

Roman Dolata (red.)

# Czy szkoła ma znaczenie?

Analiza zróżnicowania efektywności  
nauczania na pierwszym etapie  
edukacyjnym

Tom 1



**KAPITAŁ LUDZKI**  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



*entuzjaści  
edukacji*

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY



Roman Dolata (red.)

# Czy szkoła ma znaczenie?

Analiza różnicowania  
efektywności nauczania  
na pierwszym etapie edukacyjnym

Tom I



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



entuzjaści  
edukacji

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY



**Redakcja naukowa:**

*dr hab. Roman Dolata*

**Autorzy:**

*dr Dorota Campfield*

*dr hab. Roman Dolata*

*Anna Hawrot*

*Kamila Hernik*

*Ewelina Jarnutowska*

*Aleksandra Jasińska-Maciąg*

*dr Radosław Kaczan*

*Karolina Malinowska*

*Michał Modzelewski*

*Michał Maluchnik*

*Jolanta Pisarek*

*Ludmiła Rycielska*

*dr Piotr Rycielski*

*Kamil Sijko*

*dr Magdalena Szpotowicz*

**Recenzenci:**

*dr hab. Ewa Filipiak*

*prof. dr hab. Bogdan Zawadzki*

© Copyright by: Instytut Badań Edukacyjnych, Warszawa, grudzień 2014

ISBN-978-83-61693-73-4

**Wzór cytowania:**

Dolata, R. (red.). (2014). *Czy szkoła ma znaczenie? Zróżnicowanie wyników nauczania po pierwszym etapie edukacyjnym oraz jego pozaszkolne i szkolne uwarunkowania (T. 1)*. Warszawa: Instytut Badań Edukacyjnych.

**Wydawca:**

*Instytut Badań Edukacyjnych, ul. Górczewska 8, 01-180 Warszawa, tel. (22) 241 71 00*

*www.ibe.edu.pl*

**Skład i druk:**

*Drukarnia TINTA, Z. Szymański, ul. Żwirki i Wigury 22, 13-200 Działdowo*

*www.drukarniatinta.pl*

Publikacja opracowana w ramach projektu systemowego: *Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego*, współfinansowanego przez Unię Europejską ze środków Europejskiego Funduszu Społecznego, realizowanego przez Instytut Badań Edukacyjnych.

Publikacja zastała wydrukowana na papierze ekologicznym.

Egzemplarz bezpłatny

# spis treści

## Tom I

<b>Wprowadzenie: założenia i cele badania szkolnych uwarunkowań efektywności kształcenia SUEK (Roman Dolata, Piotr Rycielski)</b> .....	<b>7</b>
Literatura cytowana .....	13

## Rozdział 1

<b>Próba badawcza i proces zbierania danych (Michał Maluchnik, Michał Modzelewski)</b> ...	<b>15</b>
1.1. Populacja badana .....	15
1.2. Próba .....	16
1.3. Schemat doboru próby .....	16
1.4. Próba zrealizowana .....	18
1.5. Reprezentatywność próby .....	19
1.6. Grupy uczestniczące w badaniu i najważniejsze narzędzia badawcze .....	23
1.7. Wagi zastosowane w analizach .....	24
Literatura cytowana .....	26

## Rozdział 2

<b>Pomiar wyników nauczania w zakresie czytania, świadomości językowej i matematyki (Aleksandra Jasińska-Maciążek, Michał Modzelewski)</b> .....	<b>27</b>
2.1. Podstawowe założenia i cel pomiaru .....	28
2.2. Charakterystyka zawartości treściowej testów .....	29
2.2.1. Test umiejętności czytania .....	30
2.2.2. Test świadomości językowej .....	31
2.2.3. Test umiejętności matematycznych .....	33
2.3. Konstrukcja testów osiągnięć szkolnych .....	35
2.4. Opis zawartości treściowej ostatecznej wersji testów .....	36
2.5. Model skalowania testów .....	39
2.5.1. Kalibracja zadań .....	40
2.5.2. Szacowanie wyników uczniów .....	41
2.5.3. Ocena dopasowania danych do modelu .....	42
2.5.4. Standaryzacja .....	44
2.6. Właściwości psychometryczne stworzonych testów .....	45
2.6.1. Właściwości zadań .....	45
2.6.2. Właściwości skal pomiarowych .....	46
2.6.3. Rzetelność i błąd pomiaru .....	48
2.6.4. Trafność pomiaru osiągnięć szkolnych .....	52
2.7. Podsumowanie .....	60
Literatura cytowana .....	62

## Rozdział 3

<b>Modele analizy zróżnicowania wyników nauczania (Aleksandra Jasińska-Maciążek, Michał Modzelewski)</b> .....	<b>65</b>
3.1. Problem przyczynowości w badaniach edukacyjnych .....	65

## CZY SZKOŁA MA ZNACZENIE?

3.2. Wielopoziomowe modele analizy a struktura danych edukacyjnych .....	66
3.3. Zastosowane modele analizy .....	67
3.4. Problem wag w modelach wielopoziomowych .....	71
3.5. Opis modeli kontrolnych opracowanych na potrzeby dalszych analiz .....	72
3.6. Prezentacja i omówienie wyników .....	75
3.6.1. Wyniki analiz dla testu umiejętności czytania .....	77
3.6.2. Wyniki analiz dla testu świadomości językowej .....	82
3.6.3. Wyniki analiz dla testu umiejętności matematycznych .....	85
3.6.4. Dyskusja wyników .....	88
3.7. Podsumowanie .....	90
Literatura cytowana .....	91

