno-ekonomicznym, inteligencji, składowe skali pomocny nauczyciel, mogą zmieniać swoje znaczenie i oznaczać aktywność nauczyciela w pracy z uczniami słabszymi lub też wyręczanie uczniów w realizacji zadań, zamiast ograniczania się do stymulowania indywidualnej pracy uczniów.

Po dodaniu charakterystyk nauczycieli (2MN), do modelu uwzględniającego już zmienne konstruujące wskaźnik EWD dla wyników z części matematycznej egzaminu (2M), model poprawia dopasowanie w stosunku do modelu pustego (1M) o 10,13%. Zmienne na poziomie ucznia wyjaśniają 54% wariancji wyników. Charakterystyki nauczyciela wyjaśniają 98% wariancji na poziomie oddziałów.

Uczniowie nie są losowo przypisani do oddziałów i nauczycieli, należy więc kontrolować zmienne odpowiedzialne za procesy segregacji. Zazwyczaj są one oparte na wcześniejszych

osiągnięciach uczniów oraz statusie społeczno-ekonomicznym. W celu kontroli tych czynników, wprowadzono jako zmienne poziomu oddziałów: średni wynik sprawdzianu uczniów w oddziale oraz średni status społeczno-ekonomiczny.

Model (3MNK) zawierający wszystkie zmienne z baterii konstruującej wskaźnik EWD, charakterystyki społeczno-ekonomiczne ucznia, miarę inteligencji, skale autorytet nauczyciela/utrzymanie dyscypliny i pomocny nauczyciel, a także zmienne kontrolujące skład oddziału pod względem wcześniejszych wyników i statusu społeczno-ekonomicznego, poprawia dopasowanie modelu o 11,67%, w porównaniu do modelu pustego. Kontrola składu oddziału poprawia dopasowanie modelu jedynie o 0,06%. Jednak chcąc wypowiadać się o wynikach w kategoriach przyczynowych, powinno się raportować współczynniki z modelu kontrolującego skład oddziału (3MNK).

Przy kontroli zmiennych konstruujących wskaźnik EWD, charakterystyk społeczno-ekonomicznych ucznia, inteligencji, składu oddziału, jedynymi istotnymi predyktorami wyników części matematycznej egzaminu są skale autorytet nauczyciela/utrzymanie dyscypliny i pomocny nauczyciel. Wyjaśniają one 98% wariancji na poziomie oddziałów w stosunku do modelu pustego (1M). 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 (Bryk i Raudenbush, 2002: 73–79). Pseudo r^2 mierzy stopień redukcji niewyjaśnionej wariancji na danym poziomie analizy (uczniowie, oddziały klasowe, szkoły) w modelu pustym w stosunku do modelu ze zmiennymi wyjaśniającymi. Sugeruje się jednak ostrożność w interpretacji pseudo r^2 . Nie należy przywiązywać dużej wagi do tej miary (Kreft i de Leeuw, 1998, s. 119).

Interpretacja ujemnej korelacji skali pomocny nauczyciel z wynikami egzaminu nastrocza pewne problemy, o czym była mowa już wcześniej. Natomiast stabilnym efektem, zgodnym z przewidywaniem, jest istotny efekt dyscypliny na lekcji. Wzrost o jedno odchylenie standardowe na skali dyscypliny lekcyjnej generuje średnio o 3,05 punktu (0,203 odch. std.) wyższy wynik egzaminu z części matematycznej.

10.8. Wyniki dla części polonistycznej

Tabela 7 prezentuje współczynniki i błędy standardowe dla modelu pustego (1P) z uwzględnieniem tylko losowej stałej regresji, oraz modeli, do których dodawano kolejno predyktory na poziomie uczniów i oddziałów. Wyniki surowe ze sprawdzianu i części polonistycznej egzaminu gimnazjalnego zostały wyskalowane za pomocą modelu 3PL. Wyniki uczniów raportowane są na skali o średniej w populacji 100 i odchyleniu standardowym 15. W analizowanej próbie odpowiednio 102,6 i 14,39. Średnia w próbie różni się od stałej w modelu ze względu na braki danych.

Istotność stałej w modelu pustym (1P) potwierdza występowanie efektu grupowania uczniów zarówno w oddziałach, jak i szkołach. W modelu pustym (1P) współczynnik korelacji wewnątrzoddziałowej, który można interpretować w procentach wyjaśnionej wariancji wyników egzaminu z części matematycznej, wynosi 0,122. Zgodnie z przyjętą praktyką, można więc mówić o efekcie nauczyciela w nauczaniu języka polskiego równym 12%.

Po dodaniu do modelu pustego (1P), wyłonionych uprzednio zmiennych, opisujących nauczycieli i ich pracę, model wyjaśniający zróżnicowanie wyników uczniów z części matematycznej egzaminu, poprawił swe dopasowanie o 0,24% (1MN). Wprowadzone na poziomie drugim. zmienne wyjaśniają 18% wariancji na poziomie oddziałów, czyli trzy razy mniej niż w przypadku matematyki. Oznacza to, że efekt nauczyciela w nauczaniu matematyki jest silniejszy, a tym samym łatwiejszy do zaobserwowania i modelowania za pomocą charakterystyk nauczycieli i ich pracy.

W kolejnym kroku wyłonione zmienne charakteryzujące nauczycieli i ich pracę, zostały przetestowane wspólnie w modelu uwzględniającym baterię zmiennych, konstruujących wskaźnik EWD, adekwatny dla wyniku egzaminu gimnazjalnego z części matematycznej. Model EWD dla wyników z części polonistycznej skonstruowany jest z wyniku sprawdzianu, wyniku sprawdzianu podniesionego do potęgi drugiej, trzeciej i czwartej, z płci ucznia, dysleksji na egzaminie, dysleksji na sprawdzianie.

Odwrotnie niż oczekiwano, skala wsparcie spoza szkoły koreluje z wynikami uczniów ujemnie. Prawidłowa interpretacja skali, jest taka, że im wyższa pozycja na skali, tym intensywniej nauczyciel deklaruje intencje korzystania z sieci wsparcia w sytuacji potrzeby. Ponieważ skala ta, po wprowadzeniu kolejnych zmiennych kontrolnych, przestaje być w modelu istotna, nie pogłębiono analizy, dlaczego skala nie zachowuje się zgodnie z oczekiwaniami. Jednym z wyjaśnień jest możliwość, że nauczyciele deklarujący intencje korzystania z sieci wsparcia w sytuacji potrzeby, mogą tak naprawdę nie radzić sobie z problemami wychowawczymi i dydaktycznymi. Skala mierzy wtedy bezradność nauczycieli. W takiej interpretacji oczekiwanym kierunkiem korelacji skali wsparcie spoza szkoły z wynikami uczniów jest kierunek ujemny.

Zgodnie z oczekiwaniami zachowuje się natomiast skala otwarty styl prowadzenia zajęć. Wraz ze wzrostem o jedno odchylenie standardowe na tej skali, uczeń średnio uzyska o 3,00 punkty (0,200 odch. std. wyniku) więcej na egzaminie z części polonistycznej.

Zgodnie z oczekiwaniami zachowuje się skala autorytet nauczyciela/utrzymanie dyscypliny. Im wyższa wartość na skali, tym większa zdolność nauczyciela do utrzymywania dyscypliny na lekcjach. Wyższa dyscyplina na lekcjach języka polskiego o jedno odchylenie standardowe, generuje średnio o 4,89 punktu (0,326 odch. std.) wyższy wynik ucznia.

Skalę pomocny nauczyciel należy interpretować tak, że im wyższe wartości na skali, tym bardziej pomocny nauczyciel. Niezgodnie z oczekiwaniami skala ujemnie koreluje

Tabela 7. Uwarunkowania osiągnięć językowych oraz charakterystyki nauczycieli i ich pracy. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych.

zmienna zależna: wynik egzaminu gimnazjalnego w części polonistycznej														
oszacowanie efektów stałych														
	(1P)	(s.e.)	(1PN)	(s.e.)	(2P)	(s.e.)	(2PN)	(s.e.)	(3P)	(s.e.)	(3PN)	(s.e.)	(3PNK)	(s.e.)
poziom uczenia														
sprawdzian					11,05	(0,275)	11,03	(0,274)	9,66	(0,318)	9,63	(0,321)	9,65	(0,325)
sprawdzian ²					-1,89	(0,268)	-1,89	(0,270)	-1,83	(0,282)	-1,82	(0,283)	-1,83	(0,283)
sprawdzian ³					-0,32	(0,075)	-0,33	(0,074)	-0,36	(0,080)	-0,37	(0,080)	-0,37	(0,080)
sprawdzian ⁴					0,20	(0,043)	0,20	(0,043)	0,18	(0,040)	0,18	(0,040)	0,18	(0,040)
pleć ^a					4,14	(0,336)	4,12	(0,336)	4,11	(0,366)	4,05	(0,364)	4,07	(0,363)
dysleksja na egz. gimnazjalnym ^b					1,86	(0,771)	1,78	(0,765)	1,83	(0,851)	1,70	(0,852)	1,74	(0,821)
dysleksja na spr ^b					-3,53	(0,822)	-3,41	(0,818)	-2,62	(1,023)	-2,52	(1,023)	-2,51	(0,999)
wyższa z pozycji rodziców na międzynarodowej skali prestiżu zawodów (SIOPS)									0,06	(0,017)	0,06	(0,017)	0,06	(0,018)
zamożność gospodarstwa domowego									1,40	(0,301)	1,41	(0,301)	1,44	(0,304)
Inteligencja (test matryc Ravena)									2,43	(0,253)	2,42	(0,254)	2,42	(0,253)
stała	101,70	(0,555)	101,82	(0,483)	101,03	(0,379)	101,01	(0,379)	98,45	(0,921)	97,06	(0,908)	90,48	(4,145)
poziom oddziału											1,46	(0,517)	1,19	(0,629)
pleć nauczyciela ^a														
udział w konferencjach metodycznych														
skala wsparcie spoza szkoły (EZZEW)			-1,15	(0,478)										
skala otwarty styl prowadzenia zajęć (OPENSTNU)			3,00	(1,735)										
skala autorytet nauczyciela / utrzymanie dyscypliny - pol (DYSF)			4,89	(1,110)			2,12	(0,464)			2,40	(0,61)	2,11	(0,590)
skala pomocny nauczyciel - pol (PO-MOCNP)			-2,80	(1,526)										
skala wymagający nauczyciel - pol (WYMIP)			-3,18	(1,058)										
skala koncentracja na wymaganiach - pol (PODEGZP)													1,38	(0,710)

zmienna zależna: wynik egzaminu gimnazjalnego w części polonistycznej												
	(1P)	(s.e.)	(1PN)	(s.e.)	(2P)	(s.e.)	(2PN)	(s.e.)	(3P)	(s.e.)	(3PN)	(s.e.)
średnia pozycja rodziców uczniów w klasie na międzynarodowej skali prestiżu zawodów (SIOPS)											0,16	(0,093)
średnia zarobkowość gosp. dom. uczniów w oddziale											-0,50	(1,387)
średni wynik sprawdzianu uczniów w oddziale											-0,92	(0,869)
poziom szkoły												
oszacowanie efektów losowych												
wariancja efektów szkół	20,05		12,80		1,77		2,30		2,86		1,56	
wariancja efektów oddziałów (% wyj. war. po dodaniu predyktorów poz. oddziałów)	25,37		20,70	(18%)	8,53	(74%)	6,67	(74%)	7,36	(74%)	6,61	(74%)
wariancja na poziomie ucznia (% wyj. war. po dodaniu predyktorów poz. ucznia)	162,04		161,93		78,50	(52%)	78,46		74,79	(54%)	74,77	
współczynniki korelacji wewnątrzgrupowej												
poziom szkół	0,097		0,066		0,020		0,026		0,034		0,019	
poziom oddziałów	0,122		0,106		0,096		0,076		0,087		0,080	
poziom szkół i oddziałów łącznie	0,219		0,171		0,116		0,103		0,120		0,099	
podsumowanie												
jakość dopasowania (deviance)	28274,56		28207,04		25592,14		25569,55		25431,32		25399,22	
liczba szacowanych parametrów	4		9		11		12		14		16	
porównanie dopasowania modelu do modelu pustego	--		0,24%		9,49%		9,57%		10,06%		10,17%	
poprawa dopasowania przed dodanie predyktorów	--		0,24%		9,49%		0,09%		0,63%		0,13%	

liczba uczniów: 3520, liczba oddziałów: 223, liczba szkół: 148

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe

a – grupa odniesienia: chłopcy / mężczyźni, b – grupa odniesienia: uczniowie bez dysleksji

z wynikami uczniów. Może to wynikać z przyczyn opisanych podczas omawiania podobnej sytuacji w modelu dla wyników części matematycznej.

Niezgodnie z oczekiwaniami, im bardziej wymagający nauczyciel języka polskiego, tym gorsze wyniki osiągają uczniowie. Wraz ze wzrostem o jedno odchylenie standardowe na skali wymagający nauczyciel, uczeń uzyskuje średnio o 3,18 punktu (0,212 odch. std.) niższy wynik z części polonistycznej egzaminu. Ponieważ efekt ten przestaje być istotny w momencie wprowadzenia do modelu zmiennych na poziomie uczniów, zrezygnowano z pogłębionych prób wyjaśnienia tego sprzecznego z intuicją kierunku związku.

Po dodaniu charakterystyk nauczycieli (2PN) do modelu EWD dla wyników z części polonistycznej egzaminu (2P), model poprawia dopasowanie w stosunku do modelu pustego (1P) o 9,57%. Zmienne na poziomie ucznia wyjaśniają 52% wariancji wyników na tym poziomie. Charakterystyki nauczyciela wyjaśniają 74% wariancji na poziomie oddziałów.

Model (3PNK) zawierający wszystkie zmienne z baterii konstruującej wskaźnik EWD, charakterystyki społeczno-ekonomiczne, miarę inteligencji, płeć nauczyciela oraz skalę autorytet nauczyciela/utrzymanie dyscypliny, a także kontrolujące skład oddziału, poprawia dopasowanie modelu o 10,20%. Kontrola składu oddziału poprawia dopasowanie modelu jedynie o 0,04%. Jednak, chcąc wypowiadać się o wynikach w kategoriach przyczynowych, powinno się raportować współczynniki z modelu kontrolującego skład oddziału (3PNK).

Przy kontroli zmiennych konstruujących wskaźnik EWD, charakterystyk społeczno-ekonomicznych ucznia, inteligencji, składu oddziału, istotnym predyktorem wyników części polonistycznej egzaminu jest skala autorytet nauczyciela/utrzymanie dyscypliny. Efekt dyscypliny na lekcji jest stabilny i zgodny z oczekiwaniami. Wzrost o jedno odchylenie standardowe na skali dyscypliny, daje średnio o 2,11 punktu (0,141 odch. std.) wyższy wynik egzaminu z części polonistycznej. W modelu kontrolującym skład oddziału (3PNK) istotna okazała się skala koncentracja na wymaganiach. Jej interpretacja jest taka, że im wyższe wartości na skali, tym bardziej intensywne jest nauczanie pod kątem egzaminu. Współczynnik regresji przy tej zmiennej wskazuje, że uczniowie nauczani pod kątem egzaminu z języka polskiego, uzyskują więcej punktów na egzaminie z części polonistycznej. Przy wzroście o jedno odchylenie standardowe na skali, uczniowie uzyskują średnio o 1,38 punktu więcej (0,092 odch. std.). Na granicy istotności ($p = 0,06$) jest też płeć nauczyciela. Uczniowie nauczani języka polskiego przez mężczyznę osiągają średnio o 1,19 (0,079 odch. std.) punktu mniej na egzaminie z części polonistycznej. Odkrycie to należy traktować ostrożnie, ponieważ w analizowanej próbie odsetek mężczyzn uczących języka polskiego wynosi tylko 2,4%. Łącznie zmienne na poziomie oddziału w modelu pełnym, kontrolującym skład oddziału (3PNK) wyjaśniają 71% wariancji na poziomie oddziału w stosunku do modelu pustego (1P).

10.9. Podsumowanie

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

Charakterystyki nauczycieli wyjaśniają 98% wariancji na poziomie oddziałów w przypadku modeli dla osiągnięć matematycznych, uwzględniających baterię zmiennych, konstruujących wskaźnik EWD. W przypadku osiągnięć polonistycznych charakterystyki nauczycieli wyjaśniają 74% wariancji wyników egzaminu z części polonistycznej egzaminu na poziomie oddziałów.

W przypadku uwzględnienia wyniku sprawdzianu, płci, dysleksji, statusu społeczno-ekonomicznego rodziny, uzyskano wynik 6% (matematyka) i 9% (kompetencje językowe). Rowan, Correnti i Miller (2002) raportują 8–18% i 4–16%. Przy uwzględnieniu dodatkowo miary inteligencji i kontroli składu oddziału pod kątem kompozycji społecznej i wcześniejszych osiągnięć uzyskano odpowiednio 7% i 9%. Podane wartości to współczynniki korelacji wewnątrzgrupowej, dla których w przypadku modeli z predyktorami zastosowano taką samą interpretację jak dla modeli pustych. Wartości te to procent wyjaśnionej wariancji wyników uczniów z wariancji pozostałej do wyjaśnienia na poziomie oddziałów, po uwzględnieniu predyktorów na poziomie ucznia.

Wykazano brak liniowego związku stażu pracy z osiągnięciami uczniów. Jest to wniosek zgodny z istniejącymi dowodami (Murnane i Phillips, 1981; Klitgaard i Hall, 1975), choć dalszych analiz wymaga testowanie nieliniowego związku stażu nauczycieli z wynikami uczniów.

Odnotowano pozytywny związek wykształcenia nauczycieli z wynikami uczniów w nauczaniu matematyki. Jest to zgodne z dowodami przedstawionymi przez Nye, Konstantopoulosa, Hedgesa (2004) oraz Wayne'a i Youngsa (2003). Związek ten staje się jednak nieistotny przy kontroli inteligencji i statusu społeczno-ekonomicznego rodziny ucznia.

Doskonalenie zawodowe okazało się być nieistotnym czynnikiem w wyjaśnianiu wyników uczniów. Odnotowano jednak niewielki negatywny wpływ na wyniki uczniów uczestnictwa w szkoleniowych radach pedagogicznych i w konferencjach metodycznych. Rezultat ten niekoniecznie musi generować praktyczną konsekwencję w formie rekomendacji rezygnacji ze szkoleń nauczycieli. Powinien raczej inspirować krytyczną analizę jakości tych szkoleń.

Ciekawym efektem okazał się pozytywny wpływ płci nauczyciela na wyniki egzaminu z części polonistycznej. W nauczaniu języka polskiego nauczycielki przyczyniają się bardziej

do podniesienia wyników uczniów niż nauczyciele. Efekt ten wymaga jednak dalszej eksploracji na większych próbach, gdzie odsetek nauczycieli mężczyzn jest wyższy, niż w próbie tu analizowanej.

Należy odnotować, że informacje o wcześniejszych osiągnięciach edukacyjnych uczniów, płci, statusie społeczno-ekonomicznym i inteligencji znoszą istotność lub osłabiają efekt charakterystyk nauczycieli. Jedynym stabilnym i o jednoznacznej interpretacji efektem związanym ze sposobem prowadzenia lekcji, jest utrzymanie dyscypliny. Nauczyciele potrafiący utrzymać dyscyplinę na lekcjach, zaciekać tematem uczniów, zająć ich uwagę, przyczyniają się do podniesienia wyników uczniów średnio o 0,167 odchylenia standardowego z matematyki i 0,141 odchylenia standardowego z języka polskiego.

Podsumowując, literatura dotycząca efektu nauczyciela nie dostarcza prostych odpowiedzi na pytanie, za pomocą jakich charakterystyk samych nauczycieli, jak i charakterystyk ich pracy, można wyjaśnić zróżnicowanie w ich efektywności. Jednak, większość charakterystyk nauczycieli i ich pracy zachowuje się w modelach EWD zgodnie z oczekiwaniami, formułowanymi na podstawie wyników wcześniejszych badań. Potwierdza to prawomocność interpretacji EWD jako trafnej miary efektywności kształcenia.

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*. R-2007-LAUDS. 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.
- 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*. Learning Publications (Holmes Beach, Fla.).
- Brown, C., Smith, M. i Stein, M. (1996). Linking teacher support to enhanced classroom instruction. *Prezentowane 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. *Prezentowane na konferencji Northwestern Educational Research Association*, Ellenville, NY.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (wyd. 2.). Hillsdale, NJ: L. Erlbaum Associates.
- 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, DC: 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.
- 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 Revisted. *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:] Ladd, H. F. (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 productivityon. *The 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.
- 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.
- Klitgaard, R. E. i Hall, G. R. (1975). Are There Unusually Effective Schools. *Journal of Human Resources*, 10(1), 90–106.
- Kreft, I. i de Leeuw, J. (1998) *Introducing Multilevel Modeling*. Thousand Oaks: 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., Pickering, D. J. (2003). *Classroom management that works*. Alexandria, VA: ASCD.

McDonald, R. P. i Ho, M.-H. R. (2002). Principles and Practice in Reporting Structural Equation Analyses. *Psychological methods*, 7(1), 64–82.

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. *Prezentowane na National Research Council Conference on Teacher Quality in Science and Mathematics*, Washington, DC.

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.

Rivkin, S. G., Hanushek, E. A. i Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.

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.

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.

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, W., Horn, S. i Sanders, W. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 4(1), 3–7.

Aneks do rozdziału 1

Organizacja i metodologia badania podłużnego w gimnazjum

Międzyszkolne i międzyoddziałowe zróżnicowanie wyników egzaminacyjnych

W tabelach 1, 2 oraz 3 zawarto wyniki trzypoziomowych modeli regresji bez predyktorów (tzw. modeli pustych), w których zmienne zależne stanowiły odpowiednio: wyniki sprawdzianu po klasie szóstej, wyniki egzaminu gimnazjalnego w części humanistycznej oraz wyniki egzaminu gimnazjalnego w części matematyczno-przyrodniczej. Każdy rodzaj modelu obliczono na czterech zestawach danych: (1) wynikach w całej populacji, (2) wynikach w populacji z wykluczeniem danych uczniów uczęszczających do szkół wykluczonych z operatu losowania, (3) wynikach uczniów ze szkół wylosowanych, (4) wynikach uczniów ze szkół, w których zrealizowano badanie. Obliczenia wykonano za pomocą oprogramowania Stata 11, estymacja przeprowadzona została metodą *full maximum likelihood*.

Porównanie modeli populacyjnych i modeli dla danych z próby badawczej:

W tabelach 5 i 6 znajduje się porównanie współczynników i parametrów modeli dwupoziomowych ważonych (tzw. modeli podstawowych) uwzględniających zmienne takie jak w modelach EWD oraz zmienne z poziomu indywidualnego (wyskalowany wynik testu inteligencji, zmienne opisujące wiek ucznia, społeczno-ekonomiczny status jego rodziny, wykształcenie rodziców oraz wielkość i strukturę rodziny ucznia), wykonanych w programach Stata 11 i HLM6. Procedura wyboru zmiennych do tych modeli opisana została w rozdziale metodologicznym w punkcie drugim podrozdziału o modelach podstawowych stosowanych w analizach.

W tabeli 7 znajduje się porównanie trzypoziomowych modeli podstawowych na danych ważonych z uwzględnieniem dodatkowych zmiennych mogących wpływać na wyniki egzaminacyjne uczniów. Zbiór tych zmiennych jest analogiczny jak w dwupoziomowych modelach podstawowych. Procedura ich szacowania opisana została w rozdziale 1 w punkcie trzecim podrozdziału o modelach podstawowych stosowanych w analizach.

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 1. Wyniki trzypoziomowych pustych modeli regresji; zmienna zależna: wynik sprawdzianu po szóstej klasie

zmienna zależna: sprawdzian	1	(s.e.)	2	(s.e.)	3	(s.e.)	4	(s.e.)
oszacowania efektów stałych								
poziom ucznia								
stała	98,88	(0,078)	98,80	(0,077)	99,71	(0,502)	99,86	(0,473)
oszacowania efektów losowych								
wariancja efektów szkół	25,01	(0,724)	18,99	(0,597)	27,15	(4,28)	23,28	(3,80)
wariancja efektów oddziałów	21,89	(0,408)	21,66	(0,406)	29,17	(2,42)	27,09	(2,28)
wariancja na poziomie ucznia	179,07	(0,430)	179,45	(0,444)	168,36	(2,07)	171,87	(2,11)
ICC - poziom szkół	11,07%		8,63%		12,08%		10,47%	
ICC - poziom klas	9,69%		9,84%		12,98%		12,19%	
podsumowanie								
log pseudolikelihood	-1480415,20		-389790,40		-55924,32		-56151,02	
liczba uczniów	365290		343037		13896		13924	
liczba oddziałów	18343		16521		636		634	
liczba szkół	6650		5161		150		150	

pogrubionym drukiem zaznaczone wartości istotne statystycznie na poziomie $p < 0,05$; w nawiasach podano błędy standardowe;
 1 – populacja; 2 – populacja z wykluczeniem uczniów ze szkół małych, niepublicznych, specjalnych, przyszpitalnych, dla dorosłych;
 3 – wylosowana próba szkół; 4 – zrealizowana próba szkół; ICC – współczynnik korelacji wewnątrzgrupowej

Tabela 2. Wyniki trzypoziomowych pustych modeli regresji; zmienna zależna: wynik egzaminu gimnazjalnego, część humanistyczna

zmienna zależna: EG hum	1	(s.e.)	2	(s.e.)	3	(s.e.)	4	(s.e.)
oszacowania efektów stałych								
poziom ucznia								
stała	99,41	(0,090)	99,71	(0,080)	100,66	(0,515)	100,57	(0,478)
oszacowania efektów losowych								
wariancja efektów szkół	37,47	(1,01)	20,93	(0,652)	28,76	(4,47)	23,71	(3,86)
wariancja efektów oddziałów	25,29	(0,448)	24,54	(0,433)	31,08	(2,49)	28,90	(2,33)
wariancja na poziomie ucznia	165,24	(0,392)	166,88	(0,410)	157,46	(1,92)	157,88	(1,92)
ICC - poziom szkół	16,43%		9,85%		13,24%		11,26%	
ICC - poziom klas	11,09%		11,56%		14,30%		13,73%	

ANEKS DO ROZDZIAŁU 1

ORGANIZACJA I METODOLOGIA BADANIA PODŁUŻNEGO W GIMNAZJUM

zmienna zależna: EG hum	1 (s.e.)	2 (s.e.)	3 (s.e.)	4 (s.e.)
podsumowanie				
log pseudolikelihood	-1503678,70	-1399255,80	-56271,46	-56338,96
liczba uczniów	374204	348209	14092	14111
liczba oddziałów	18583	16565	637	635
liczba szkół	6776	5163	150	150

pogrubionym drukiem zaznaczone wartości istotne statystycznie na poziomie $p < 0,05$; w nawiasach podano błędy standardowe; 1 – populacja; 2 – populacja z wykluczeniem uczniów ze szkół małych, niepublicznych, specjalnych, przyszpitalnych, dla dorosłych; 3 – wylosowana próba szkół; 4 – zrealizowana próba szkół; ICC – współczynnik korelacji wewnątrzgrupowej

Tabela 3. Wyniki trzypoziomowych pustych modeli regresji; zmienna zależna: wynik egzaminu gimnazjalnego, część matematyczno-przyrodnicza

zmienna zależna: EG mat.-przyr.	1 (s.e.)	2 (s.e.)	3 (s.e.)	4 (s.e.)
oszacowania efektów stałych				
poziom ucznia				
stała	99,35 (0,080)	98,95 (0,078)	99,85 (0,502)	99,95 (0,484)
oszacowania efektów losowych				
wariancja efektów szkół	26,14 (0,736)	19,54 (0,608)	26,96 (4,28)	24,87 (3,99)
wariancja efektów oddziałów	23,24 (0,403)	22,97 (0,400)	30,86 (2,41)	28,33 (2,26)
wariancja na poziomie ucznia	155,43 (0,371)	155,32 (0,381)	150,19 (1,83)	153,58 (1,87)
ICC - poziom szkół	12,76%	9,88%	12,96%	12,03%
ICC - poziom klas	11,35%	11,61%	14,83%	13,70%
podsumowanie				
log pseudolikelihood	-1472548,30	-1385460,80	-55887,06	-56113,82
liczba uczniów	369485	347873	14077	14102
liczba oddziałów	18098	16564	637	635
liczba szkół	6428	5163	150	150

pogrubionym drukiem zaznaczone są wartości istotne statystycznie na poziomie $p < 0,05$; w nawiasach podano błędy standardowe; 1 – populacja; 2 – populacja z wykluczeniem uczniów ze szkół małych, niepublicznych, specjalnych, przyszpitalnych, dla dorosłych; 3 – wylosowana próba szkół; 4 – zrealizowana próba szkół; ICC – współczynnik korelacji wewnątrzgrupowej

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Tabela 4. Porównanie dwupoziomowych modeli ważonych – część humanistyczna egzaminu

	część humanistyczna		język polski		WOS i historia	
	Stata 11	HLM 6	Stata 11	HLM 6	Stata 11	HLM 6
oszacowanie efektów stałych						
poziom ucznia						
wynik sprawdzianu (skala 0;1)	9,930	9,930	9,350	9,350	9,310	9,309
druga potęga wyniku sprawdzianu	-0,399	-0,399	-1,707	-1,706		
trzecia potęga wyniku sprawdzianu	-0,336	-0,335	-0,323	-0,323	-0,413	-0,413
czwarta potęga wyniku sprawdzianu			0,171	0,171		
pleć: kobieta	1,359	1,356	4,567	4,563	-2,382	-2,384
dysleksja na egz. gimnazjalnym	1,221	1,223	1,610	1,613	0,613	0,609
dysleksja na sprawdzianie	-2,641	-2,642	-3,371	-3,372	-1,777	-1,774
dysleksja na obu egzaminach						
IQ (TMS-K etap 1)	2,440	2,438	2,490	2,490	2,162	2,160
wyższy z poziomów wyszst. rodziców	0,459	0,460	0,377	0,379	0,483	0,483
wyższy ze statusów społ.-ekon. rodziców						
wyższy z prestiżu zawodów rodziców						
wiek ucznia w miesiącach	-0,106	-0,106			-0,084	-0,084
opóźnione rozpoczęcie nauki w gimnazjum						
przyspieszone rozpoczęcie nauki w gimnazjum						
zasobność gospodarstwa domowego	1,490	1,490	1,177	1,177	1,629	1,627
liczba dzieci w rodzinie	-0,449	-0,450	-0,476	-0,478		
wielkość rodziny (l. osób w gosp. domowym)						
kolejność urodzenia ucznia (ref. jednak)						
najmłodszy					-1,369	-1,374
w środku					-1,652	-1,661
najstarszy					0,256	0,249
brak danych					-0,748	-0,736
struktura rodziny (ref. pełna dwupokoleniowa)						
pełna trzypokoleniowa						
niepełna dwupokoleniowa						
niepełna trzypokoleniowa						
inna						
brak danych						
stała	96,239	96,232	97,438	97,427	96,944	96,949
oszacowanie efektów losowych						
wariancja efektów szkół	6,625	6,681	8,400	8,497	5,472	5,513
wariancja na poziomie ucznia	71,973	72,030	78,212	78,262	91,884	92,021
podsumowanie						
deviance	1082119	33517,9	1095456,8	33928,1	1117727,6	34612,0
liczba szacowanych parametrów	14	14	14	14	16	16
liczba uczniów	4680	4680	4681	4681	4681	4681
liczba szkół	150	150	150	150	150	150

ANEKS DO ROZDZIAŁU 1

ORGANIZACJA I METODOLOGIA BADANIA PODŁUŻNEGO W GIMNAZJUM

Tabela 5. Porównanie dwupoziomowych modeli ważonych – część matematyczno-przyrodnicza

	część mat-przyr		część matematyczna		część przyrodnicza	
	Stata 11	HLM 6	Stata 11	HLM 6	Stata 11	HLM 6
oszacowanie efektów stałych						
poziom ucznia						
wynik sprawdzianu (skala 0;1)	8,879	8,875	10,217	10,214	9,126	9,122
druga potęga wyniku sprawdzianu	-0,294	-0,295	2,306	2,302	0,862	0,860
trzecia potęga wyniku sprawdzianu	-0,312	-0,312	-0,639	-0,639	-0,292	-0,292
czwarta potęga wyniku sprawdzianu			-0,240	-0,240		
płeć: kobieta	-2,084	-2,086	-1,515	-1,516	-1,395	-1,400
dysleksja na egz. gimnazjalnym	1,650	1,651	1,469	1,474	1,119	1,116
dysleksja na sprawdzianie					-2,180	-2,171
dysleksja na obu egzaminach	-3,239	-3,241	-3,781	-3,785		
IQ (TMS-K etap 1)	4,158	4,158	4,541	4,539	3,382	3,385
wyższy z poziomów wyksz. rodziców			0,246	0,246		
wyższy ze statusów społ.-ekon. rodziców					0,049	0,049
wyższy z prestiżu zawodów rodziców	0,076	0,076	0,041	0,041		
wiek ucznia w miesiącach	-0,133	-0,133	-0,129	-0,129	-0,135	-0,135
opóźnione rozpoczęcie nauki w gimnazjum						
przyspieszone rozpoczęcie nauki w gimnazjum			3,812	3,805		
zasobność gospodarstwa domowego	0,777	0,776			1,004	1,002
liczba dzieci w rodzinie					-0,242	-0,243
wielkość rodziny (l. osób w gosp. domowym)						
kolejność urodzenia ucznia (ref. jedynak)						
najmłodszy			1,047	1,043		
w środku			0,568	0,560		
najstarszy			0,606	0,601		
brak danych			-1,097	-1,102		
struktura rodziny (ref. pełna dwupokoleniowa)						
pełna trzypokoleniowa	0,205	0,197	0,282	0,272		
niepełna dwupokoleniowa	-1,368	-1,367	-1,438	-1,438		
niepełna trzypokoleniowa	-1,290	-1,290	-1,645	-1,645		
inna	-1,495	-1,483	0,124	0,135		
brak danych	-6,223	-6,223	-5,977	-5,971		
stała	98,682	98,698	94,814	94,830	99,069	99,072
oszacowanie efektów losowych						
wariancja efektów szkół	5,000	5,030	7,634	7,689	5,314	5,360
wariancja na poziomie ucznia	76,194	76,333	72,067	72,285	90,589	90,676
podsumowanie						
deviance	1108928,9	34394,8	1063158,8	32934,5	1135187,4	35218,7
liczba szacowanych parametrów	18	18	24	24	14	14
liczba uczniów	4771	4771	4596	4596	4771	4771
liczba szkół	150	150	150	150	150	150