### Rozdział 4

<b>Międzyszkolne i międzyoddziałowe zróżnicowanie wyników nauczania (Aleksandra Jasińska-Maciąg, Michał Modzelewski) .....</b>	<b>93</b>
4.1. Zróżnicowanie międzyszkolne wyników nauczania po pierwszym etapie kształcenia na podstawie wyników badania SUEK .....	96
4.2. Zróżnicowanie międzyszkolne po pierwszym etapie kształcenia na tle drugiego i trzeciego etapu edukacji .....	100
4.2.1. Dane i metoda analizy .....	101
4.2.2. Zróżnicowanie międzyszkolne na kolejnych etapach edukacji .....	103
4.3. Zróżnicowanie międzyoddziałowe .....	104
4.4. Podsumowanie .....	108
Literatura cytowana .....	109

### Rozdział 5

<b>Inteligencja a wyniki nauczania (Anna Hawrot, Jolanta Pisarek) .....</b>	<b>111</b>
5.1. Koncepcje inteligencji .....	111
5.2. Inteligencja a osiągnięcia szkolne .....	113
5.3. Zróżnicowanie siły zależności pomiędzy osiągnięciami a inteligencją w szkołach .....	115
5.4. Problem badawczy .....	118
5.5. Pomiar inteligencji .....	119
5.6. Wyniki .....	120
5.6.1. Ogólny plan analiz .....	120
5.6.2. Statystyki opisowe .....	121
5.6.3. Inteligencja płynna a osiągnięcia w zakresie matematyki .....	122
5.6.4. Inteligencja płynna a osiągnięcia szkolne w zakresie czytania .....	124
5.6.5. Inteligencja płynna a osiągnięcia szkolne w zakresie świadomości językowej .....	127
5.7. Podsumowanie .....	129
Literatura cytowana .....	130

### Rozdział 6

<b>Lęk a wyniki nauczania (Anna Hawrot, Radosław Kaczan) .....</b>	<b>135</b>
6.1. Koncepcje i rodzaje lęku .....	136

6.2. Lęk a osiągnięcie szkolne .....	139
6.3. Metoda .....	144
6.4. Wyniki .....	146
6.4.1. Ogólny plan analiz .....	146
6.4.2. Analiza psychometryczna narzędzi i tworzenie wskaźników .....	147
6.4.3. Statystyki opisowe: natężenie lęku i napięcia emocjonalnego w zbadanej grupie .....	151
6.4.4. Lęk a osiągnięcia z matematyki .....	154
6.4.5. Lęk a osiągnięcia w zakresie czytania .....	156
6.4.6. Lęk a osiągnięcia w zakresie świadomości językowej .....	158
6.5. Podsumowanie .....	161
Literatura cytowana .....	167

## Tom II

### Rozdział 7

#### **Czynniki statusowe a wyniki nauczania (Roman Dolata, Ewelina Jarnutowska) ..... 177**

7.1. Nierówności edukacyjne i mechanizmy ich powstawania .....	178
7.2. Szkolne czynniki wpływające na nasilenie nierówności edukacyjnych i dynamika nierówności .....	186
7.3. Metoda .....	187
7.3.1. Pomiar cech statusowych .....	187
7.3.2. Charakterystyka pozostałych zmiennych użytych w analizach .....	191
7.4. Wyniki .....	194
7.4.1. Międzyodziałowe i międzyoddziałowe zróżnicowanie statusu ekonomiczno-społecznego rodziny ucznia .....	194
7.4.2. Nasilenie nierówności edukacyjnych: SES rodziny ucznia a osiągnięcia szkolne ..	197
7.4.3. Znaczenie inteligencji ucznia dla nasilenia nierówności edukacyjnych .....	204
7.4.4. Pozycja rodziców na rynku pracy a wyniki nauczania .....	208
7.4.5. Struktura rodziny ucznia a wyniki nauczania .....	212
7.4.6. Lokalizacja szkoły a wyniki nauczania .....	218
7.4.7. Międzyodziałowe zróżnicowanie nasilenia nierówności edukacyjnych i jego szkolne uwarunkowania .....	221
7.5. Podsumowanie .....	227
Literatura cytowana .....	229

### Rozdział 8

#### **Cechy nauczyciela a wyniki nauczania (Kamila Hernik, Karolina Malinowska, Kamil Sijko) ..... 235**

8.1. Problem wpływu nauczyciela na wyniki nauczania .....	235
8.2. Charakterystyka nauczycieli edukacji wczesnoszkolnej .....	236
8.2.1. Cechy socjodemograficzne .....	236
8.2.2. Przekonania i postawy nauczycieli edukacji wczesnoszkolnej .....	239
8.2.3. Pochodzenie społeczne badanych nauczycieli .....	242
8.2.4. Nauczyciele – wychowawcy badanych oddziałów .....	249

## CZY SZKOŁA MA ZNACZENIE?

8.3. Wyniki .....	249
8.3.1. Doświadczenie zawodowe nauczyciela a wyniki nauczania .....	250
8.3.2. Przekonania i postawy nauczyciela a wyniki nauczania .....	256
8.3.3. Pochodzenie społeczne nauczyciela a wyniki nauczania .....	261
8.4. Podsumowanie .....	266
Literatura cytowana .....	266

### Rozdział 9

#### **Podręcznik a wyniki nauczania (Kamila Hernik, Karolina Malinowska, Kamil Sijko) .....**

9.1. Metoda .....	269
9.2. Wyniki .....	270
9.3. Podsumowanie .....	275
Literatura cytowana .....	276

### Rozdział 10

#### **Wyniki nauczania w zakresie języka angielskiego i ich uwarunkowania**

#### **(Magdalena Szpotowicz, Dorota Campfield, Ludmiła Rycielska) .....**

10.1. Problem .....	277
10.2. Pomiar przyswojenia języka obcego u dzieci .....	278
10.3. Czynniki wpływające na umiejętności językowe uczniów .....	281
10.4. Cele badania .....	283
10.5. Metoda .....	283
10.6. Wyniki .....	290
10.6.1. Międzyszkolne zróżnicowanie wyników nauczania z języka angielskiego .....	290
10.6.2. Uwarunkowania wyników nauczania z języka angielskiego .....	290
10.7. Podsumowanie .....	299
Literatura cytowana .....	302