Tabela 6. Porównanie trzypoziomowych modeli ważonych z uwzględnieniem dodatkowych zmiennych mogących wpływać na wyniki egzaminacyjne uczniów

	część					
	matematyczno- -przyrodnicza	matematyczna	przyrodnicza	humanistyczna	WOS i historia	humanistyczna
oszacowanie efektów stałych						
poziom ucznia						
wynik sprawdzianu	8,875	10,791	9,132	9,903	9,399	9,333
druga potęga wyniku sprawdzianu	-0,285	2,160	0,849	-0,764	-1,819	0,211
trzecia potęga wyniku sprawdzianu	-0,310	-1,071	-0,290	-0,321	-0,343	-0,408
czwarta potęga wyniku sprawdzianu		-0,200		0,070	0,190	
piąta potęga wyniku sprawdzianu		0,058				
pleć: kobieta	-1,890	-1,485	-1,233	1,322	4,517	-2,395
dysleksja na egz. gimnazjalnym	1,798	1,325		1,232	1,661	
dysleksja na sprawdzianie			-1,352	-2,472	-3,224	-0,983
dysleksja na obu egzaminach	-3,410	-3,717				
IQ (TMS-K etap 1)	4,030	4,438	3,282	2,365	2,382	2,121
wyższy z poziomów wyksz. rodziców	0,193	0,228	0,182	0,422	0,330	0,458
wyższy ze statusów społ.-ekon. rodziców			0,036			
wyższy z prestiżu zawodów rodziców	0,050	0,042		0,031	0,032	0,028
wiek ucznia w miesiącach	0,571	-0,141	0,879	1,467		1,588
opóźnione rozpoczęcie nauki w gimnazjum					1,119	
zasobność gospodarstwa domowego	-0,131		-0,147	-0,591	-3,233	-0,751
wyposażenie gospodarstwa w sprzęt elektroniczny				-0,106		-0,088
kolejność urodzenia ucznia (ref. jedynak)						
najmłodszy		1,159		-1,134	-0,979	-1,298
w środku		0,747		-1,937	-1,925	-1,500
najstarszy		0,728		-0,066	-0,638	0,409
brak danych		-0,682		-0,383	-0,898	-0,533
struktura rodziny (ref. pełna 2-pokolenia)						
pełna trzypokoleniowa	0,271	0,219				
niepełna dwupokoleniowa	-1,355	-1,550				
niepełna trzypokoleniowa	-1,152	-1,518				
inna	-1,108	-0,076				
brak danych	-5,480	-5,272				
stała	97,341	94,911	96,632	95,063	96,345	95,670
oszacowanie efektów losowych						
wariancja efektów szkół	4,04	5,49	3,43	5,06	6,05	3,48
wariancja efektów klas	1,70	3,90	3,10	1,96	3,32	2,45
wariancja na poziomie ucznia	75,64	70,31	89,43	70,02	76,12	89,58

ANEKS DO ROZDZIAŁU 1

ORGANIZACJA I METODOLOGIA BADANIA PODŁUŻNEGO W GIMNAZJUM

	część					
	matematyczno- -przyrodnicza	matematyczna	przyrodnicza	humanistyczna	WOS i historia	humanistyczna
podsumowanie						
deviance	33099,7	32898,2	33872,6	32776,7	33207,3	33863,8
liczba szacowanych parametr.	20	25	14	21	20	19
liczba uczniów	4591	4596	4591	4590	4591	4591
liczba klas	291	291	291	291	291	291
liczba szkół	150	150	150	150	150	150

Aneks do rozdziału 2

Trafność testów egzaminacyjnych

Kwestionariusz integracji uczniów (KIU)

Narzędzie to jest adaptacją szwajcarskiego kwestionariusza *Fragebogen zur Erfassung von Dimensionen der Integration von Schülern FDI 4–6* (Haeberlin, Moser i Bless, 1989), służącego do badania funkcjonowania szkolnego uczniów. Polska adaptacja została opracowana przez Grzegorza Szumskiego. Test składa się z 45 pozycji podzielonych na trzy skale: 1) skalę integracji emocjonalnej wyrażającą nastawienie uczniów do szkoły, 2) skalę integracji motywacyjnej, która odnosi się do obrazu siebie jako ucznia za pomocą samooceny własnych szkolnych kompetencji oraz 3) skalę integracji społecznej, która mierzy satysfakcję uczniów z kontaktów z kolegami z klasy i szkoły. Badany ustosunkowuje się do każdego stwierdzenia zaznaczając jedną z czterech możliwych odpowiedzi (tak/raczej tak/raczej nie/nie). W założeniu autorów narzędzia, każda skala jest wskaźnikowana przez 15 pozycji. Jednak przeprowadzona eksploracyjna analiza czynnikowa wykazała, że ostatnie stwierdzenie (*Chętnie się uczę*) jest związane przede wszystkim z wymiarem integracji emocjonalnej, a nie jak zakładali autorzy, z integracją motywacyjną. Pozycja ta została więc potraktowana w modelu jako wskaźnik integracji emocjonalnej.

W celu weryfikacji właściwości narzędzia oraz wyznaczenia wyników dla poszczególnych wymiarów integracji przeprowadzono konfirmacyjną analizę czynnikową dla zmiennych porządkowych. Obliczenia wykonano za pomocą oprogramowania MPlus 6.1 (Muthén i Muthén, 1998–2010), z wykorzystaniem estymatora WLSMV. W modelu uwzględniono zgrupowanie uczniów w klasach. W analizach wykorzystane zostały dane wszystkich uczniów, którzy wypełnili kwestionariusz (5081 uczniów). Tak wyspecyfikowany model osiągnął akceptowalne dopasowanie do danych (patrz poniższa tabela).

Tabela 1. Miary dopasowania ostatecznego modelu konfirmacyjnej analizy czynnikowej

miara dopasowania	wartość
chi ²	12622,237 (df = 942, p < 0,001)
RMSEA	0,049 (PU: 0,049-0,050)
CFI	0,928
TLI	0,924

ANEKS DO ROZDZIAŁU 2: TRAFNOŚĆ TESTÓW EGZAMINACYJNYCH

W poniższej tabeli podano wartości standaryzowanych ładunków czynnikowych poszczególnych pozycji dla trzech skal.

Tabela 2. Standaryzowane ładunki czynnikowe wraz z błędem standardowym ich oszacowań dla poszczególnych wymiarów integracji szkolnej

skala	pozycja	standaryzowany ładunek czynnikowy	błąd standardowy
Integracja emocjonalna	KIU_1	0,802	0,006
	KIU_4	0,661	0,009
	KIU_7	0,656	0,011
	KIU_10	0,845	0,005
	KIU_13	0,713	0,008
	KIU_16	0,616	0,011
	KIU_19	0,681	0,009
	KIU_22	0,813	0,006
	KIU_25	0,815	0,006
	KIU_28	0,669	0,013
	KIU_31	0,541	0,013
	KIU_34	0,741	0,01
	KIU_37	0,752	0,009
	KIU_40	0,532	0,013
	KIU_43	0,797	0,007
	KIU_45	0,637	0,011
Integracja motywacyjna	KIU_3	0,782	0,007
	KIU_6	0,436	0,014
	KIU_9	0,731	0,007
	KIU_12	0,595	0,011
	KIU_15	0,741	0,008
	KIU_18	0,754	0,007
	KIU_21	0,769	0,006
	KIU_24	0,675	0,009
	KIU_27	0,786	0,007
	KIU_30	0,721	0,008
	KIU_33	0,675	0,01
	KIU_36	0,769	0,007
	KIU_39	0,674	0,01
	KIU_42	0,795	0,007

TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

skala	pozycja	standaryzowany ładunek czynnikowy	błąd standardowy
Integracja społeczna	KIU_2	0,73	0,01
	KIU_5	0,737	0,008
	KIU_8	0,659	0,01
	KIU_11	0,687	0,009
	KIU_14	0,799	0,007
	KIU_17	0,56	0,011
	KIU_20	0,701	0,009
	KIU_23	0,571	0,015
	KIU_26	0,698	0,009
	KIU_29	0,675	0,012
	KIU_32	0,809	0,007
	KIU_35	0,775	0,008
	KIU_38	0,686	0,009
	KIU_41	0,697	0,01
	KIU_44	0,68	0,011

Poszczególne wymiary integracji szkolnej są ze sobą skorelowane. Macierz korelacji latentnych znajduje się w poniższej tabeli.

Tabela 3. Korelacje latentne między poszczególnymi wymiarami integracji szkolnej

	int. emocjonalna	int. motywacyjna	int. społeczna
int. emocjonalna	1		
int. motywacyjna	0,431	1	
int. społeczna	0,307	0,254	1

Na podstawie tak wyspecyfikowanego modelu wygenerowano wartości czynnikowe dla poszczególnych obserwacji, które następnie wystandaryzowano (wyrażono w jednostkach odchylenia standardowego). Statystyki opisowe dla poszczególnych wymiarów integracji szkolnej podano w poniższej tabeli.

Tabela 4. Statystyki opisowe dla oszacowań wyników skal integracji (N = 5081)

skala integracji	średnia	odch. std.	min	mediana	max	kurtoza	skośność
emocjonalnej	0	1	-2,92	0,042	3,79	2,97	-0,099
motywacyjnej	0	1	-3,43	0,017	3,16	3,26	0,028
społecznej	0	1	-3,79	0,047	2,68	3,19	-0,276

Aneks do rozdziału 4

Wiek i inteligencja uczniów

Wiek uczniów – charakterystyka zmiennych

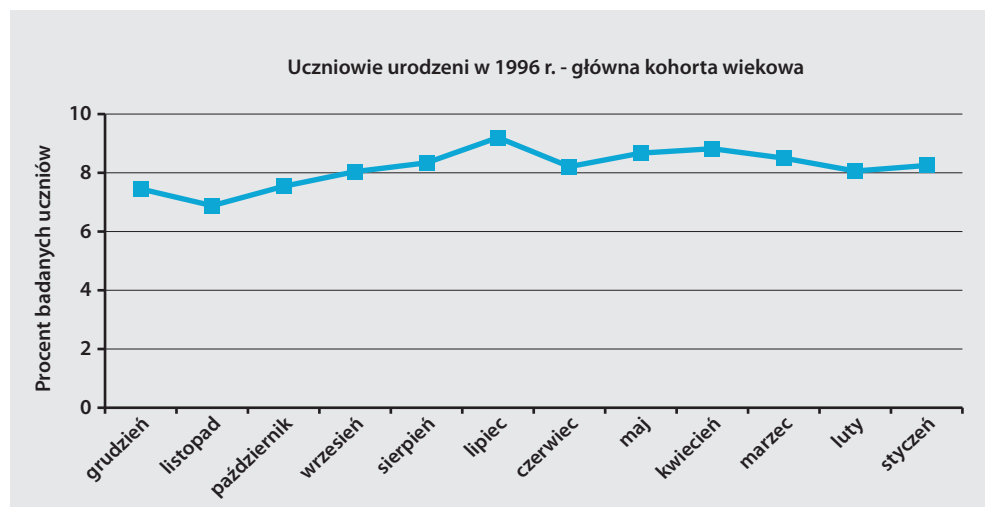
Tabela 1. Rozkład zmiennej rok urodzenia oraz wskaźnika bycia młodszym lub starszym niż główna kohorta wiekowa

	rok urodzenia	liczba uczniów w próbie	procent całości	procent całości
starsi niż główna kohorta wiekowa	1993	1	0,02%	1,28%
	1994	4	0,08%	
	1995	62	1,18%	
główna kohorta wiekowa	1996	5142	97,96%	97,96%
młodszy niż główna kohorta wiekowa	1997	40	0,76%	0,76%
razem		5249	100%	100%

Tabela 1 prezentuje rozkład zmiennej rok urodzenia oraz sposób tworzenia i rozkład zmiennych starszy niż główna kohorta i młodszy niż główna kohorta.

Rysunki 1–3 prezentują proporcje uczniów urodzonych w poszczególnych miesiącach roku w głównej kohorcie wiekowej oraz wśród uczniów młodszych niż główna kohorta i o rok starszych niż główna kohorta. Brak miesiąca na wykresie oznacza brak w próbie uczniów urodzonych w tym okresie.

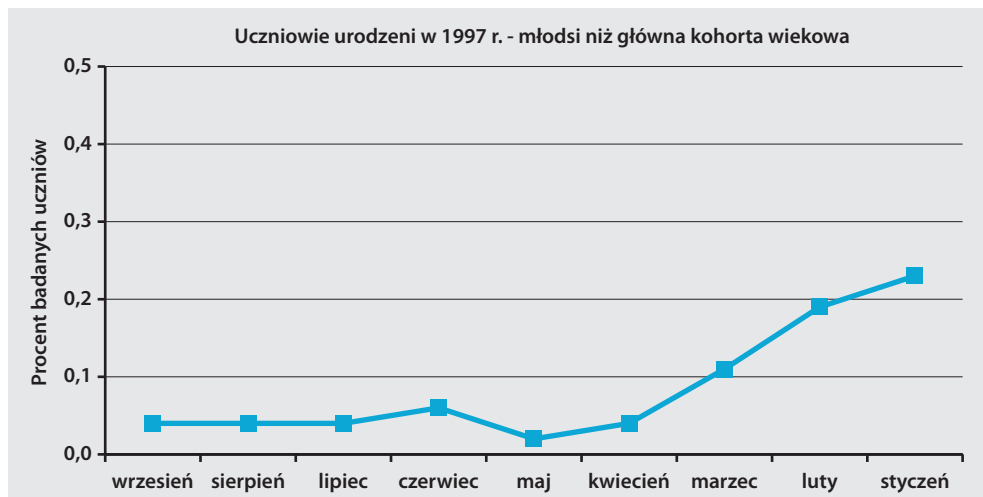
Rysunek 1. Procent uczniów urodzonych w poszczególnych miesiącach w głównej kohorcie wiekowej



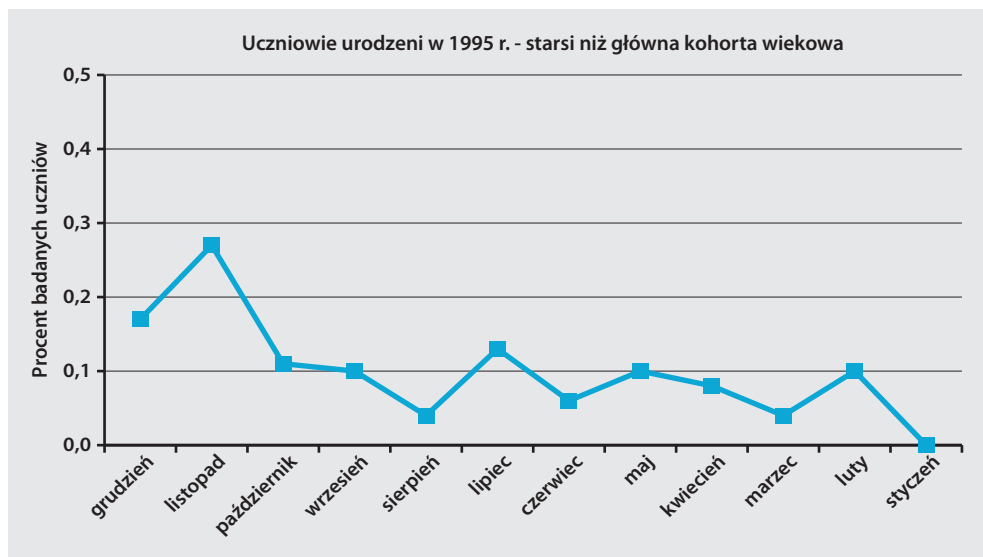
TRAFNOŚĆ METODY EDUKACYJNEJ WARTOŚCI DODANEJ DLA GIMNAZJÓW

Na wykresie prezentującym dane dotyczące głównej kohorty wiekowej widoczne są pewne odstępstwa od rozkładu równomiernego. W szczególności zauważalny jest niższy odsetek uczniów urodzonych w ostatnich miesiącach roku. Jednak porównanie z danymi GUS pokazuje, że odchylenia te są zbliżone do fluktuacji urodzeń w całych kohortach wiekowych (Dolata i Pokropek, 2012).

Rysunek 2. Procent uczniów urodzonych w poszczególnych miesiącach spośród urodzonych w 1997 r. (młodszych niż główna kohorta)



Rysunek 3. Procent uczniów urodzonych w poszczególnych miesiącach spośród urodzonych w 1995 r. (starszych niż główna kohorta)



Rozkład miesiąca urodzenia wśród uczniów młodszych niż główna kohorta wiekowa pokazuje, że to raczej uczniom urodzonym na początku roku kalendarzowego, czyli uczniom starszym w swojej kohorcie wyznaczonej ze względu na rok urodzenia, przyspiesza się start szkolny. Wśród uczniów starszych niż główna kohorta wiekowa mamy do czynienia z sytuacją odwrotną.

Inteligencja – charakterystyka zmiennych, model analizy i właściwości narzędzi

Plany testów i schemat kotwiczenia

W badaniu zostały wykonane dwa pomiary inteligencji płynnej uczniów. W pierwszym pomiarze wykorzystano test matryc Ravena w wersji standard, forma klasyczna. W drugim serie C – E z testu w wersji standard, oraz pierwsze 24 pozycje z wersji dla zaawansowanych. Miało to na celu dostosowanie trudności testu do poziomu inteligencji badanej populacji. Zadania, które zostały wykorzystane w obu pomiarach potraktowano jako kotwiczące podczas przedstawiania wyników na wspólnej skali, o ile spełniały odpowiednie kryteria (por. niżej). W poniższej tabeli doprecyzowano, które zadania zostały wykorzystane w którym pomiarze oraz które z nich zostały potraktowane jako kotwiczące w modelu zrównującym wyniki.

Tabela 2. Struktura testów inteligencji w obu pomiarach ze wskazaniem zadań kotwiczących

wersja testu matryc Ravena	seria testu i nr zadania	pozycje wykorzystane na I etapie badania	pozycje wykorzystane na III etapie badania	kotwiczące
standard	A3-A12	+		
	B1-B12	+		
	C1	+	+	
	C2-C10	+	+	+
	C11, C12	+	+	
	D1-D12	+	+	+
	E1-E6	+	+	+
	E7	+	+	
	E8	+	+	+
	E9, E10	+	+	
	E11, E12	+	+	+
dla zaawansowanych	1-24		+	

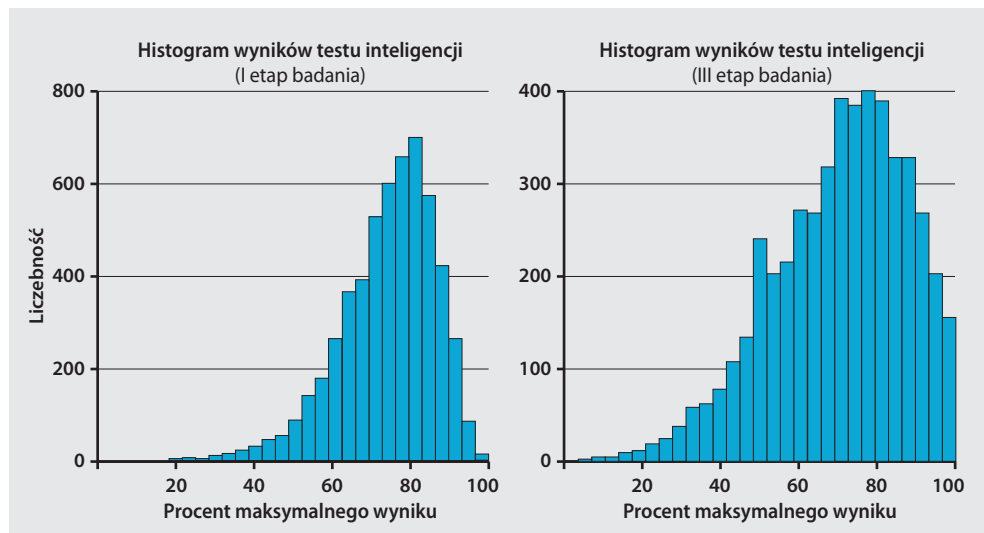
Zadania A1 i A2 z testu w wersji standard, mimo iż zgodnie z instrukcją realizacji badania zostały przedstawione uczniom do rozwiązania, zostały wykluczone z obliczeń ze względu na ich treningowy charakter oraz fakt, że rozwiązywane są razem z osobą prowadzącą badanie.