#### **Podsumowanie (Roman Dolata) .....**

Charakterystyka metodologiczna badania .....	308
Najważniejsze wyniki pierwszej fazy badania SUEK .....	310

## Wprowadzenie: założenia i cele badania szkolnych uwarunkowań efektywności kształcenia SUEK

Badania nad uwarunkowaniami szkolnej efektywności kształcenia (*school effectiveness research – SER*) mają wieloletnią, międzynarodową tradycję. Doczekały się już opracowań encyklopedycznych, których przykładem może być obszerna, dwutomowa praca pod redakcją Tony Townsenda *International Handbook of School Effectiveness and Improvement* (2007). W ciągu ostatnich dwudziestu lat projekty badawcze dotyczące efektywności kształcenia wprowadzane były do programu badań edukacyjnych w bardzo dużej liczbie krajów. Badania takie przynoszą wymierne korzyści w postaci ewaluacji systemów oświaty oraz dostarczają wiedzy o znaczeniu różnych charakterystyk szkół dla efektywności kształcenia. Wiedza ta pozwala lepiej planować różnego typu interwencje publiczne mające na celu doskonalenie pracy szkół (*school improvement research*).

Dla badań SER kluczową kwestią metodologiczną jest zdolność do odkrywania związków przyczynowo-skutkowych. Zawodność w tym zakresie schematu badań przekrojowych sprawia, że coraz częściej wykorzystywane są „mocniejsze” plany badawcze. Choć nic nie zastąpi schematu eksperymentalnego, to względy praktyczne sprawiają, że realną alternatywą dla badań przekrojowych stają się badania longitudinalne.

W ostatnich latach uruchomiono wiele podłużnych badań edukacyjnych. Poniższa tabela pozwala sięgnąć głębiej w historię i przedstawia zestawienie tego typu studiów empirycznych od lat 50. ubiegłego wieku.

**Tabela 1. Przykłady podłużnych badań edukacyjnych**

Nazwa badania (nazwy międzynarodowe)	Kraj	Rok rozpoczęcia badań	Początkowa wielkość próby
NCDS, National Child Development Study	Wielka Brytania	1958	17 500
BCS 70, British Cohort Study	Wielka Brytania	1970	17 198
Mannheim Study of Children at Risk	Niemcy	1986	384
NICHD, Study of Early Child Care	USA	1991	1364
JLD, Jyväskylä Longitudinal Study of Dyslexia	Finlandia	1993	200
DALSC, Danish Longitudinal Survey of Children	Dania	1995	7200



## CZY SZKOŁA MA ZNACZENIE?

Nazwa badania (nazwy międzynarodowe)	Kraj	Rok rozpoczęcia badań	Początkowa wielkość próby
QLSCD, Quebec Longitudinal Study of Child Development	Kanada	1998	2817
ECLS-B, Early Childhood Longitudinal Study	USA	2001	10 700
MCS, Millenium Cohort Study	Wielka Brytania	2001	18 818
LASC, Longitudinal Study of Australian Children	Australia	2004	5107
GUS, Growing Up in Scotland	Szkocja	2005	5000
GUI, Growing Up in Ireland	Irlandia	2008	11 000
ELFE, French Longitudinal Study of Children	Francja	2011	20 000

Źródło: Schlessiger i in. (2011)

Badania podłużne uruchamiane są w celu systematycznego obserwowania pewnej kohorty osób badanych. Cel może być specyficzny, jak w przypadku dzieci z dysleksją badanych w JLD, lub ogólny, jak na przykład w przypadku badań o wdzięcznej nazwie „Dorastanie w Irlandii”. Trzeba zauważyć, że i polska pedagogika ma w swoim dorobku liczebnie skromne, ale poznawczo ważne badanie tego typu (badanie toruńskie, Kwieciński, 2002).

Badania prowadzone w obrębie tradycji school effectiveness research i school improvement research pozwalają formułować dobrze uzasadnione twierdzenia dotyczące zarówno szkolnych jak i pozaszkolnych czynników odpowiedzialnych za osiągnięcia szkolne (Townsend, 2007, Hattie, 2009). Jednak niewielka liczba badań edukacyjnych prowadzonych w polskich realiach nie pozwala na ich weryfikację w odniesieniu do polskich szkół. Uogólnienia empiryczne zbudowane głównie na podstawie wyników badań w USA i krajach zachodniej Europy należy zatem traktować jako hipotezy wymagające weryfikacji w polskich warunkach kulturowych.

Celem podłużnego badania edukacyjnego realizowanego w Instytucie Badań Edukacyjnych jest identyfikacja kluczowych czynników szkolnych warunkujących efektywność kształcenia na II etapie edukacyjnym (klasy IV–VI szkoły podstawowej). Program pod nazwą *Szkolne Uwarunkowania Efektywności Kształcenia (SUEK)* rozpoczął się na jesieni 2010 roku, gdy badani uczniowie uczęszczali do III klas szkół podstawowych. We wrześniu 2014 roku uczniowie z realizowanego badania podłużnego rozpoczynają naukę w gimnazjum. Taki dobór trwania badania umożliwi uwzględnienie najważniejszych dla edukacji na poziomie podstawowym progów szkolnych: między III i IV klasą oraz między szkołą podstawową i gimnazjum.

## WPROWADZENIE: ZAŁOŻENIA I CELE BADANIA SZKOLNYCH UWARUNKOWAŃ EFEKTYWNOŚCI KSZTAŁCENIA SUEK

**Tabela 2. Kalendarz badania SUEK**

<b>Rok szkolny:</b>	<b>Klasa</b>
2010/11	III klasa
2011/12	IV klasa
2012/13	V klasa
2013/14	VI klasa
2014/15	I klasa gimnazjum

Mimo że plan finansowania projektu obejmuje okres do czerwca 2015 roku, uważamy, że rozpoczęte badanie podłużne SUEK powinno być kontynuowane tak długo, jak to będzie możliwe. Umożliwi to podjęcie w badaniu wielu istotnych problemów daleko wykraczających poza uwarunkowania efektów kształcenia w pierwszym okresie edukacji.