Zadania, które wykorzystano w obu pomiarach, zostały poddane analizie zróżnicowanego funkcjonowania zadania (*Differential Item Functioning*, DIF) w celu weryfikacji założeń o posiadaniu przez nie tych samych parametrów w obu pomiarach. Tylko zadania, dla których można było uznać, że spełniają te założenia, zostały potraktowane jako zadania kotwiczące umożliwiające zrównanie wyników z obu pomiarów. Do skalowania wyników testu matryc Ravena zdecydowano się wykorzystać dwuparametryczny model IRT. Tak więc w modelu zrównującym istniała potrzeba założenia istnienia w obu pomiarach tej samej wartości parametru trudności i dyskryminacji dla zadań kotwiczących. Dlatego wykonano analizę niejednorodnego efektu DIF, gdzie zakłada się, że wielkość efektu może zmienić się w zależności od poziomu mierzonej testem cechy ukrytej. Analizę DIF przeprowadzono na podstawie testu ilorazu wiarygodności w podejściu IRT (Kondratka i Grudniewska, 2013); wykorzystano niepublikowane oprogramowanie do analiz IRT autorstwa Bartosza Kondratka (Instytut Badań Edukacyjnych). Na podstawie wyników analiz 6 zadań, które okazały się mieć istotny i relatywnie największy efekt DIF, wykluczono z puli zadań kotwiczących, uznając, że będą traktowane tylko jako zadania swoiste dla każdego z pomiarów.

Dostosowanie trudności testów do poziomu inteligencji badanej populacji

Wyniki z I etapu badania pozwoliły przewidzieć problem niedostosowanej trudności testu matryc Ravena w wersji Standard (Jaworowska i Szustrowa, 1991a) do badanej populacji w momencie, gdy uczniowie będą kończyć naukę w gimnazjum. Rozkład wyników testu inteligencji z I etapu badania (testu w wersji standard) charakteryzowała asymetria lewostronna. Oznaczało to, że test jest za łatwy i z mniejszą precyzją pozwala różnicować poziom cechy ukrytej uczniów o wyższym poziomie inteligencji. Przewidywanie wzrostu średniego poziomu inteligencji uczniów biorących udział w badaniu w okresie nauki w gimnazjum, pozwalało przypuszczać, że zastosowanie dokładnie tego samego testu na III etapie badania będzie skutkowało silnym efektem sufitowym i uniemożliwi dobre oszacowanie poziomu inteligencji uczniów najzdolniejszych. Należało jednak zadbać o to, żeby wyniki obu pomiarów można było wyrazić na tej samej skali. Dlatego na III etapie badania zdecydowano się wykorzystać część pytań z testu w wersji dla zaawansowanych (Jaworowska i Szustrowa, 1991b) oraz trzy serie z testu w wersji standard, będące zadaniami kotwiczącymi. Rozkład wyników z III etapu badania pokazuje, że nie udało się uniknąć problemu efektu sufitowego, choć na pewno został złagodzony przez zastąpienie części prostszych zadań z wersji standard trudniejszymi zadaniami z wersji dla zaawansowanych. Niemniej należy mieć na uwadze, że efekt sufitowy wyników z III etapu badania może zaniżać oszacowania przyrostów inteligencji uczniów najzdolniejszych.

Rysunek 4. Rozkłady surowych wyników uczniów w teście inteligencji na I i III etapie badania



Należy zwrócić uwagę na różny zakres skali na osiach pionowych wykresów.

Zrównanie wyników z obu pomiarów oraz oszacowanie poziomu inteligencji uczniów¹

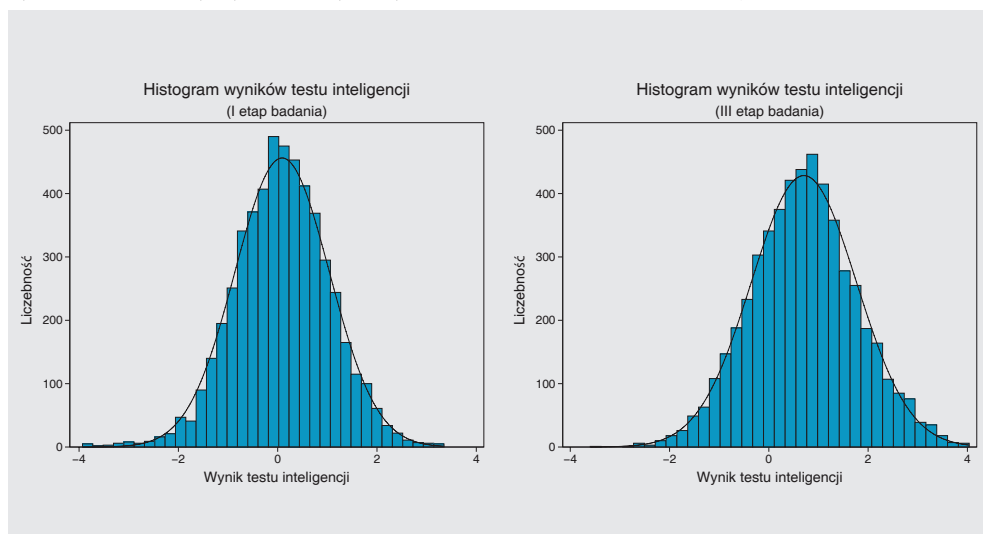
W celu zrównania wyników testu inteligencji z I i III etapu badania przyjęto plan zrównywania nierównoważnych grup z testem kotwiczącym z kotwicą wewnętrzną (patrz np.: Pokropek, Kondrątek, 2012). Modyfikacja klasycznej wersji tego planu polegała na tym, że w przypadku omawianych wyników mamy do czynienia nie z dwiema populacjami osób rozwiązującymi dwa testy, ale z tą samą populacją rozwiązującą testy w pewnym odstępie czasu. Informacja o powtarzalności pomiaru została uwzględniona w przyjętym modelu. Do zrównania wyników wykorzystano metodę kalibracji łącznej, zgodnie z którą w jednym kroku szacowano model IRT dla obu pomiarów. Analizy wykonano w programie Mplus 6, definiując strukturalny model czynnikowy dla zmiennych porządkowych. Wykorzystano estymator WLSMV. Model taki jest równoważny dwuparametrycznemu modelowi IRT (patrz np.: Bartholomew, Knott i Moustaki, 2011). Model ten przyjmował, że poziom inteligencji z I etapu (cecha ukryta przejawiająca się w rozwiązaniach zadań wykorzystanych w teście na I etapie badania) wpływa na poziom inteligencji z III etapu (cechę ukrytą przejawiającą się w rozwiązaniach zadań wykorzystanych w teście na III etapie badania). W modelu zdefiniowano, że zadania kotwiczące (30 zadań) mają w obu pomiarach takie same parametry trudności i dyskryminacji (czyli

¹ Autorki składają podziękowania Tomaszowi Żółtakowi (Instytut Badań Edukacyjnych, Pracownia EWD) za pomoc w specyfikacji modelu strukturalnego.

w terminologii analizy czynnikowej: wartości progowe i ładunki czynnikowe). Model dopuszczał różne wartości średnich i odchyłeń standardowych dla obu pomiarów. Dla pierwszego pomiaru zostały one zafiksowane na wartościach odpowiednio 0 i 1, natomiast dla drugiego zostały uwolnione i wyrażone na skali pierwszego. Dodatkowo uwzględniono pogrupowanie uczniów w klasach. Model uzyskał dobre dopasowanie do danych ($RMSEA = 0,011$, $CFI = 0,954$, $TLI = 0,954$). Nie przedstawiamy rysunku modelu ze względu na jego znaczne rozmiary.

Korelacja latentna między poziomem inteligencji uczniów na I etapie i III etapie badania jako cechami ukrytymi wyniosła 0,71.

Rysunek 5. Rozkłady wyskalowanych wyników uczniów w teście inteligencji na I i III etapie badania



Na podstawie tego modelu wyznaczono wyniki dla uczniów dla obu pomiarów jako wartości czynnikowe dla zmiennych ukrytych. Uczniom, którzy mieli dane tylko z jednego pomiaru, przypisano najbardziej wiarygodny wynik na podstawie rekordu odpowiedzi z drugiego pomiaru i zależności między pomiarami obserwowanymi w danych. Imputacji tej dokonano dla 6,61% wyników dla pierwszego pomiaru i 4,46% wyników dla drugiego pomiaru. Uczniowie z brakami danych dla obu pomiarów, nie mieli imputowanych wyników (uczniów tych w próbie wyznaczonej do dalszych analiz jest 26, co daje 0,49%). Uzyskane wyniki przeliczono na skalę o średniej 0 i odchyleniu standardowym 1 w próbie uczniów z I etapu badania. Wyniki z III etapu badania wyrażone są na skali wyników z pierwszego pomiaru. Tak wyznaczone wyniki są wykorzystywane w analizach jako wskaźniki poziomu inteligencji uczniów. Rozkłady wyskalowanych wyników pokazano na poniższym rysunku. Na wykresy histogramów nałożono krzywe rozkładu normalnego.

Przyrosty poziomu inteligencji

Na podstawie wyników uczniów z dwóch etapów badania policzono dla każdego ucznia przyrost poziomu inteligencji rozumiany jako różnica między wyskalowanymi wynikami drugiego i pierwszego pomiaru. Na potrzeby analiz wyznaczono także wartość dodaną przyrostów inteligencji dla szkół i oddziałów. W poniższej tabeli przedstawiono statystyki opisowe dla tych zmiennych.

Tabela 3. Statystyki opisowe dla wskaźników przyrostów poziomu inteligencji uczniów

	średnia	odchylenie standardowe	minimum	maksimum	skośność
przyrost poziomu inteligencji	0,60	0,64	-3,33	4,64	-0,06
wartość dodana w zakresie przyrostu inteligencji dla szkół	0	0,223	-0,779	0,655	-0,413
wartość dodana w zakresie przyrostu inteligencji dla oddziałów	0	0,238	-0,831	0,681	-0,207

Specyfikacja i wyniki modeli służących wyznaczeniu wskaźników wartości dodanej przyrostów inteligencji i EWD

Tabela 4. Wyniki dwupoziomowego modelu przyrostów inteligencji z uwzględnieniem pogrupowania uczniów na szkoły

zmienna zależna:	poziom inteligencji w III klasie (test matryc Ravena – II pomiar)	
	(1)	(s.e.)
oszacowanie efektów stałych		
poziom ucznia		
test matryc Ravena (I pomiar)	0,858	(0,016)
stała	0,588	(0,028)
oszacowanie efektów losowych		
wariancja efektów szkół	0,060	(0,019)
wariancja na poziomie ucznia	0,354	(0,012)
podsumowanie		
log pseudolikelihood	-155681,74	
liczba uczniów	5222	
liczba szkół	150	

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe;

Tabela 5. Wyniki dwupoziomowego modelu przyrostów inteligencji z uwzględnieniem pogrupowania uczniów na klasy

zmienna zależna:	poziom inteligencji w III klasie (test matryc Ravena – II pomiar)	
	(1)	(s.e.)
oszacowanie efektów stałych		
poziom ucznia		
test matryc Ravena (I pomiar)	0,851	(0,012)
stała	0,600	(0,018)
oszacowanie efektów losowych		
wariancja efektów oddziałów	0,072	(0,010)
wariancja na poziomie ucznia	0,333	(0,012)
podsumowanie		
log pseudolikelihood	-155681,74	
liczba uczniów	5222	
liczba oddziałów	291	

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe;

Tabela 6. Wyniki dwupoziomowych modeli EWD z uwzględnieniem pogrupowania uczniów na szkoły dla zmiennych zależnych: wyniki egzaminu w części matematyczno-przyrodniczej (1) i części humanistycznej (2)

zmienna zależna:	(1) GMP	(s.e.)	(2) GPH	(s.e.)
oszacowanie efektów stałych				
poziom ucznia				
sprawdzian	11,47	(0,262)	11,97	(0,236)
sprawdzian ²	-0,285	(0,165)	-0,407	(0,132)
sprawdzian ³	-0,387	(0,091)	-0,366	(0,078)
pleć ^a	-1,42	(0,316)	1,73	(0,329)
dysleksja na egz. gimnazjalnym ^b	2,43	(0,840)	1,93	(0,700)
dysleksja na sprawdzianie ^b			-3,28	(0,803)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym	-4,21	(0,866)		
stała	102,17	(0,333)	100,91	(0,370)
oszacowanie efektów losowych				
wariancja efektów szkół	5,94	(1,37)	6,82	(1,21)
wariancja na poziomie ucznia	86,06	(2,94)	79,26	(2,56)
podsumowanie				
log pseudolikelihood	-614673,65		-608073,74	
liczba uczniów	5222		5222	
liczba szkół	150		150	

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji

Tabela 7. Wyniki dwupoziomowych modeli EWD z uwzględnieniem pogrupowania uczniów na oddziały dla zmiennych zależnych: wyniki egzaminu w części matematyczno-przyrodniczej (1) i części humanistycznej (2)

zmienna zależna:	(1) GMP	(s.e.)	(2) GPH	(s.e.)
oszacowanie efektów stałych				
poziom ucznia				
sprawdzian	11,42	(0,240)	11,80	(0,215)
sprawdzian ²	-0,326	(0,132)	-0,376	(0,102)
sprawdzian ³	-0,364	(0,080)	-0,346	(0,061)
pleć ^a	-1,34	(0,274)	1,48	(0,292)
dysleksja na egz. gimnazjalnym ^b	1,88	(0,680)	1,49	(0,594)
dysleksja na sprawdzianie ^b			-3,11	(0,674)
interakcja: dysleksja na sprawdzianie i dysleksja na egz. gimnazjalnym	-3,44	(0,846)		
stała	102,17	(0,284)	100,89	(0,298)
oszacowanie efektów losowych				
wariancja efektów oddziałów	7,64	(1,28)	8,46	(1,10)
wariancja na poziomie ucznia	84,24	(2,58)	76,08	(1,95)
podsumowanie				
log pseudolikelihood	-1075355,7		-1061220,9	
liczba uczniów	5222		5222	
liczba oddziałów	291		291	

pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji

Aneks do rozdziału 5

Status społeczny rodziny ucznia

Skalowanie wskaźnika zasobności rodziny ucznia

W modelu bifactor (Holzinger i Swineford, 1937) istnieją dwa rodzaje czynników: pojedynczy czynnik główny oraz czynniki lokalne (specyficzne subczynniki). Czynnik główny jest zdefiniowany przez ładunki wszystkich pozycji tworzących skalę (itemów), podczas gdy subczynniki definiowane są przez grupy pozycji przypisanych do danej podskali. Taka metoda analizy wymaga istnienia założeń odnośnie struktury czynnikowej i jest specjalnym przypadkiem konformacyjnej analizy czynnikowej (*Confirmatory Factor Analysis, CFA*).

W modelu bifactor istnieje założenie o braku korelacji pomiędzy czynnikiem głównym oraz subczynnikami. Cała współzmiennność jest podzielona na ładunki na czynniku głównym lub na subczynnikach. W efekcie czynniki lokalne ujmują wspólną wariancję pozycji pozostałą ponad wariancją ujętą w czynniku głównym czyli wariancję niewyjaśnioną przez ten czynnik (Chen, West i Sousa, 2006). Jeśli ładunki na czynniku głównym oraz na subczynnikach są znaczne ($> 0,3$), to interpretować można zarówno jedno, jak i drugie. Wskazuje to na istnienie generalnego wymiaru skali i jednocześnie pozwala stwierdzić istnienie dobrze zdefiniowanych czynników lokalnych. W konsekwencji sensownym jest przedstawianie i analiza tak uzyskanych wyników zarówno na wymiarze głównym, jak i na wymiarach lokalnych (Brown, Finney i France, 2011; Lai i inni, 2009).

W odróżnieniu od klasycznego modelu bifactor analizowanego w ramach analizy konformacyjnej (CFA) w analizie zamożności zastosowana została stosunkowo nowa metoda analizy, tj. eksploracyjne modele równań strukturalnych (*exploratory structural equation modeling* – ESEM; Asparouhov i Muthén, 2009). Daje ona nowe możliwości analizy struktury czynnikowej skal pomiarowych (Marsh, Liem, Martin, Morin i Nagengast, 2011; Herbert, Marsh i inni, 2009; 2010; 2011; Herbert, Marsh, Nagengast i Morin, 2013; Mattsson, 2012; Morin i Maïano, 2011; Myers, Chase, Pierce i Martin, 2011; Sánchez-Carracedo i in., 2012). Jej podstawową zaletą jest duża elastyczność w estymowaniu ładunków czynnikowych. W kontekście analizy modeli bifactor daje możliwość zbudowania modeli, w których poszczególne pozycje skali mogą łączyć więcej niż jeden subczynnik. W efekcie uzyskać można nie tylko lepsze dopasowanie modelu do danych, lecz, co istotne, dobre odzwierciedlenie struktury czynnikowej.

Konstrukcja pomiarowego modelu bifactor w oparciu o ESEM przeprowadzona została z wykorzystaniem programu MPlus 6.12.