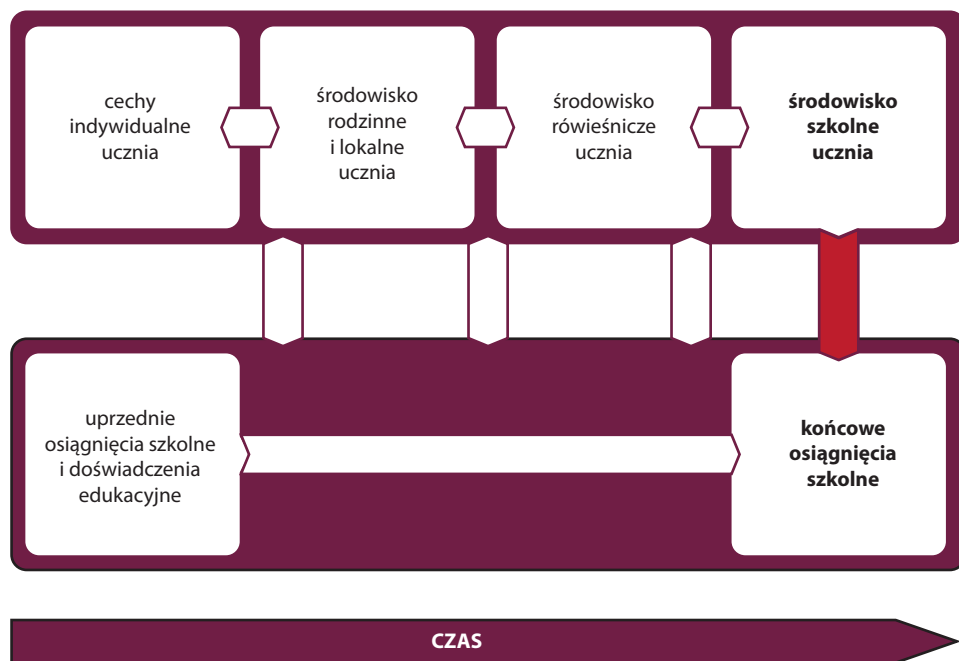
Planowane studium empiryczne poza klasyczną dla badań podstawowych funkcją poznawczą ma spełnić również funkcję monitorującą – dostarczyć rzetelnych, trafnych i reprezentatywnych danych na potrzeby długofalowego monitorowania stanu edukacji w Polsce.

Posługując się sformułowaniem autorów raportu Plowden (Central Advisory Council for Education, 1967) można powiedzieć, że osiągnięcia szkolne są efektem działania trudnej do rozplątania sieci zależności czynników biologicznych, psychospołecznych, społecznych i szkolnych. Gdy próbujemy zatem zidentyfikować kluczowe czynniki szkolne odpowiedzialne za efektywność kształcenia napotykamy na problem zwany endogenicznością. Na poziomie technicznym endogeniczność oznacza pozorny związek między zmiennymi niezależnymi a zmiennymi zależnymi, wynikający ze skorelowania tych zmiennych z cechami nieuwzględnionymi w modelu wyjaśniającym. Inaczej mówiąc, zmienne pominięte w modelu mogą być faktyczną przyczyną obserwowanej zależności, a przez to oszacowania siły związku mogą być nietrafne (Lee, 2005). Jest to problem, który napotyka każdy badacz zjawisk społecznych nie dających się eksplorować za pomocą eksperymentu, ale w wypadku badań edukacyjnych jest on szczególnie ważny i trudny do rozwiązania. Zadowalające poradzenie sobie z problemem endogeniczności wymaga starannej kontroli pozaszkolnych czynników odpowiedzialnych za osiągnięcia szkolne uczniów oraz stosowania adekwatnych modeli analizy danych.

Planowane jest badanie osiągnięć szkolnych i różnego rodzaju charakterystyk psychologicznych uczniów, cech ich środowiska rodzinnego, charakterystyk nauczycieli i środowiska

## CZY SZKOŁA MA ZNACZENIE?

szkolnego. Podstawowy schemat analizy dla głównego cyklu (klasy IV–VI) badania SUEK przedstawia poniższy rysunek.

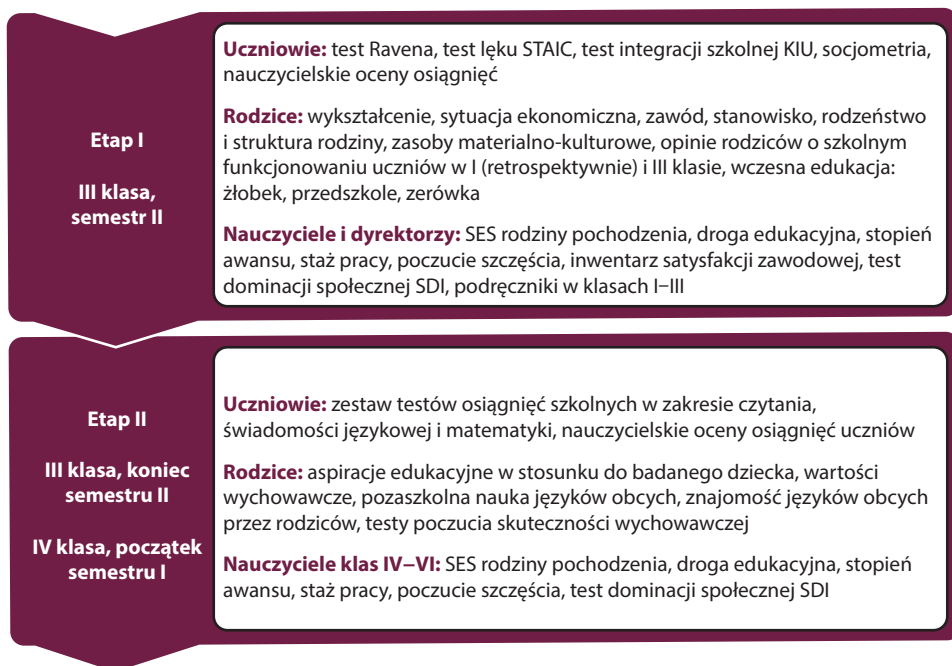


Rysunek 1. Model analizy danych w badaniu SUEK.