Do analizy poziomu zamożności gospodarstwa domowego ucznia zaadaptowana została skala wyposażenia gospodarstwa domowego wykorzystywana w badaniach PISA. W polskiej wersji składa się ona z 17 pytań dychotomicznych, w których respondenci pytani są o posiadanie następujących elementów wyposażenia: biurka do nauki dla dziecka; własnego pokój dla dziecka; spokojnego miejsca do nauki dla dziecka; komputera, który dziecko może używać do nauki; edukacyjnych programów komputerowych; dostępu do Internetu; kalkulatora; literatury pięknej (np. Sienkiewicz); zbiorów poezji; dzieł sztuki (np. obrazy); encyklopedii; słowników ; zmywarki do naczyń; DVD albo wideo; anteny satelitarnej lub telewizji kablowej; aparatu cyfrowego; mikroskopu lub lunety. Do skali tej wliczane jest również pytanie o liczbę książek w domu.

Ze względu na problemy z estymacją modelu pomiarowego tak poprzez klasyczną confirmacyjną analizę czynnikową (CFA) jak i z wykorzystaniem eksploracyjnych modeli równań strukturalnych (ESEM) z oryginalnej skali usunięta została pozycja 4 „dostęp do komputera, który dziecko może używać do nauki”.

Estymacja modelu odzwierciedlającego podejście wykorzystywane w analizach danych PISA (OECD, 2009) z wykorzystaniem CFA napotkała dwa problemy. Pierwszy to znaczne skorelowanie zmiennych latentnych (wartości od 0,64 do 0,90). Drugi to wysoka wartość ważonego pierwiastka średnich reszt (*weighted root-mean-square residual* – WRMR, Yu, 2002) – 2,61 – wskazującego na niedopasowanie modelu do danych.

Estymacja modelu bifactor w ramach podejścia ESEM pozwoliła na uzyskanie modelu dobrze dopasowanego do danych: RMSEA (*Root Mean Square Error Of Approximation*) 0,021, CFI – 0,989, TLI – 0,980, WRMR – 1,06. Ładunki czynnikowe dla modelu przedstawia tabela 1. Poza czynnikiem głównym mierzącym zasobność gospodarstwa domowego rodziny ucznia wyróżnione zostały trzy czynniki lokalne: miejsce do nauki, sprzęt elektroniczny oraz dobra kulturowe. Przy interpretacji znaczenia subczynników należy pamiętać, iż ich charakterystyka opiera się na tej części współzmienności ładujących je pozycji skali, która nie jest częścią czynnika głównego. Czynniki lokalne w takim modelu charakteryzują się niewielkim skorelowaniem (w granicach od 0,11 do 0,39).

W dalszych analizach, jako zmienne opisujące ogólną zasobność gospodarstwa domowego oraz różnego rodzaju jej przejawy uchwycone poprzez czynniki lokalne, wykorzystywane będą wartości oszacowane w ww. modelu.

Tabela 1. Ładunki czynnikowe w modelu bifactor estymowanego metodą ESEM dla zasobności gospodarstwa domowego rodzin uczniów

zmienna	zasobność	(s.e.)	miejsce	(s.e.)	sprzęt	(s.e.)	kultura	(s.e.)
biurko do nauki dla dziecka	0,460	0,045	0,558	0,058	0,128	0,052	0,018	0,062
własny pokój dla dziecka	0,268	0,028	0,822	0,057	-0,016	0,035	-0,079	0,039
spokojne miejsce do nauki dla dziecka	0,322	0,041	0,760	0,071	0,010	0,046	0,153	0,046
edukacyjne programy komputerowe	0,522	0,025	0,083	0,034	0,395	0,041	0,156	0,042
dostęp do internetu	0,507	0,028	0,022	0,030	0,636	0,034	0,055	0,036
kalkulator	0,490	0,049	0,023	0,071	0,268	0,078	0,419	0,083
literatura piękna, np. Sienkiewicz	0,887	0,013	-0,024	0,035	-0,018	0,036	0,215	0,035
zbiory poezji	0,885	0,012	0,035	0,037	-0,118	0,036	-0,031	0,037
dzieła sztuki, np. obrazy	0,499	0,019	0,151	0,040	0,019	0,043	-0,329	0,041
encyklopedie	0,649	0,025	0,094	0,042	0,134	0,044	0,283	0,045
słowniki	0,541	0,039	0,105	0,037	0,049	0,031	0,697	0,066
zmywarka do naczyń	0,430	0,026	0,157	0,038	0,410	0,042	-0,442	0,046
DVD albo video	0,265	0,031	0,002	0,041	0,489	0,041	0,092	0,048
antena satelitarna lub telewizja kablowa	0,244	0,028	0,040	0,035	0,482	0,031	0,032	0,038
aparat cyfrowy	0,499	0,024	0,064	0,030	0,544	0,031	-0,011	0,035
mikroskop lub luneta	0,511	0,021	0,058	0,042	0,140	0,041	-0,166	0,042
liczba książek w domu	0,765	0,011	-	-	-	-	-	-

pogrubieniem oznaczono ładunki istotne statystycznie na poziomie 0,05. Na niebiesko zaznaczono ładunki czynnikowe, które substancjalnie przynależą do czynnika głównego (wszystkie) oraz poszczególnych subczynników

Aneks do rozdziału 6

Korepetycje i rodzinne wsparcie edukacyjne

Skala motywacji ASRS

Narzędzie jest wstępną adaptacją holenderskiego testu motywacji autorstwa Vansteenkiste i współpracowników (2009). Stanowi ona modyfikację narzędzia Ryana i Connella (1989), lecz stosuje się go do badania uczniów starszych (od 7 klasy) i studentów¹. Oryginalna wersja testu składa się z 16 pozycji podzielonych na 4 równoliczne podskale mierzące 4 typy motywacji: motywację zewnętrzną (*extrinsic regulation*), introjekcyjną (*introjected*), identyfikacyjną (*identified*) i wewnętrzną (*intrinsic*) w rozumieniu teorii autodeterminacji (por. np. Ryan i Deci, 2000). Zachowania motywowane zewnętrznie są podejmowane po to, by spełnić zewnętrzne wobec jednostki wymagania, uzyskać konkretną korzyść czy uniknąć sankcji. Towarzyszy im poczucie przymusu, braku wyboru. Ten rodzaj motywacji jest bliski tradycyjnemu ujęciu motywacji zewnętrznej. Motywacja introjekcyjna dotyczy podejmowania działań w celu uniknięcia emocji i uczuć o negatywnym zabarwieniu (lęk, poczucie winy) lub doświadczania uczuć o zabarwieniu pozytywnym (np. duma, poczucie bycia wartościową osobą). Wiąże się z odczuwaniem pewnej presji przymusu, gdyż działanie nie jest w pełni odczuwane jako własny wybór. Nieco bardziej autonomiczne, czyli odczuwane jako wynikające z własnego wyboru, są działania motywowane identyfikacyjnie. Wynikają one z przekonania o ważności, istotności danego działania z perspektywy osobistych celów i planów i dlatego są podejmowane. Motywacja wewnętrzna dotyczy podejmowania działania dla niego samego, bez konkretnego instrumentalnego celu stojącego za aktywnością. Jednostka wykazuje nią zainteresowanie, czerpie z niej satysfakcję, jest w nie zaangażowana (Ryan i Deci, 2000).

Każda pozycja testu ma formę stwierdzenia dotyczącego przyczyn uczenia się szkolnego. Respondent na pięciostopniowej skali (zupełnie nieważne/raczej nieważne/ani ważne ani nieważne/raczej ważne/bardzo ważne) odpowiada, na ile każdy z tych powodów jest dla niego ważny, tj. odzwierciedla to, dlaczego uczy się. Prace adaptacyjne zostały przeprowadzone przez zespół EWD. W pierwszym kroku przeanalizowano podstawy teoretyczne narzędzia. Następnie dwóch psychologów dokonało niezależnych przekładów testu na język polski. Tłumaczenia zostały porównane, a różnice między nimi – przedyskutowane. Na tej podstawie przygotowana została druga wersja narzędzia, którą zwrotnie przełożył na język angielski psycholog biegły posługujący się językiem angielskim

¹ Autor nie podaje dokładnego przedziału wiekowego.

(mieszkający na stałe w Wielkiej Brytanii). Porównanie obu anglojęzycznych wersji testów (oryginalnego i przełożonego zwrotnie) nie wykazało istotnych różnic znaczeniowych. Tak przygotowane narzędzie wykorzystano w badaniach pilotażowych zrealizowanych na incydentalnej próbie uczniów gimnazjów ($N = 85$). Test okazał się mieć dobre właściwości psychometryczne, stąd wykorzystano go w badaniu panelowym.

W ramach weryfikacji właściwości narzędzia w badaniu panelowym przeprowadzona została konfirmacyjna analiza czynnikowa (CFA) dla zmiennych porządkowych (por. Konarski, 2009). Obliczenia przeprowadzono za pomocą oprogramowania MPlus 6.1 (Muthén i Muthén, 1998–2010), z wykorzystaniem estymatora WLSMV. W modelu uwzględniono zgrupowanie uczniów w szkołach. W analizach wykorzystane zostały dane wszystkich uczniów, którzy wypełnili kwestionariusz i jednocześnie zanotowano dla nich nie więcej niż jeden brak danych na pozycjach wchodzących w skład każdej podskali (łącznie 5021 osób)². Podczas samych analiz braki danych usuwano parami³. W pierwotnej wersji modelu pozycja 1 okazała się łądować na każdą z podskal, w związku z czym została usunięta z analiz. Zmodyfikowany model osiągnął on akceptowalne dopasowanie do danych (RMSEA = 0,078, TLI = 0,965; tabela 1).

Tabela 1. Skala Motywacji ASRS – miary dopasowania ostatecznego modelu konfirmacyjnej analizy czynnikowej (CFA) dla zmiennych porządkowych

miara dopasowania	wartość
chi ²	2672,31 (df = 85, $p < 0,001$)
RMSEA	0,078 (PU: 0,075–0,08)
CFI	0,971
TLI	0,965

Standaryzowane ładunki czynnikowe poszczególnych pozycji skali osiągnęły wysokie wartości, wahając się w przedziale od 0,624 do 0,899. Wartości 0,8 lub więcej osiągnęło 13 z nich. Szczegółowe dane znajdują się w tabeli 2.

² Z powodu braków danych wykluczono 50 obserwacji.

³ Dla każdej pary pozycji zanotowano nie więcej niż 1,6% braków danych.

Tabela 2. Skala Motywacji ASRS – miary dopasowania ostatecznego modelu konfirmacyjnej analizy czynnikowej (CFA) dla zmiennych porządkowych

podskala	pozycja	ładunek czynnikowy	
		standaryzowany	niestandaryzowany
motywacja zewnętrzna	ASRS2	0,624	1
	ASRS3	0,850	1,36
	ASRS4	0,813	1,30
motywacja introjekcyjna	ASRS5	0,660	1
	ASRS6	0,822	1,24
	ASRS7	0,841	1,27
	ASRS8	0,800	1,21
motywacja identyfikacyjna	ASRS9	0,802	1
	ASRS10	0,889	1,11
	ASRS11	0,864	1,08
	ASRS12	0,849	1,06
motywacja wewnętrzna	ASRS13	0,867	1
	ASRS14	0,899	1,04
	ASRS15	0,842	0,97
	ASRS16	0,842	0,97

Wszystkie ładunki są istotne statystycznie.

Poszczególne wymiary (typy) motywacji korelują ze sobą, przy czym wzorec korelacji jest dość klarowny. Im dwa typy motywacji leżą dalej od siebie na wymiarze autonomii, tym słabiej są skorelowane. Motywacja wewnętrzna jest silniej skorelowana z motywacją identyfikacyjną niż z introjekcyjną. Ponadto motywacja identyfikacyjna z introjekcyjną silniej niż motywacja identyfikacyjna z zewnętrzną. Dane znajdują się w tabeli 3.

Tabela 3. Skala Motywacji ASRS – korelacje latentne między podskalami testu

	wewnętrzna	identyfikacyjna	Introjekcyjna	zewnętrzna
wewnętrzna	1	0,667*	0,471*	0 ^a
identyfikacyjna		1	0,602*	0,056*
introjekcyjna			1	0,373*
zewnętrzna				1

* istotne statystycznie na poziomie $p < 0,001$

^a w modelu nałożono ograniczenie w postaci braku zależności między podskalami motywacji zewnętrznej i wewnętrznej

Na podstawie tak wyspecyfikowanego modelu wygenerowano wartości czynnikowe dla poszczególnych obserwacji, które następnie wystandaryzowano (wyrażono w jednostkach odchylenia standardowego). Statystyki opisowe dla poszczególnych podskal testu znajdują się w tabeli 4.

Tabela 4. Wyniki Testu Motywacji ASRS – statystyki opisowe (N = 5021)

podskala ASRS	średnia	mediana	odch. std.	min	max	kurtoza	skośność
motywacja zewnętrzna	0	0,042	1	-2,56	2,65	-0,134	-0,064
motywacja introlekcyjna	0	0,012	1	-2,79	2,61	-0,131	0,242
motywacja identyfikacyjna	0	0,018	1	-2,82	2,33	-0,222	0,053
motywacja wewnętrzna	0	0,122	1	-2,32	2,61	-0,003	-0,152

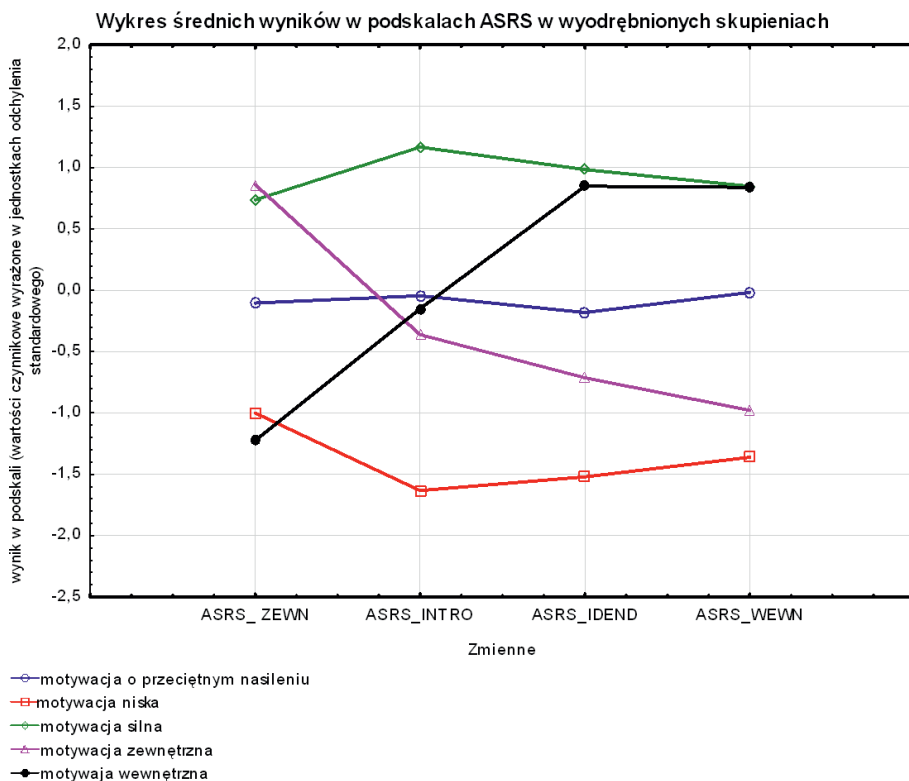
Tak przygotowane wartości czynnikowe dla poszczególnych obserwacji zostały wykorzystane w ramach analizy skupień. W pierwszym kroku przeprowadzono hierarchiczną analizę skupień metodą Warda (por. np. Stanisław, 2007) w celu określenia optymalnej liczby skupień. Następnie dokonano klasyfikacji metodą *k*-średnich (por. np. Stanisław, 2007) z opcją wyłonienia 5 skupień. Profile wyłonionych skupień znajdują się na rysunku 1.

Wyłonione skupienia mają klarowną strukturę. Widoczne są trzy skupienia odpowiednio: niskim, przeciętnym i wysokim nasileniu wszystkich typów motywacji. Ponadto wyodrębnione zostało skupienie o niskim poziomie motywacji zewnętrznej, przeciętnym – introlekcyjnej oraz wysokim identyfikacyjnej i wewnętrznej, które zinterpretowano jako skupienie uczniów o motywacji wewnętrznej. Piąte skupienie charakteryzuje się wysokim poziomem motywacji zewnętrznej, przeciętnym – introlekcyjnej oraz niskim motywacji identyfikacyjnej i wewnętrznej. Zinterpretowane zostało jako skupienie uczniów o motywacji zewnętrznej. Informacje na temat liczebności skupień znajdują się w tabeli 5.

Tabela 5. Liczebności wyłonionych skupień

skupienie	liczebność	liczebność skumulowana	%
motywacja o przeciętnym nasileniu	1826	1826	36,37
motywacja niska	521	2347	10,38
motywacja silna	1145	3492	22,80
motywacja zewnętrzna	833	4325	16,59
motywacja wewnętrzna	696	5021	13,86

Rysunek 1. Profile skupień (średnie wyniki w podskalach testu ASRS w poszczególnych skupieniach).



ASRS_ZEWN – podskala „motywacja zewnętrzna”

ASRS_INTRO – podskala „motywacja introlekcyjna”

ASRS_IDENT – podskala „motywacja identyfikacyjna”

ASRS_WEWN – podskala „motywacja wewnętrzna”

Wyniki w podskalach to wartości czynnikowe wygenerowane na podstawie modelu konfirmacyjnej analizy czynnikowej Skali ASRS (wyrażone w jednostkach odchylenia standardowego).