Powyższy model analizy pozwoli uwzględnić podłużny charakter badania. Modele tego typu nazywamy modelami edukacyjnej wartości dodanej. Pozwalają one na podstawie pomiarów osiągnięć szkolnych na „wejściu” i na „wyjściu” oraz kontroli czynników pozaszkolnych lepiej szacować „czysty” wkład czynników szkolnych w osiągnięcia szkolne (czerwona strzałka na schemacie).

**W poniższym raporcie przedstawiono wyniki pierwszej, wstępnej fazy badania SUEK.** Głównym celem poznawczym tej fazy programu SUEK był swoisty „bilans zamknięcia” pierwszego etapu edukacyjnego (klasy I–III) w zakresie efektywności nauczania. Przyjęto różnicowy model oceny efektywności nauczania, czyli skoncentrowano się na określeniu międzyszkolnego i międzyoddziałowego zróżnicowania wyników nauczania i zbadaniu pozaszkolnych i szkolnych uwarunkowań tego zróżnicowania. W pierwszej fazie (etap I i II badania) zebrano następujące dane:

## WPROWADZENIE: ZAŁOŻENIA I CELE BADANIA SZKOLNYCH UWARUNKOWAŃ EFEKTYWNOŚCI KSZTAŁCENIA SUEK



Rysunek 2. Najważniejsze pomiary i obszary eksploracji w badaniu SUEK we wstępnej fazie badania<sup>1</sup>.

Zrealizowane w klasie III i na początku klasy IV pomiary dostarczyły danych pozwalających na:

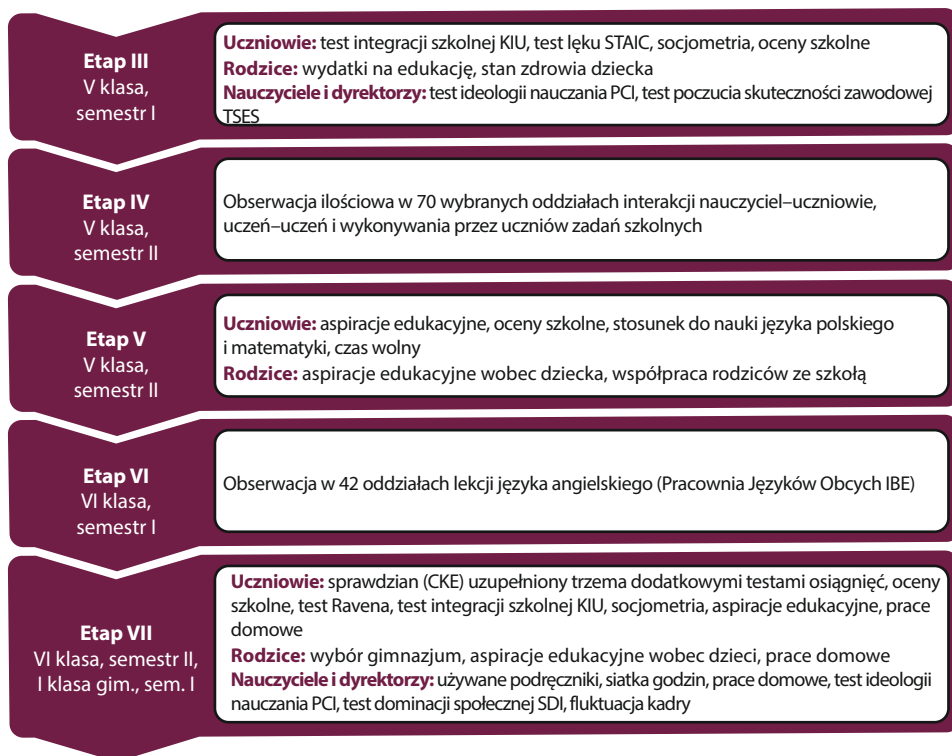
- ilościowy opis międzyszkolnego, międzyoddziałowego i wewnątrzoddziałowego zróżnicowania wyników nauczania;
- wyjaśnienie części tego zróżnicowania niezależnymi od szkoły czynnikami biopsychologicznymi i rodzinnymi i oszacowanie w ten sposób, w jakim stopniu to system szkolny odpowiada za to zróżnicowanie;
- oszacowanie wyników nauczania z wytrąceniem wpływu czynników pozaszkolnych, czyli wyznaczenie efektywności nauczania w danej szkole lub oddziale;
- oszacowanie wpływu wybranych czynników szkolnych na międzyszkolne i międzyoddziałowe zróżnicowanie efektywności nauczania.

Uzyskane w fazie wstępnej wyniki będą punktem wyjścia dla analiz w głównej fazie badania SUEK poświęconej II etapowi edukacyjnemu (klasy IV–VI szkoły podstawowej).

<sup>1</sup> Szczegółowe informacje znajdują się w rozdziale 1.

## CZY SZKOŁA MA ZNACZENIE?

W głównej fazie badania przewidzianych jest 5 kolejnych etapów zbierania danych (czyli w całym badaniu etapy od III do VII). Plan najważniejszych pomiarów w kolejnych etapach przedstawia poniższy schemat:



Rysunek 3. Najważniejsze pomiary i obszary eksploracji w badaniu SUEK w głównej fazie badania<sup>2</sup>.