Skale bezradności intelektualnej

Narzędzia służą do pomiaru poziomu bezradności intelektualnej rozumianej jako utrwalony stan, „którego istotą jest „(a) demobilizacja poznawcza, tj. zahamowanie tendencji do aktywnego myślenia problemowego (...), oraz (b) deficyt rozumienia (jakościowe pogorszenie zdolności do asymilowania nowych informacji do już istniejących struktur wiedzy)” (Sędek, 1995, s. 25–26). Stan ten powstaje w wyniku tzw. poznawczego treningu bezradności, czyli długotrwałego wysiłku poznawczego, który nie prowadzi do zrozumienia materiału omawianego na lekcji (lub poza lekcjami, np. w domu, podczas korepetycji). Jeżeli stan ten powtarza się, ma tendencję do stabilizowania i utrwalania się. Jest on specyficzny do nauczanego przedmiotu (tj. uczeń może doświadczać bezradności intelektualnej w zakresie matematyki, lecz nie

mieć żadnych problemów z innymi przedmiotami) i towarzyszyć mu może utrata wewnętrznej motywacji do nauki. Jego źródeł należy upatrywać w wadliwych sposobach nauczania, zaburzeniach relacji między nauczycielem a uczniem oraz niewłaściwych sposobach uczenia się (Sędek, 1995). Oba testy składają się z 20 stwierdzeń dotyczących doświadczeń na lekcjach z danego przedmiotu oraz w z zetknięciu z wiedzą z danej dziedziny (typowych dla stanu bezradności intelektualnej lub jego braku). Uczeń miał za zadanie ustosunkować się do nich, zaznaczając na pięciopunktowej skali (nigdy/rzadko/czasem/często/zawsze), jak często ich doświadcza. Narzędzie zostało dostarczone przez Interdyscyplinarne Centrum Badań i Rozwoju Organizacji działające przy Zakładzie Psychologii Społecznej Instytutu Psychologii UJ.

W zakresie języka polskiego

Przeprowadzono konfirmacyjną analizę czynnikową (CFA) dla zmiennych porządkowych (por. Konarski, 2009). Obliczenia przeprowadzono za pomocą oprogramowania MPlus 6.1 (Muthén i Muthén, 1998–2010), z wykorzystaniem estymatora WLSMV. W modelu uwzględniono zgrupowanie uczniów w szkołach oraz zezwolono na skorelowanie błędów. W analizach wykorzystano dane wszystkich uczniów, którzy wypełnili kwestionariusz i zanotowano dla nich nie więcej niż 4 braki danych na pojedynczych pozycjach⁴ (łącznie 5045 osób). W samej analizie braki danych usuwano parami.⁵ Testowany model osiągnął dobre dopasowanie do danych (RMSEA = 0,048, TLI = 0,974; tabela 6).

Tabela 6. Skala bezradności intelektualnej (język polski) – miary dopasowania modelu konfirmacyjnej analizy czynnikowej (CFA) dla zmiennych porządkowych

miara dopasowania	wartość
chi ²	1024,69 (df = 81, p < 0,001)
RMSEA	0,048 (0,045-0,051)
CFI	0,989
TLI	0,974

PU – przedział ufności

W modelu zezwolono na skorelowanie błędów.

Standaryzowane ładunki czynnikowe poszczególnych pozycji skali okazały się akceptowalne. Ich wartości wahały się w przedziale od 0,392 do 0,809, przy czym wartości powyżej 0,5 osiągnęło 16 z nich. Szczegółowe dane znajdują się w tabeli 7.

⁴ Z powodu braków danych wykluczono 36 obserwacji.

⁵ Dla każdej pary zmiennych zanotowano nie więcej niż 5,1% braków danych.

Tabela 7. Skala bezradności intelektualnej (język polski) – wartości ładunków czynnikowych w konfirmacyjnej analizie czynnikowej (CFA dla zmiennych porządkowych)

pozycja	ładunek czynnikowy	
	standaryzowany	niestandaryzowany
SBJP1	0,546	1,000
SBJP2	0,527	0,966
SBJP3	0,597	1,095
SBJP4	0,530	0,972
SBJP5	0,425	0,780
SBJP6	0,740	1,357
SBJP7	0,567	1,039
SBJP8	0,661	1,212
SBJP9	0,712	1,305
SBJP10	0,654	1,199
SBJP11	0,445	0,816
SBJP12	0,392	0,718
SBJP13	0,799	1,464
SBJP14	0,418	0,766
SBJP15	0,795	1,457
SBJP16	0,770	1,411
SBJP17	0,806	1,478
SBJP18	0,812	1,488
SBJP19	0,780	1,430
SBJP20	0,562	1,031

Wszystkie ładunki są istotne statystycznie.

Na podstawie tak wyspecyfikowanego modelu wygenerowano wartości czynnikowe dla poszczególnych obserwacji, które następnie wystandaryzowano (wyrażono w jednostkach odchylenia standardowego). Statystyki opisowe obliczone dla tak przygotowanego wskaźnika znajdują się w tabeli 8.

Tabela 8. Wyniki Skali bezradności intelektualnej (język polski) – statystyki opisowe (N = 5045)

test	średnia	mediana	odch. std.	min	max	skośność	kurtosa	kwartył 1	kwartył 3
SBI (język polski)	0	-0,042	1	-3,89	3,05	0,155	0,600	-0,708	0,632

W zakresie matematyki

Przeprowadzona została confirmacyjna analiza czynnikowa (CFA) dla zmiennych porządkowych (por. Konarski, 2009). Obliczenia przeprowadzono za pomocą oprogramowania MPlus 6.1 (Muthén i Muthén, 1998–2010), z wykorzystaniem estymatora WLSMV. W modelu uwzględniono zgrupowanie uczniów w szkołach oraz zezwolono na skorelowanie błędów. W analizach wykorzystano dane wszystkich uczniów, którzy wypełnili kwestionariusz i zanotowano dla nich nie więcej niż 4 braki danych na pojedynczych pozycjach⁶ (łącznie 5043 osoby). Podczas samej analizy braki danych usuwane były parami⁷. Testowany model osiągnął dobre dopasowanie do danych (RMSEA = 0,05, TLI = 0,971; tabela 9).

Tabela 9. Skala bezradności intelektualnej (matematyka) – miary dopasowania modelu confirmacyjnej analizy czynnikowej (CFA) dla zmiennych porządkowych

miara dopasowania	wartość
chi ²	1359,72 (df = 101, p < 0,001)
RMSEA	0,050 (PU: 0,047-0,052)
CFI	0,984
TLI	0,971

PU – przedział ufności

W modelu zezwolono na skorelowanie błędów.

Standaryzowane ładunki czynnikowe poszczególnych pozycji skali okazały się akceptowalne. Ich wartości wahały się w przedziale od 0,294 do 0,809, przy czym wartości powyżej 0,5 osiągnęło 16 z nich. Szczegółowe dane znajdują się w tabeli 10.

Na podstawie tak wyspecyfikowanego modelu wygenerowane zostały wartości czynnikowe dla poszczególnych obserwacji, które następnie wystandaryzowano (wyrażono w jednostkach odchylenia standardowego). Statystyki opisowe obliczone dla tak przygotowanego wskaźnika znajdują się w tabeli 11.

⁶ Z powodu braków danych wykluczono 38 obserwacji.

⁷ Dla każdej pary pozycji zanotowano nie więcej niż 6,2% braków danych.

Tabela 10. Skala bezradności intelektualnej (matematyka) – wartości ładunków czynnikowych w konfirmacyjnej analizie czynnikowej (CFA) dla zmiennych porządkowych

pozycja	ładunek czynnikowy	
	standaryzowany	niestandaryzowany
SBM1	0,615	1
SBM2	0,562	0,914
SBM3	0,616	1,00
SBM4	0,581	0,945
SBM5	0,458	0,745
SBM6	0,766	1,24
SBM7	0,635	1,03
SBM8	0,575	0,934
SBM9	0,724	1,18
SBM10	0,626	1,02
SBM11	0,343	0,558
SBM12	0,294	0,478
SBM13	0,776	1,26
SBM14	0,403	0,654
SBM15	0,774	1,26
SBM16	0,795	1,29
SBM17	0,794	1,29
SBM18	0,809	1,31
SBM19	0,728	1,18
SBM20	0,574	0,934

Wszystkie ładunki są istotne statystycznie.

Tabela 11. Wyniki Skali Bezradności Intelektualnej (matematyka) – statystyki opisowe (N = 5043)

test	średnia	mediana	odch. std.	min	max	skośność	kurtosis	kwartył 1	kwartył 3
SBI (matematyka)	0	-0,055	1	-3,74	3,26	0,166	0,558	-0,618	0,609

Korelaty uczęszczania na korepetycje z przedmiotów humanistycznych

W tabeli 12 przedstawiono wyniki analizy korelatów uczęszczania na korepetycje z przedmiotów humanistycznych. Przypominamy, że z powodu niskiego odsetka uczniów uczęszczających na korepetycje z przedmiotów humanistycznych, oszacowanie modeli z tą zmienną wyjaśnianą okazało się problematyczne. Część modeli nie osiągnęła zbieżności (model pusty, model z predyktorami z poziomu szkół), stąd nie są one prezentowane.

Bezpłatne wsparcie w nauce a wyniki egzaminacyjne i EWD

W tabeli 13 zaprezentowano W wyniki analiz, w których zmienną zależną jest łączny wynik egzaminu gimnazjalnego w części humanistycznej, wyrażony na skali o średniej 100 i odchyleniu standardowym 15 w populacji uczniów, którzy przystąpili do egzaminu. Analizy te miały przede wszystkim na celu znalezienie odpowiedzi na pytanie o to, jak kontrola w modelach EWD zmiennych charakteryzujących ucznia pozwala zmodyfikować zależność wyników egzaminacyjnych od ilości otrzymywanego bezpłatnego wsparcia w nauce z tych przedmiotów.

W tabeli 14 znajdują się modele uwzględniające dodatkowe zmienne kontrolowane.

Tabela 12. Korelaty uczęszczania na korepetycje z przedmiotów humanistycznych. Wyniki trzypoziomowych analiz regresji logistycznych, modele z losową stałą, odporne (robust) oszacowania błędów standardowych

zmienna zależna: uczęszczanie na korepetycje z przedmiotów humanistycznych						
	(1)	(s.e.)	iloraz szans	(2)	(s.e.)	iloraz szans
oszacowanie efektów stałych						
poziom ucznia						
pleć ^a	-0,373	(0,170)	0,688	-0,439	(0,163)	0,644
sprawdzian	-0,495	(0,134)	0,610	-0,549	(0,098)	0,577
test matryc Ravena	-0,142	(0,100)	0,867			
bezradność intelektualna (pol)	0,129	(0,085)	1,137			
motywacja niska ^b	0,370	(0,283)	1,448			
motywacja wysoka ^b	0,358	(0,224)	1,431			
motywacja zewnętrzna ^b	0,162	(0,257)	1,176			
motywacja wewnętrzna ^b	0,275	(0,282)	1,317			
aspiracje rodziców	-0,002	(0,047)	0,998			
zasobność gosp. domowego	0,153	(0,092)	1,165	0,208	(0,096)	1,23
wykształcenie rodziców	0,054	(0,042)	1,05			
HSIOPS	0,002	(0,010)	1,003			
stała	-4,01	(0,800)	0,018	-3,05	(0,110)	0,047
				-4,13	(0,526)	0,016
oszacowanie efektów losowych						
wariancja efektów szkół	0,122			0,126		
wariancja efektów oddziałów	0,244			0,285		

liczba uczniów: 4114, liczba oddziałów: 291, liczba szkół: 150
pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie o profilu motywacji o przeciętnym nasileniu; c – grupa odniesienia: wieś

Tabela 13. Znaczenie bezpłatnej pomocy w nauce dla wyników egzaminu gimnazjalnego w części humanistycznej i EWD. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (*robust*) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części humanistycznej						
	(1)	(s.e.)	(2)	(s.e.)	(3)	(s.e.)
oszacowanie efektów stałych						
poziom ucznia						
bezpłatne wsparcie z hum.			-2,62	(0,196)	-0,717	(0,112)
sprawdzian					11,31	(0,243)
sprawdzian ²					-0,135	(0,132)
sprawdzian ³					-0,353	(0,084)
pleć ^a					2,03	(0,332)
dysleksja na egz. gimnazjalnym ^b					2,04	(0,607)
dysleksja na sprawdzianie ^b					-2,67	(0,768)
stała	100,30	(0,478)	96,66	(0,470)	98,91	(0,375)
90,65				(1,24)	97,48	(0,839)
poziom oddziałów						
średnia z bezpłatnego wsparcia z hum.					-7,07	(0,951)
-1,76				(0,534)	-1,21	(0,543)
oszacowanie efektów losowych						
wariancja efektów szkół	17,15		15,62		4,87	
13,47					4,78	
wariancja efektów oddziałów	16,74		13,95		2,92	
11,06					2,62	
wariancja na poziomie ucznia	162,58		153,72		72,14	
162,53					72,79	
podsumowanie						
deviance	33565,05		33312,36		30036,11	
30030,30					3062,51	
liczba szacowanych parametrów	4		5		11	
11					11	
liczba uczniów: 4185, liczba oddziałów: 291, liczba szkół: 150					5	
30030,30					12	

liczba uczniów: 4185, liczba oddziałów: 291, liczba szkół: 150
pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe; a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji

Tabela 14. Znaczenie bezpłatnej pomocy w nauce dla wyników egzaminu gimnazjalnego w części humanistycznej z uwzględnieniem dodatkowych zmiennych kontrolowanych. Wyniki trzypoziomowych analiz regresji, modele z losową stałą (7–11) i losowym nachyleniem (12), odporne (*robust*) oszacowania błędów standardowych

zmienna zależna: wynik egzaminu gimnazjalnego w części humanistycznej												
	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)
oszacowanie efektów stałych												
poziom ucznia												
bezpłatne wsparcie z hum.	-0,634	(0,110)	-0,646	(0,105)	-0,706	(0,103)	-0,749	(0,099)	-0,744	(0,098)	-0,743	(0,098)
sprawdzian	10,14	(0,266)	9,83	(0,271)	9,80	(0,269)	8,68	(0,260)	8,70	(0,258)	8,70	(0,258)
sprawdzian ²	-0,146	(0,135)	-0,086	(0,133)	-0,085	(0,133)	-0,020	(0,127)	-0,014	(0,128)	-0,014	(0,128)
sprawdzian ³	-0,347	(0,080)	-0,339	(0,073)	-0,340	(0,073)	-0,304	(0,073)	-0,306	(0,073)	-0,307	(0,073)
pleć ^a	1,73	(0,310)	0,864	(0,304)	0,610	(0,323)	0,421	(0,303)	0,428	(0,304)	0,428	(0,304)
dysleksja na egz. gimnazjalnym ^b	1,76	(0,580)	1,72	(0,610)	1,71	(0,590)	1,09	(0,661)	1,09	(0,658)	1,09	(0,658)
dysleksja na sprawdzianie ^b	-2,33	(0,741)	-2,39	(0,760)	-2,37	(0,738)	-1,80	(0,779)	-1,75	(0,778)	-1,76	(0,777)
test matryc Ravena (I pomiar)	2,12	(0,203)	2,12	(0,185)	2,09	(0,180)	1,84	(0,179)	1,84	(0,178)	1,84	(0,178)
bezzadność intelektualna (pol)			-2,01	(0,149)	-1,86	(0,150)	-1,68	(0,145)	-1,67	(0,145)	-1,67	(0,144)
motywacja niska ^c					-0,767	(0,641)	-0,773	(0,617)	-0,722	(0,616)	-0,721	(0,616)
motywacja wysoka ^c					1,00	(0,484)	0,744	(0,457)	0,731	(0,453)	0,724	(0,453)
motywacja zewnętrzna ^c					-0,675	(0,450)	-0,699	(0,441)	-0,683	(0,439)	-0,683	(0,438)
motywacja wewnętrzna ^c					1,24	(0,445)	0,86	(0,428)	0,85	(0,430)	0,850	(0,430)
aspiracje rodziców							0,640	(0,076)	0,645	(0,076)	0,644	(0,076)
zasobność gosp. domowego							0,780	(0,174)	0,801	(0,174)	0,801	(0,174)
wykształcenie rodziców							0,279	(0,081)	0,293	(0,081)	0,293	(0,081)
HSIOPS							0,029	(0,017)	0,030	(0,017)	0,030	(0,017)
stała	97,86	(0,818)	97,87	(0,820)	97,62	(0,900)	83,30	(1,48)	83,70	(1,50)	83,73	(1,50)

zmienna zależna: wynik egzaminu gimnazjalnego w części humanistycznej												
	(7)	(s.e.)	(8)	(s.e.)	(9)	(s.e.)	(10)	(s.e.)	(11)	(s.e.)	(12)	(s.e.)
poziom oddziałów												
średnia z bezpłatnego wsparcia z hum.	-0,886	(0,533)	-1,06	(0,517)	-1,12	(0,521)	-0,905	(0,491)	-0,834	(0,496)	-0,966	(0,564)
poziom szkół												
lok.: miasto do 20 tys. ^d									-0,969	(0,565)	-0,976	(0,568)
lok.: miasto od 20 do 100 tys. ^d									-1,80	(0,640)	-1,80	(0,638)
lok.: miasto powyżej 100 tys. ^d									-1,12	(0,749)	-1,13	(0,748)
oszacowanie efektów losowych												
wariancja efektów szkół	4,68		5,05		5,00		4,72		4,31		4,29	
wariancja efektów oddziałów	2,43		1,98		1,89		1,47		1,42		1,13	
wariancja na poziomie ucznia	70,03		66,70		66,30		62,26		62,28		62,26	
poziom oddziałów												
losowe nachylenie dla korepetycje z hum.											0,022	
podsumowanie												
deviance	29899,57		29697,01		29668,081		29396,24		29387,91		29387,80	
liczba szacowanych parametrów	13		14		18		22		25		27	

liczba uczniów: 4186, liczba oddziałów: 291, liczba szkół: 150
pogrubionym drukiem opisywane są wartości istotne na poziomie istotności $p < 0,05$, w nawiasach podano błędy standardowe a – grupa odniesienia: chłopcy; b – grupa odniesienia: uczniowie bez dysleksji; c – grupa odniesienia: uczniowie o profilu motywacji o przeciętnym nasileniu; d – grupa odniesienia: wieś

Aneks do rozdziału 7

Społeczno-gospodarcze charakterystyki gmin

Konstrukcja wskaźnika ubóstwa na poziomie gminy

W celu przygotowania syntetycznego wskaźnika ubóstwa na poziomie gmin wykorzystane zostały następujące informacje: udział osób w gospodarstwach domowych korzystających ze środowiskowej pomocy społecznej w ludności ogółem w 2010 roku, wpływy z PIT na osobę w gminie w 2010 roku, wpływy z CIT na osobę w gminie w 2010 roku oraz świadczenia na rzecz osób fizycznych na osobę w 2010. Na danych tych została przeprowadzona analiza głównych składowych.