Na koniec wprowadzenia krótkie przedstawienie struktury raportu. W rozdziale 1. Michał Maluchnik omawia dobór próby i system wag, problem reprezentatywności próby, realizację badania w terenie i dokonywane pomiary. W kolejnym rozdziale Aleksandra Jasińska i Michał Modzelewski referują koncepcję pomiaru osiągnięć szkolnych i psychometryczną charakterystykę stworzonego na potrzeby badania SUEK zestawu testów osiągnięć po I etapie edukacyjnym. Ci sami Autorzy w rozdziale 3. prezentują wykorzystywane w pracy modele statystycznej analizy danych. To bardzo ważna część raportu, ponieważ od trafnego wyboru modeli analizy w dużej mierze zależy wartość uzyskanych wyników. Kolejny rozdział

<sup>2</sup> Szczegółowe informacje znajdują się w rozdziale 1.

## WPROWADZENIE: ZAŁOŻENIA I CELE BADANIA SZKOLNYCH UWARUNKOWAŃ EFEKTYWNOŚCI KSZTAŁCENIA SUEK

autorstwa Jasińskiej i Modzelewskiego zawiera kluczową dla raportu analizę międzyszkolnego i międzyoddziałowego zróżnicowania wyników i efektywności nauczania. W następnej części raportu Anna Hawrot i Jolanta Pisarek analizują związki inteligencji z osiągnięciami szkolnymi. Rozdział 6. napisany przez Annę Hawrot i Radosława Kaczana prezentuje wyniki badania pokazujące znaczenie lęku dla szkolnych osiągnięć uczniów. Rozdział 7. wprowadza nas w niezwykle ważną problematykę statusowej determinacji osiągnięć szkolnych – Roman Dolata i Ewelina Jarnutowska określają siłę tej determinacji i szukają czynników wyjaśniających. Liczą się nauczyciele – to prawda powszechnie uznawana, ale jakie mierzalne cechy nauczyciela decydują o tym, że efektywnie naucza? W rozdziale 8. próbują na to pytanie odpowiedzieć Kamila Hernik, Karolina Malinowska i Kamil Sijko. Kolejna część książki dotyczy wpływu na efektywność nauczania podręcznika szkolnego. Czy wybór tego, a nie innego podręcznika daje większe szanse na efektywne nauczanie języka polskiego czy matematyki? Na to pytanie odpowiada znany już zespół autorski: Kamila Hernik, Karolina Malinowska i Kamil Sijko. Rozdział 10. wyłamuje się ze struktury raportu, ale jest to usprawiedliwione specyfiką problemu. Zróżnicowanie efektywności nauczania języka angielskiego na I etapie edukacyjnym naświetlają Magdalena Szpotowicz, Dorota Campfield i Ludmiła Rycielska. Raport kończy Podsumowanie, w którym redaktor naukowy książki wraca do głównych pytań badawczych i próbuje syntetycznie na nie odpowiedzieć.

### Literatura cytowana

- Hattie, J. (2009). *Visible learning. A synthesis of over, 800 Meta-Analyses Relating to Achievement*. Routledge.
- Kwieciński (2002). *Nieuniknione? Funkcje alfabetyzacji w dorosłości*. Toruń-Olsztyn, Wydawnictwo UMK.
- Plowden Committee. (1967). *Children and their primary schools*. Central Advisory Council for Education, London.
- Schlesiger, C., Lorenz, J., Weinert, S., Schneider, T., & Roßbach, H. G. (2011). From birth to early child care. *Zeitschrift für Erziehungswissenschaft*, 14(2), 187–202.
- Townsend, T. (red.). (2007). *International handbook of school effectiveness and improvement*. Dordrecht, The Netherlands: Springer.



## Rozdział 1

### Próba badawcza i proces zbierania danych

#### 1.1. Populacja badana

Badanie szkolnych uwarunkowań efektywności kształcenia (SUEK) jest badaniem podłużnym, realizowanym na próbie pozwalającej na uprawnione dla kraju wnioskowanie statystyczne na poziomie szkół, oddziałów i uczniów szkół podstawowych. Badanie objęło uczniów szkół podstawowych uczących się w roku szkolnym 2010/2011 w III klasie.

Ze względu na podłużny charakter badania, konieczne było rozwiązanie problemu zmian próby w trakcie badania. Przyjęto jedno ze stosowanych w takich przypadkach rozwiązań, w którym próba uzupełniana jest o nowych uczniów dołączających do badanych oddziałów. Jest to tzw. *fixed panel plus 'births' design*, (Elliot, 2009). Zgodnie z tym rozwiązaniem, niektórzy uczniowie mogli przestać uczestniczyć w badaniu (np. zmienili szkołę lub nie otrzymali promocji do następnej klasy), możliwe jest też dołączanie do badania nowych uczniów (np. uczniów, którzy w danym semestrze pojawili się w szkole lub uczniów, którzy powtarzają klasę i tym samym dołączyli do badanych oddziałów). Dodatkową komplikacją jest przejście uczniów z pierwszego etapu kształcenia w szkole podstawowej do drugiego etapu, co może wiązać się ze zmianami oddziału przez uczniów.

Operat losowania skonstruowano na podstawie danych pochodzących z Systemu Informacji Oświatowej (SIO, stan z 30 września 2009). Na podstawie informacji z roku szkolnego 2009/2010 stworzono listę szkół podstawowych z informacją o liczbie oddziałów klasowych i liczbie uczniów na poziomie II klasy. Założono, że większość uczniów otrzyma promocję do następnej klasy, a ewentualne migracje międzyszkolne lub międzyoddziałowe nie zmienią znacząco stanów liczebnościowych klas.

Z operatu zostały wyłączone szkoły specjalne i integracyjne, szkoły mistrzostwa sportowego i sportowe, przyszpitalne oraz eksperymentalne (jedna szkoła w skali kraju). Oddziały integracyjne nie znalazły się w próbie ze względu na swoją specyfikę. Są to zazwyczaj małe oddziały z nauczycielem wspomagającym, którym powinny być dedykowane oddzielne badania. Ze względu na małą precyzję estymowanych parametrów w operacie nie znalazły się również szkoły o prognozowanej liczbie uczniów poniżej 11 osób. Dokonane na operacie zabiegi pomniejszyły populację uczniów o około 7,3%. Finalnie uzyskano operat składający się z 309 285 uczniów, uczęszczających do 8924 szkół. Należy dodać, że już na etapie realizacji



## CZY SZKOŁA MA ZNACZENIE?

terenowej wykluczano z badania dzieci z problemami zdrowotnymi (intelektualnymi lub fizycznymi), które uniemożliwiały im rozwiązywanie testów lub uzupełnianie kwestionariuszy.

### 1.2. Próba

W badaniu przyjęto, że oszacowania wyników odbywać się będą na trzech poziomach analizy: dla szkół, oddziałów oraz uczniów. Implementując rozwiązania z badań takich jak CivEd (Schulz & Sibberns, 2004), PIRLS (Kennedy, Martin, & Mullis, 2007), TIMSS (Olson, Martin, & Mullis, 2008) czy PISA (Biedrzycki, i inni, 2009) założono, że aby uzyskać jak najlepsze oszacowania na każdym z wymienionych poziomów, należy wylosować co najmniej 150 szkół. Próba uczniów z tych szkół da oszacowania nie gorsze niż te, które by uzyskano, losując w sposób prosty niezależny co najmniej 400 uczniów. Mając na uwadze, że badanie jest podłużne i ma trwać 5 lat, wylosowano ich 180. Pozwoli to w jakimś stopniu uniknąć spadku precyzji oszacowań. Doświadczenia terenowe wielu podobnych badań wskazują, że *response rate* nigdy nie wynosi 100%, co było dodatkowym argumentem za wylosowaniem większej liczby szkół.

### 1.3. Schemat doboru próby

Schemat doboru uczniów do badania SUEK miał charakter doboru grupowego, warstwowego z zastosowaniem losowania systematycznego. W pierwszym etapie losowania podzielono operat na sześć warstw, uwzględniających wielkość miejsca zamieszkania (dwie kategorie: 1. wieś i miasto do 5 tys. mieszkańców oraz 2. miasto powyżej 5 tys. mieszkańców) oraz wielkość szkoły, mierzoną liczbą oddziałów w szkole (trzy kategorie: 1. szkoły jednooddziałowe, 2. szkoły dwuoddziałowe oraz 3. szkoły trzyoddziałowe i większe).

**Tabela 1.1. Liczba szkół i uczniów na poziomie klasy II w operacie losowania z uwzględnieniem warstw**

warstwa	liczba szkół	liczba uczniów	%
Wieś i miasto do 5 tysięcy mieszkańców Szoły jednooddziałowe	4312	73 290	23,70%
Wieś i miasto do 5 tysięcy mieszkańców Szoły dwuoddziałowe	1156	42 364	13,70%
Wieś i miasto do 5 tysięcy mieszkańców Szoły trzyoddziałowe i większe	223	14 162	4,60%

warstwa	liczba szkół	liczba uczniów	%
Miasto powyżej 5 tysięcy mieszkańców Szkoly jednooddziałowe	713	14 444	4,70%
Miasto powyżej 5 tysięcy mieszkańców Szkoly dwuoddziałowe	1105	47 774	15,40%
Miasto powyżej 5 tysięcy mieszkańców Szkoly trzyoddziałowe i większe	1415	117 251	37,90%
ogółem	8924	309 285	100,00%

Źródło: opracowanie własne

W każdej warstwie wylosowano w sposób systematyczny szkoły zgodnie z zasadą alokacji proporcjonalnej do liczby uczniów w warstwie. Przedstawia to poniższa tabela.

**Tabela 1.2. Alokacja szkół w próbie**

warstwa	liczba szkół w próbie
Wieś i miasto do 5 tysięcy mieszkańców Szkoly jednooddziałowe	42
Wieś i miasto do 5 tysięcy mieszkańców Szkoly dwuoddziałowe	25
Wieś i miasto do 5 tysięcy mieszkańców Szkoly trzyoddziałowe i większe	8
Miasto powyżej 5 tysięcy mieszkańców Szkoly jednooddziałowe	9
Miasto powyżej 5 tysięcy mieszkańców Szkoly dwuoddziałowe	28
Miasto powyżej 5 tysięcy mieszkańców Szkoly trzyoddziałowe i większe	68
ogółem	180

Źródło: opracowanie własne

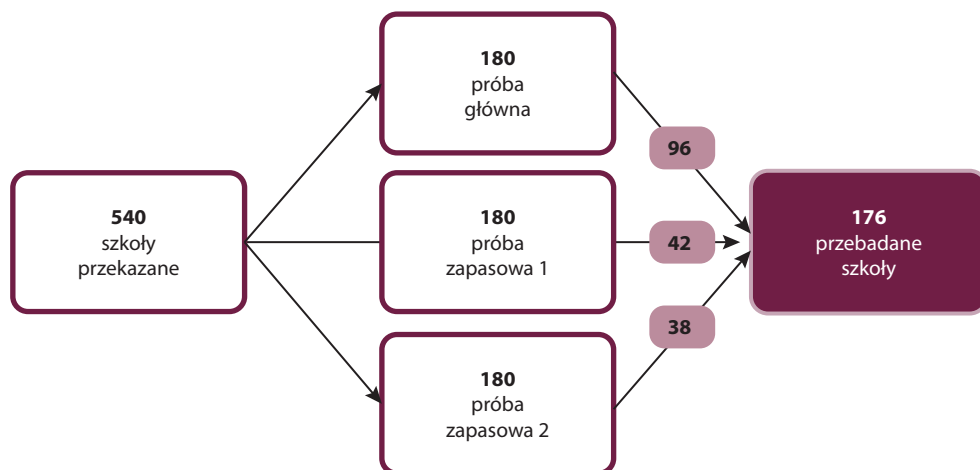
Po uporządkowaniu wszystkich szkół w ramach warstwy rosnąco po kodzie pocztowym, otrzymano listę, z której wylosowano w sposób systematyczny 180 szkół z próby głównej. Szkoła, która występowała na liście powyżej wylosowanej, trafiła do próby zapasowej 1, zaś ta poniżej – do zapasowej 2. Zabieg ten umożliwił stworzenie zastępników szkół o maksymalnie podobnych (warstwa, lokalizacja geograficzna) do wytypowanych do próby głównej.