Test Kaisera-Mayera-Olkina przyjął wartości 0,630, co świadczy o dobrym doborze zmiennych do analizy. Na podstawie kryterium wartości własne, większej od 1 przyjęto do wyodrębnienia 1 główną składową wyjaśniającą 68% wariancji. Największy ładunek czynnikowy (0,884) ma zmienna opisująca średnie wpływy na osobę z podatku od osób fizycznych. Niewiele mniejszy ładunek mają zmienne opisujące wielkość świadczeń na rzecz osób fizycznych per capita (0,841) oraz udział osób w gospodarstwach domowych korzystających ze środowiskowej pomocy społecznej w ludności ogółem (0,837). Nieco mniejsze znaczenie ma zmienna opisująca średnie wpływy na osobę z podatku od osób prawnych (ładunek - 0,730).

Tabela. Macierz czynników^a

zmienne	czynnik 1
wpływy z PIT na osobę w gminie w 2010 roku	-0,884
świadczenia na rzecz osób fizycznych na osobę w 2010	0,841
udział osób w gospodarstwach domowych korzystających ze środowiskowej pomocy społecznej w ludności ogółem 2010	0,837
wpływy z CIT na osobę w gminie w 2010 roku	-0,730

metoda wyodrębniania czynników – głównych składowych.

^a. 1 – liczba wyodrębnionych czynników. Wymagana liczba iteracji: 11.

Oszacowania wartości na skali syntetycznej dla poszczególnych gmin dokonane zostały metodą regresyjną. Ze względu na znak przy poszczególnych ładunkach czynnikowych wskaźnik ten należy traktować jako informację o relatywnym poziomie ubóstwa mieszkańców poszczególnych gmin. Im wyższa wartość tym wyższe świadczenia, większy udział osób objętych świadczeniami w ogóle mieszkańców, niższe dochody z pracy najemnej oraz niższe dochody firm działających na obszarze gminy.

Aneks do rozdziału 10

Cechy nauczyciela

Opis skal wykorzystywanych jako wskaźniki stylów nauczania i przebiegu lekcji

Skale tu prezentowane pochodzą z baterii pytań P2-P5, P8-P10 z kwestionariusza ankiety dla ucznia gimnazjum z 3. etapu badania. Są to zmienne oznaczone jako g3_u_MNP_2_1-7, g3_u_MNP_5_1-18, g3_u_MNP_3_1-8, g3_u_MNM_8_1-7, g3_u_MNM_10_1-19, g3_u_MNM_9_1-8. Do Eksploracyjnej Analizy Czynnikowej (EFA) wykorzystano program SPSS, do Konfirmacyjnej Analizy Czynnikowej (CFA) wykorzystano program Mplus. Ponieważ w programie Mplus zmienne nie mogą przekraczać 8 znaków, nazwy zmiennych zmieniono odpowiednio p1-p7, p16-p33, p8-p15, m1-m7, m16-m34, m8-m15. Zmienne, które w nazwie mają MNP (lub „p” w zmienionej nazwie), dotyczą stylów prowadzenia lekcji języka polskiego. Zmienne, które w nazwie mają MNM (lub „m” w zmienionej nazwie), dotyczą stylów prowadzenia lekcji matematyki. Na pytania z baterii g3_u_MNP_2_1-7 (p1-p7), g3_u_MNP_5_1-18 (p16-p33), g3_u_MNM_8_1-7 (m1-m7), g3_u_MNM_10_1-19 (m16-m34) uczniowie odpowiadali na 4-punktowej skali: 1-„na każdej lekcji”, 2-„na większości lekcji”, 3-„czasami”, 4-„nigdy lub prawie nigdy”. Na pytania z baterii g3_u_MNP_3_1-8 (p8-p15), g3_u_MNM_9_1-8 (m8-m15) uczniowie odpowiadali na 5-punktowej skali: 1-„bardzo często”, 2-„często”, 3-„czasami”, 4-„rzadko”, 5-„nigdy”.

Do EFA wykorzystano estymację *Maximum Likelihood* (ML). Stosowano wykluczenie braków danych parami (PAIRWISE). Wszystkie rozwiązania nie były notowane.

Do CFA wykorzystano estymację *Weighted Least Squares* (WLS). Mplus oferuje dla zmiennych kategoryalnych następujące metody estymacji: MLR, MLF, WLS, WLSM, WLSMV, ULMSV. Stosowane w programie nazwy mogą być mylące, ponieważ MLR jest dokładnie tą samą metodą co MML (Marginal ML), MLF jest tą samą metodą co *Full Information ML*, WLSM jest skorygowaną do średniej metodą WLS, WLSMV jest metodą *Diagonally Weighted LS*, ULMSV jest tą samą metodą co Unweighted LS.

Zdecydowano się na wykorzystanie metody WLS, ponieważ jest metodą rekomendowaną dla dużych prób (~ 1000 i większych) i wysokich ładunków czynnikowych (> 0,7) (Nestler, 2013). Metoda WLS umożliwia także ocenę dopasowania modeli z wykorzystaniem statystyki χ^2 w tradycyjny sposób.

W przypadku małej liczby wskaźników, jak ma to miejsce w przypadku prezentowanych tu skal, WLS jest rekomendowaną metodą, zamiast innej często wykorzystywanej metody dla zmiennych kategoryalnych, tj. DWLS (Woods, 2002). DWLS ma tendencję do przeszacowywania parametrów (Forero, 2009; Nestler, 2013), jak również do akceptowania źle wyspecyfikowanych

modeli (Nestler, 2013). Estymację WLS zastosowano do wszystkich prezentowanych w aneksie skal, chyba że zaznaczono inaczej.

W CFA zastosowano domyślną w Mplus parametryzację Delta, w której ładunki czynnikowe są wolnymi parametrami, a reszty ustalone do jedności. Parametryzacja ta jest odzwierciedlona w sposobie wyliczania macierzy korelacji polychorycznych, stąd popularna w CFA dla zmiennych kategoryalnych.

Jako miary dopasowania raportowane są statystyka χ^2 , miara relatywnego dopasowania CFI oraz miara przybliżonego dopasowania RMSEA. Stosuje się te miary jako standardowe i najczęściej stosowane w publikacjach raportujących wyniki CFA (McDonald i Ho, 2002: 72).

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. Jednak w przypadku dużych prób (> 200) poziom istotności zazwyczaj jest zawsze istotny. Dlatego w ocenie dopasowania kierować należy się miarami RMSEA i CFI.

Zgodnie uznaje się RMSEA niższe niż 0,05, jako wskaźnik dobrego dopasowania, a wartość w przedziale 0,05–0,08 jako akceptowalne dopasowanie. 90% przedział ufności dla RMSEA nie powinien zawierać wartości > 0,8. Miara CFI powyżej 0,9 wskazuje na dobre dopasowanie (McDonald i Ho, 2002: 72).

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. Wszystkie prezentowane skale są jednowymiarowe.

Skala: „Autorytet nauczyciela/utrzymanie dyscypliny” (DYSP; DYSM)

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i kursywą błędy standardowe dla jęz. polskiego/i matematyki):

- Nauczyciel musi długo czekać, zanim uczniowie się uspokoją i w klasie będzie cisza (0.743; 0.010/0.820; 0.007).
- Uczniowie nie słuchają tego, co mówi nauczyciel (0.796; 0.010/0.818; 0.007).
- Uczniowie podczas lekcji zajmują się swoimi sprawami (0.792; 0.010/0.865; 0.007).

Dla języka polskiego skala ma następujące właściwości (przy nałożeniu ograniczenia w postaci równych ładunków czynnikowych): $n = 5066$; $\chi^2 = 19.293$; $df = 2$; $p = 0.000$; RMSEA = 0.041 (0.90 CI = 0.026 do 0.059; $p = 0.773$); CFI = 0.997). Bardzo dobre dopasowanie. Im wyższe wartości skali, tym większa dyscyplina.

Dla matematyki skala ma następujące właściwości (przy nałożeniu ograniczenia w postaci równych ładunków czynnikowych): $n = 5071$; $\chi^2 = 35.861$; $df = 2$; $p = 0.000$; RMSEA = 0.058

(0.90 CI = 0.042 do 0.075; $p = 0.194$); CFI = 0.997). Dobre dopasowanie. Im wyższe wartości skali, tym większa dyscyplina.

Skala: „Wymagający nauczyciel” (WYMP; WYMM)

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i kursywą błędy standardowe dla jęz. polskiego/i matematyki):

- Nauczyciel wymaga od uczniów dużo pracy (0.362; 0.018/0.272; 0.019).
- Nauczyciel robi niezapowiedziane kartkówki (0.642; 0.017/0.639; 0.016).
- Nauczyciel odpytuje uczniów z poprzedniej lekcji (0.678; 0.017/0.680; 0.017).
- Uczniowie przygotowują referaty lub prezentacje związane z tematem lekcji (0.530; 0.017/0.642; 0.018).

Dla języka polskiego skala ma następujące właściwości: $n = 5071$; $\chi^2 = 0.386$; $df = 2$; $p = 0.824$; RMSEA = 0.000 (0.90 CI = 0.000 do 0.017; $p = 1.000$); CFI = 1.000). Idealne dopasowanie. Im wyższe wartości na skali, tym mniej wymagający nauczyciel.

Dla matematyki skala ma następujące właściwości: $n = 5071$; $\chi^2 = 84,345$; $df = 2$; $p = 0.000$; RMSEA = 0.090 (0.90 CI = 0.074 do 0.107; $p = 0.000$); CFI = 0.954). Słabe dopasowanie. Im wyższe wartości na skali, tym mniej wymagający nauczyciel.

Skala: „Pomocny nauczyciel” (POMOCNP; POMOCNM)

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i kursywą błędy standardowe dla jęz. polskiego/i matematyki):

- Nauczyciel wykazuje zainteresowanie postępami w nauce każdego z uczniów (0.660; 0.010/0.709; 0.009).
- Nauczyciel tłumaczy tak długo, aż wszyscy uczniowie zrozumieją (0.796; 0.007/0.871; 0.006).
- Nauczyciel chętnie pomaga uczniom w nauce (0.874; 0.007/0.896; 0.006).
- Zajęcia są starannie zaplanowane (0.650; 0.010/0.652; 0.010).
- Nauczyciel podczas lekcji zachęca uczniów do zadawania pytań, dotyczących przerabianego materiału (0.603; 0.011/0.669; 0.010).

Dla języka polskiego skala ma następujące właściwości: $n = 5071$; $\chi^2 = 86.771$; $df = 5$; $p = 0.000$; RMSEA = 0.057 (0.90 CI = 0.047 do 0.068; $p = 0.131$); CFI = 0.991). Akceptowalne dopasowanie. Im wyższe wartości na skali, tym mniej pomocny nauczyciel.

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). Akceptowalne dopasowanie. Im wyższe wartości na skali, tym mniej pomocny nauczyciel.

Skala: „Koncentracja na wymaganiach” (PODEGZP; PODEGZM)

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i kursywą błędy standardowe dla jęz. polskiego/i matematyki; w nawiasach kwadratowych podano ładunki obliczone metodą głównych składowych odpowiednio dla jęz. polskiego i matematyki):

- Nauczyciel przypominał o czekającym Was egzaminie gimnazjalnym (0.849; 0.008/0.919; 0.005) [0.716; 0.794].
- Nauczyciel zwracał uwagę na umiejętności, które będą sprawdzane podczas egzaminu gimnazjalnego (0.860; 0.007/0.932; 0.004) [0.755; 0.820].
- Nauczyciel podkreślał, jak bardzo wybór szkoły ponadgimnazjalnej zależy od wyniku na egzaminie gimnazjalnym (0.656; 0.010/0.769; 0.007) [0.662; 0.734].
- Nauczyciel przygotował sprawdziany podobne do testów egzaminacyjnych, które rozwiązywaliśmy podczas lekcji w szkole (0.841; 0.007/0.873; 0.006) [0.736; 0.753].
- Nauczyciel omawiając sprawdziany nawiązywał do wymagań egzaminu gimnazjalnego (0.758; 0.008/0.840; 0.006) [0.742; 0.795].
- Podczas całej lekcji rozwiązywaliśmy arkusze egzaminacyjne (0.697; 0.009/0.764; 0.007) [0.616; 0.664].

Dla języka polskiego skala ma następujące właściwości: $n = 5069$; $\chi^2 = 884.435$; $df = 9$; $p = 0.000$; $RMSEA = 0.139$ (0.90 CI = 0.131 do 0.146; $p = 0.000$); $CFI = 0.927$). Dopasowanie niezadowalające. Zdecydowano się na zapisanie wartości na skali dla respondentów, wyrażonych jako główna składowa. Im wyższe wartości na skali, tym mniej intensywna koncentracja na wymaganiach.

Dla języka polskiego skala ma następujące właściwości: $n = 5072$; $\chi^2 = 970.797$; $df = 9$; $p = 0.000$; $RMSEA = 0.145$ (0.90 CI = 0.138 do 0.153; $p = 0.000$); $CFI = 0.966$). Dopasowanie niezadowalające. Zdecydowano się na zapisanie wartości na skali dla respondentów, wyrażonych jako główna składowa. Im wyższe wartości na skali, tym mniej intensywna koncentracja na wymaganiach.

Skala: „Aktywizujący styl nauczania” (ACTIVEP5; ACTIVEM5)

Jednowymiarowa skala dwustopniowa. Czynnik nadrzędny składa się z 4 czynników podrzędnych, każdy ładowany przez 2 wskaźniki.

Czynnik podrzędny: „Indywidualizacja nauczania” (ACTIVEP1; ACTIVEM1) (0.924; 0.011/0.863; 0.010). Jego wskaźnikami to odpowiedzi na pytania:

- Nauczyciel przydziela uczniom różne zadania, biorąc pod uwagę ich wyniki w nauce (0.756; 0.010/0.804; 0.009).

- Nauczyciel przydziela uczniom różne zadania, biorąc pod uwagę ich zainteresowania (0.797; 0.010/0.832; 0.009).

Czynnik podrzędny: „Multimedia” (ACTIVEP2; ACTIVEM2) (0.840; 0.013/0.807; 0.012).

Jego wskaźnikami to odpowiedzi na pytania:

- Nauczyciel zachęca do korzystania z dodatkowych materiałów, np. artykułów, informacji z Internetu (0.744; 0.012/0.864; 0.010).
- Nauczyciel korzysta z materiałów multimedialnych, programów edukacyjnych (0.650; 0.013/0.701; 0.011).

Czynnik podrzędny: „Podsumowania” (ACTIVEP3; ACTIVEM3) (0.880; 0.014/0.862;

0.010). Jego wskaźnikami to odpowiedzi na pytania:

- Pod koniec lekcji nauczyciel podsumowuje najważniejsze zagadnienia (0.724; 0.012/0.824; 0.009).
- Na początku lekcji nauczyciel podsumowuje to, co zrobiliśmy na poprzednich zajęciach (0.657; 0.012/0.755; 0.009).

Czynnik podrzędny: „Praca w grupach” (ACTIVEP4; ACTIVEM4) (0.699; 0.015/0.782;

0.013). Jego wskaźnikami to odpowiedzi na pytania:

- Nauczyciel dzieli uczniów na grupy, w których uczniowie wspólnie pracują (0.675; 0.013/0.770; 0.011).
- Uczniowie wspólnie tworzą plakaty odnoszące się do zagadnień poruszanych na lekcji (0.800; 0.014/0.877; 0.013).

Dla języka polskiego skala ma następujące właściwości: $n = 5069$; $\chi^2 = 241.969$; $df = 16$; $p = 0.000$; $RMSEA = 0.053$ (0.90 CI = 0.047 do 0.059; $p = 0.208$); $CFI = 0.966$). Akceptowalne dopasowanie. Im wyższe wartości na skali, tym mniej aktywizujący styl nauczania.

Tabela 8. Korelacje czynników podrzędnych i czynnika głównego w skali „Aktywizujący styl nauczania” dla jęz. polskiego

	ACTIVEP1	ACTIVEP2	ACTIVEP3	ACTIVEP4	ACTIVEP5
ACTIVEP1	1				
ACTIVEP2	0,906	1			
ACTIVEP3	0,936	0,896	1		
ACTIVEP4	0,799	0,755	0,757	1	
ACTIVEP5	0,986	0,946	0,968	0,831	1

Wartości istotne na poziomie $p < 0.01$ zaznaczono pogrubieniem

Dla matematyki skala ma następujące właściwości: ($n = 5071$; $\chi^2 = 232.110$; $df = 16$; $p = 0.000$; $RMSEA = 0.052$ ($0.90\text{ CI} = 0.046$ do 0.058 ; $p = 0.314$); $CFI = 0.983$). Akceptowalne dopasowanie. Im wyższe wartości na skali, tym mniej aktywizujący styl nauczania.

Tabela 9. Korelacje czynników podrzędnych i czynnika głównego w skali „Aktywizujący styl nauczania” dla matematyki

	ACTIVEM1	ACTIVEM2	ACTIVEM3	ACTIVEM4	ACTIVEM5
ACTIVEM1	1				
ACTIVEM2	0,811	1			
ACTIVEM3	0,863	0,818	1		
ACTIVEM4	0,812	0,778	0,796	1	
ACTIVEM5	0,950	0,909	0,949	0,894	1

Wartości istotne na poziomie $p < 0.01$ zaznaczono pogrubieniem

Prezentowane powyżej skale zostały zbudowane w oparciu o dane z kwestionariuszy uczniowskich. Uczniowie odpowiadali na pytania dotyczące stylów prowadzenia zajęć przez swoich nauczycieli. Poniżej raportowane są korelacje skal stylów prowadzenia zajęć osobno dla nauczania języka polskiego i matematyki.

Tabela 10. Korelacje skal stylów prowadzenia lekcji języka polskiego

	ACTIVEP5	DYSP	POMOCNP	WYMP	PODEGZP
ACTIVEP5	1				
DYSP	-0,151	1			
POMOCNP	0,564	-0,376	1		
WYMP	0,363	0,040	0,074	1	
PODEGZP	0,372	-0,115	0,414	0,070	1

Wartości istotne na poziomie $p < 0.01$ zaznaczono pogrubieniem

Tabela 11. Korelacje skal stylów prowadzenia lekcji matematyki

	ACTIVEM5	DYSM	POMOCNM	WYMM	PODEGZM
ACTIVEM5	1				
DYSM	-0,123	1			
POMOCNM	0,551	-0,377	1		
WYMM	0,502	-0,004	0,169	1	
PODEGZM	0,381	-0,196	0,460	0,191	1

Wartości istotne na poziomie $p < 0.01$ zaznaczono pogrubieniem

Kolejne skale powstały w oparciu o dane z kwestionariuszy nauczycielskich.

Skala: „Otwarty styl prowadzenia zajęć” (OPENSTNU)

Wskaźnikami skali są odpowiedzi na następujące pytania (w nawiasach podano ładunki czynnikowe i kursywą błędy standardowe).

Jak często podczas lekcji:

- nauczyciele zachęcają uczniów do wypracowania własnego zdania? (0.722; 0.009)
- uczniowie zgłaszają własne pomysły na wykonanie jakiegoś zadania szkolnego? (0.710; 0.009)
- uczniowie wyrażają swoje opinie podczas lekcji, nawet jeśli różnią się one od zdania większości uczniów? (0.708; 0.010)
- nauczyciele zachęcają uczniów do dyskusowania z osobami mającymi odmienne opinie w danej sprawie? (0.772; 0.008)
- nauczyciele tłumacząc jakiś problem, przedstawiają różne punkty widzenia? (0.708; 0.009)

Skala ma następujące właściwości: $n = 5042$; $\chi^2 = 162.802$; $df = 5$; $p = 0.000$; $RMSEA = 0.079$ (0.90 CI = 0.069 do 0.090; $p = 0.000$); $CFI = 0.975$). Akceptowalne dopasowanie. Im wyższe wartości na skali, tym bardziej otwarty styl prowadzenia zajęć.

Skala: „Utrzymanie dyscypliny” (DYSCNAU)

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i błędy standardowe):

- Muszę długo czekać, zanim uczniowie się uspokoją i w klasie będzie cisza (0.700; 0.021).
- Uczniowie nie zwracają uwagi na to, co mówię (0.818; 0.021).
- Uczniowie podczas lekcji zajmują się swoimi sprawami (0.797; 0.021).

Skala ma następujące właściwości (przy nałożeniu ograniczenia w postaci równych ładunków czynnikowych): $n = 2604$; $\chi^2 = 18.774$; $df = 2$; $p = 0.000$; $RMSEA = 0.057$ (0.90 CI = 0.035 do 0.081; $p = 0.275$); $CFI = 0.988$). Dobre dopasowanie. Im wyższe wartości na skali, tym większa dyscyplina na lekcjach.

Skala: „Aktywizujący styl nauczania” (ACTVNAU5)

Jednowymiarowa skala dwustopniowa.

Czynnik podrzędny: „Indywidualizacja nauczania” (ACTVNAU1) (0.770; 0.014). Jego wskaźnikami to odpowiedzi na pytania:

- Przydzielam uczniom różne zadania, w zależności od ich możliwości i poziomu umiejętności (0.765; 0.013).

- Przydzielam uczniom różne zadania, biorąc pod uwagę ich zainteresowania (0.838; 0.013).
- Przydzielam uczniom różne zadania, biorąc pod uwagę ich wyniki w nauce (0.690; 0.014).

Czynnik podrzędny: **„Multimedia” (ACTVNAU2)** (0.998; 0.020). Jego wskaźnikami to odpowiedzi na pytania:

- Korzystam z materiałów multimedialnych, programów edukacyjnych (0.547; 0.017).
- Korzystam z dodatkowych źródeł informacji i materiałów, np. artykułów z magazynów, Internetu (0.639; 0.017).
- Prowadzę zajęcia z wykorzystaniem gier dydaktycznych (0.678; 0.018).

Czynnik podrzędny: **„Podsumowania” (ACTVNAU3)** (0.625; 0.025). Jego wskaźnikami to odpowiedzi na pytania:

- Pod koniec lekcji podsumowuję najważniejsze zagadnienia (0.707; 0.024).
- Na początku lekcji podsumowuję to, co zrobiliśmy na poprzednich zajęciach (0.626; 0.024).

Czynnik podrzędny: **„Praca w grupach” (ACTVNAU4)** (0.900; 0.014). Jego wskaźnikami to odpowiedzi na pytania:

- Uczniowie przygotowują referaty lub prezentacje związane z lekcją (0.742; 0.017).
- Dzielę uczniów na grupy, w których uczniowie wspólnie pracują (0.717; 0.015).
- Uczniowie wspólnie tworzą plakaty odnoszące się do zagadnień poruszanych na lekcji (0.767; 0.013).
- Uczniowie tworzą pomoce dydaktyczne (0.681; 0.016).

Skala ma następujące właściwości: $n = 2607$; $\chi^2 = 589.008$; $df = 50$; $p = 0.000$; $RMSEA = 0.064$ ($0.90\text{ CI} = 0.060$ do 0.069 ; $p = 0.000$); $CFI = 0.895$). Akceptowalne dopasowanie. Im wyższe wartości na skali, tym mniej aktywizujący styl nauczania.

Tabela 12. Korelacje czynników podrzędnych i czynnika głównego w skali „Aktywizujący styl nauczania” (na bazie kwestionariuszy nauczycielskich)

	ACTVNAU1	ACTVNAU2	ACTVNAU3	ACTVNAU4	ACTVNAU5
ACTVNAU1	1				
ACTVNAU2	0,870	1			
ACTVNAU3	0,654	0,772	1		
ACTVNAU4	0,807	0,973	0,717	1	
ACTVNAU5	0,871	1	0,773	0,974	1

Wartości istotne na poziomie $p < 0.01$ zaznaczono pogrubieniem

Skala: „Pomocny nauczyciel” (POMOCNAU)

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i błędy standardowe):

- Uczniowie mają możliwość wyrażania swoich opinii (0.691; 0.032).
- Tłumaczę materiał do momentu, gdy zrozumie go większość uczniów w klasie (0.491; 0.030).
- Zachęcam uczniów do poszukiwania dodatkowych informacji na własną rękę, np. w Internecie (0.285; 0.030).
- Zdecydowana większość uczniów aktywnie uczestniczy w lekcji (0.384; 0.029).
- Zezwalam na swobodne zadawanie pytań w każdym momencie trwania lekcji (0.473; 0.026).

Skala ma następujące właściwości: $n = 2606$; $\chi^2 = 61.680$; $df = 5$; $p = 0.000$; $RMSEA = 0.066$ (0.90 CI = 0.052 do 0.081; $p = 0.032$); $CFI = 0.898$). Akceptowalne dopasowanie. Im wyższe wartości na skali, tym mniej pomocny nauczyciel.

Skala: „Wymagający nauczyciel” (WYMNAU)

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i błędy standardowe):

- Robię niezapowiedziane kartkówki (0.316; 0.052).
- Odpytuję uczniów z materiału z poprzednich lekcji (0.671; 0.106).
- Wymagam od uczniów dużo pracy (0.263; 0.048).

Skala ma następujące właściwości (przy nałożeniu ograniczenia w postaci równych ładunków czynnikowych): $n = 2609$; $\chi^2 = 18.037$; $df = 2$; $p = 0.000$; $RMSEA = 0.055$ (0.90 CI = 0.034 do 0.080; $p = 0.306$); $CFI = 0.888$). Dobre dopasowanie. Im wyższe wartości na skali, tym mniej wymagający nauczyciel.

Tabela 13. Korelacje skal stylów prowadzenia lekcji (na bazie kwestionariuszy nauczycielskich)

	ACTVNAU5	DYSCNAU	POMOCNAU	WYMNAU
ACTVNAU5	1			
DYSCNAU	-0,065	1		
POMOCNAU	0,237	-0,141	1	
WYMNAU	0,197	-0,038	0,109	1

Wartości istotne na poziomie $p < 0.01$ zaznaczono pogrubieniem

Skala: „Wsparcie nauczyciela”

Skala jest dwustopniowa (model hierarchicznej analizy czynnikowej). Składa się z 5 czynników pierwszego rzędu i 2 drugiego rzędu. W przeciwieństwie do wyżej prezentowanych skal, została oszacowana metodą WLSMV. W opisach pytań przypisanych do poszczególnych czynników (...) zastępuje tekst: *Do kogo zwróciłaby/zwróciłby się Pani/Pan z prośbą o radę/wsparcie w przypadku...?* Wszystkie pytania na skali 0–1 (nie wybrane-wybrane).

Wskaźnikami skali są odpowiedzi na pytania (w nawiasach podano ładunki czynnikowe i błędy standardowe):

Czynnik 1. rzędu: **„Wsparcie od innych nauczycieli tej samej szkoły” (E2NAUW)** (0.368; 0.048). Jego wskaźnikami są odpowiedzi na pytania:

- (...) kłopotów wychowawczych w klasie – inny nauczyciel z Pana/i szkoły (0.602; 0.022).
- (...) konfliktowej sytuacji z rodzicem ucznia – inny nauczyciel z Pana/i szkoły (0.545; 0.025).
- (...) wątpliwości merytorycznych dotyczących nauczanego przedmiotu – inny nauczyciel z Pana/i szkoły (0.839; 0.023).
- (...) wątpliwości metodycznych dotyczących nauczanego przedmiotu – inny nauczyciel z Pana/i szkoły (0.829; 0.017).
- (...) wyboru nowego podręcznika – inny nauczyciel z Pana/i szkoły (0.721; 0.019).
- (...) opiniowania przygotowanego narzędzia (np. pracy klasowej, sprawdzianu) – inny nauczyciel z Pana/i szkoły (0.738; 0.019).
- (...) wprowadzenia nowych metod pracy – inny nauczyciel z Pana/i szkoły (0.745; 0.017).
- (...) modyfikacji sposobu oceniania – inny nauczyciel z Pana/i szkoły (0.722; 0.018).
- (...) interpretacji wyników egzaminacyjnych – inny nauczyciel z Pana/i szkoły (0.668; 0.025).

Czynnik 1. rzędu: **„Wsparcie od dyrektora szkoły” (E2DYR)** (0.517; 0.072). Jego wskaźnikami są odpowiedzi na pytania:

- (...) kłopotów wychowawczych w klasie – dyrektor szkoły (0.603; 0.029).
- (...) konfliktowej sytuacji z rodzicem ucznia – dyrektor szkoły (0.538; 0.030).
- (...) wątpliwości merytorycznych dotyczących nauczanego przedmiotu – dyrektor szkoły (0.655; 0.031).
- (...) wątpliwości metodycznych dotyczących nauczanego przedmiotu – dyrektor szkoły (0.713; 0.029).
- (...) wyboru nowego podręcznika – dyrektor szkoły (0.509; 0.033).
- (...) opiniowania przygotowanego narzędzia (np. pracy klasowej, sprawdzianu) – dyrektor szkoły (0.589; 0.038).

- (...) wprowadzenia nowych metod pracy – dyrektor szkoły (0.618; 0.030).
- (...) modyfikacji sposobu oceniania – dyrektor szkoły (0.658; 0.026).
- (...) interpretacji wyników egzaminacyjnych – dyrektor szkoły (0.666; 0.027).

Czynnik 1. rzędu: **„Wsparcie od pedagoga/psychologa szkolnego” (E2PEDPS)** (1.474; 0.463).

Jego wskaźnikami są odpowiedzi na pytania:

- (...) kłopotów wychowawczych w klasie – pedagog szkolny/psycholog (0.480; 0.106).
- (...) konfliktowej sytuacji z rodzicem ucznia – pedagog szkolny/psycholog (0.469; 0.104).
- (...) wątpliwości merytorycznych dotyczących nauczanego przedmiotu – pedagog szkolny/psycholog (0.269; 0.072).
- (...) wątpliwości metodycznych dotyczących nauczanego przedmiotu – pedagog szkolny/psycholog (0.338; 0.083).
- (...) wyboru nowego podręcznika – pedagog szkolny/psycholog (0.388; 0.088).
- (...) opiniowania przygotowanego narzędzia (np. pracy klasowej, sprawdzianu) – pedagog szkolny/psycholog (0.427; 0.093).
- (...) wprowadzenia nowych metod pracy – pedagog szkolny/psycholog (0.403; 0.095).
- (...) modyfikacji sposobu oceniania – pedagog szkolny/psycholog (0.440; 0.103).
- (...) interpretacji wyników egzaminacyjnych – pedagog szkolny/psycholog (0.435; 0.096).

Czynnik 1. rzędu: **„Wsparcie z ośrodków doskonalenia nauczycieli” (E2ODN)** (0.498; 0.130).

Jego wskaźnikami są odpowiedzi na pytania:

- (...) kłopotów wychowawczych w klasie – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.575; 0.051).
- (...) konfliktowej sytuacji z rodzicem ucznia – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.530; 0.054).
- (...) wątpliwości merytorycznych dotyczących nauczanego przedmiotu – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.828; 0.043).
- (...) wątpliwości metodycznych dotyczących nauczanego przedmiotu – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.867; 0.048).
- (...) wyboru nowego podręcznika – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.653; 0.035).
- (...) opiniowania przygotowanego narzędzia (np. pracy klasowej, sprawdzianu) – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.680; 0.039).
- (...) wprowadzenia nowych metod pracy – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.766; 0.042).

- (...) modyfikacji sposobu oceniania – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.736; 0.042).
- (...) interpretacji wyników egzaminacyjnych – konsultant/doradca w ośrodku doskonalenia nauczycieli (0.691; 0.038).

Czynnik 1. rzędu: **„Wsparcie od nauczycieli innych szkół” (E2NAUZ)** (0.885; 0.324).
Jego wskaźnikami są odpowiedzi na pytania:

- (...) kłopotów wychowawczych w klasie – nauczyciel ze szkoły innej niż Pana/i (0.529; 0.088).
- (...) konfliktowej sytuacji z rodzicem ucznia – nauczyciel ze szkoły innej niż Pana/i (0.592; 0.101).
- (...) wątpliwości merytorycznych dotyczących nauczanego przedmiotu – nauczyciel ze szkoły innej niż Pana/i (0.700; 0.114).
- (...) wątpliwości metodycznych dotyczących nauczanego przedmiotu – nauczyciel ze szkoły innej niż Pana/i (0.724; 0.116).
- (...) wyboru nowego podręcznika – nauczyciel ze szkoły innej niż Pana/i (0.590; 0.097).
- (...) opiniowania przygotowanego narzędzia (np. pracy klasowej, sprawdzianu) – nauczyciel ze szkoły innej niż Pana/i (0.618; 0.101).
- (...) wprowadzenia nowych metod pracy – nauczyciel ze szkoły innej niż Pana/i (0.648; 0.105).
- (...) modyfikacji sposobu oceniania – nauczyciel ze szkoły innej niż Pana/i (0.639; 0.102).
- (...) interpretacji wyników egzaminacyjnych – nauczyciel ze szkoły innej niż Pana/i (0.586; 0.098).

Czynnik 2. rzędu: **„Wsparcie w ramach szkoły” (E2SZKOL)**. Ładuje czynniki 1. rzędu:

- „Wsparcie od innych nauczycieli tej samej szkoły” (E2_naucz_w – WN)
- „Wsparcie od dyrektora szkoły” (E2_dyr – WD)
- „Wsparcie od pedagoga/psychologa szkolnego” (E2_pedpsych – WPP).

Czynnik 2. rzędu: **„Wsparcie spoza szkoły” (E2ZEW)**. Ładuje czynniki 1. rzędu:

- „Wsparcie z ośrodków doskonalenia nauczycieli” (E2_odn – WD)
- „Wsparcie od nauczycieli innych szkół” (E2_naucz_z – WNI).

Skala ma następujące właściwości: $n = 630$; $\chi^2_{\text{WLSM}} = 8281.705$; $\chi^2_{\text{WLSMV}} = 3011.963$; $df = 939$;
 $p_{\text{WLSM}} < 0,00005$; $p_{\text{WLSMV}} < 0,00005$; $RMSEA_{\text{WLSM}} = 0.055$ (0.90 CI = 0.053 do 0.056; $p < 0,00005$);

$RMSEA_{WLSMV} = 0.029$ (0.90 CI = 0.028 do 0.030; $p > 0,9995$); $CFI_{WLSM} = 0.924$; $CFI_{WLSMW} = 0.919$).

Bardzo dobre dopasowanie. Im wyższe wartości na skali, tym bardziej nauczyciel ma bardziej rozbudowane sieci wsparcia.

Instytut Badań Edukacyjnych (IBE) jest placówką badawczą prowadzącą interdyscyplinarne badania naukowe nad funkcjonowaniem i efektywnością systemu edukacji w Polsce.

Metoda edukacyjnej wartości dodanej to instrument polityki oświatowej pozwalający lepiej wykorzystać dane egzaminacyjne do ewaluacji nauczania. Wskaźniki EWD mogą służyć zarówno bezpośrednio szkole do autoewaluacji i tworzenia planów rozwojowych, jak i do ewaluacji zewnętrznej dokonywanej przez wizytatora.

Projekt EWD realizowany jest w ramach działania 3.2. Rozwój systemu egzaminów zewnętrznych priorytetu III. Wysoka jakość systemu oświaty Programu Operacyjnego Kapitał ludzki. Celem głównym projektu jest rozwijanie metod wykorzystywania wyników egzaminów zewnętrznych do oceny efektywności nauczania.

www.ewd.edu.pl

Instytut Badań Edukacyjnych

ul. Górczewska 8, 01-180 Warszawa | tel. +48 22 241 71 00 | ewd@ewd.edu.pl | www.ewd.edu.pl
Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.