## CZY SZKOŁA MA ZNACZENIE?

W kolejnym kroku następowało losowanie oddziałów w ramach szkół. W przypadku szkół jedno- i dwuoddziałowych, kwalifikowano wszystkie oddziały do badania. W przypadku szkół trzyoddziałowych i większych losowano dwa oddziały. Losowano je w sposób prosty, zależny, wykorzystując liczby losowe.

### 1.4. Próba zrealizowana

Z wylosowanej próby głównej 53% szkół przystąpiło do badania, pozostałe zostały zrekrutowane z prób zapasowych. Ostatecznie w pierwszym etapie badania próba liczyła 176 szkół, gdyż Wykonawca badania w terenie<sup>1</sup> nie zdążył przeprowadzić rekrutacji w przewidzianym czasie. Oznacza to, że wielkość próby zrealizowanej stanowi 98% próby zakładanej. Jednak *response rate* liczony jako odsetek szkół zrekrutowanych do liczby placówek, które proszono o udział w badaniu, wyniósł 67%. Poniższy schemat przedstawia wykorzystanie poszczególnych prób.



Rysunek 1.1. Schemat realizacji prób w Badaniu SUEK.

Tabela 1.3. przedstawia, jaki był odsetek uczniów w poszczególnych warstwach w próbie głównej.

<sup>1</sup> Z uwagi na skalę *Badania szkolnych uwarunkowań efektywności kształcenia* Instytut Badań Edukacyjnych, afiliujący ten projekt badawczy, na drodze przetargu nieograniczonego wybierał firmy odpowiedzialne za realizację poszczególnych etapów badania w terenie. Na dwa pierwsze etapy na ogłoszenia przetargowe odpowiedziała – wygrywając je – firma PBS DGA Sp. z o.o..

Tabela 1.3. Alokacja uczniów w próbie głównej

warstwa	liczba i odsetek uczniów w próbie		odsetek w całej populacji
Wieś i miasto do 5 tysięcy mieszkańców	683	13,25%	23,70%
Szkoły jednooddziałowe			
Wieś i miasto do 5 tysięcy mieszkańców	770	14,94%	13,70%
Szkoły dwuoddziałowe			
Wieś i miasto do 5 tysięcy mieszkańców	265	5,14%	4,60%
Szkoły trzyoddziałowe i większe			
Miasto powyżej 5 tysięcy mieszkańców	177	3,43%	4,70%
Szkoły jednooddziałowe			
Miasto powyżej 5 tysięcy mieszkańców	917	17,79%	15,40%
Szkoły dwuoddziałowe			
Miasto powyżej 5 tysięcy mieszkańców	2343	45,45%	37,90%
Szkoły trzyoddziałowe i większe			
Ogółem	5155	100,00%	100,00%

Źródło: opracowanie własne

W niniejszym raporcie analizowane są dane pochodzące z I i II etapu badania. W pierwszym etapie badania wzięło udział 5625 uczniów, w drugim etapie, w testach osiągnięć szkolnych – kluczowej zmiennej w badaniu – 5155 uczniów. W drugim etapie próba szkół pomniejszyła się o cztery jednostki: jedna placówka zrezygnowała z udziału w badaniu, a trzy nie uruchomiły oddziałów klas IV.

## 1.5. Reprezentatywność próby

Istotnymi problemami związanymi z próbą są jej wiarygodność (*validity*) oraz możliwość szacowania wyników dla całej populacji, tj. uogólniania (*generalizability*). Dochowując staranności próbkowania oraz pilnując procedur realizacji badania w terenie, można założyć, że próba spełnia te dwa wymagania. W przypadku omawianym w tym rozdziale dokonano dodatkowo empirycznej weryfikacji tych dwóch założeń. Dysponując wynikami sprawdzianu w klasie VI szkoły podstawowej z roku 2012 dla wszystkich szkół w Polsce, sprawdzono, jak na ich tle wyglądają wyniki szkół, które były wylosowane do badania, w których zrealizowano badanie oraz tych, które odmówiły udziału w badaniu. Ta ostatnia grupa szkół została podzielona na dwie

## CZY SZKOŁA MA ZNACZENIE?

kategorie: szkoły, w przypadku których dyrektor nie udzielił zgody na udział szkoły w badaniu oraz szkoły wykluczone z badania ze względu na bardzo niski odsetek zgód na udział w badaniu wśród nauczycieli oraz rodziców uczniów. Uzasadnieniem dla sprawdzenia rozkładu wyników w wymienionych kategoriach było podejrzenie, że szkoły odmówiły zgody na udział w badaniu ze względu na czynniki powiązane z poziomem osiągnięć uczniów. Próba byłaby wówczas obciążona błędami systemowymi (*selection biased*). Użycie wyników ze sprawdzianu jest uzasadnionym kryterium weryfikacji reprezentatywności próby, gdyż w badaniu SUEK główną zmienną wyjaśnianą są wyniki standaryzowanych testów osiągnięć szkolnych.

Zasadność użycia wyników sprawdzianu w klasie VI do analizy reprezentatywności na poziomie pierwszego etapu edukacyjnego potwierdzają korelacje średnich dla szkół wyników sprawdzianu i średnich dla trzech mierzonych umiejętności w badaniu SUEK. Poniższa tabela pokazuje wartości r-Pearsona.

**Tabela 1.4. Wartości współczynników r-Pearsona**

	Średni wynik testu z matematyki	Średni wynik testu umiejętności czytania	Średni wynik testu świadomości językowej
Średnia ze sprawdzianu dla szkoły w roku 2012	0,499	0,476	0,517

Źródło: opracowanie własne

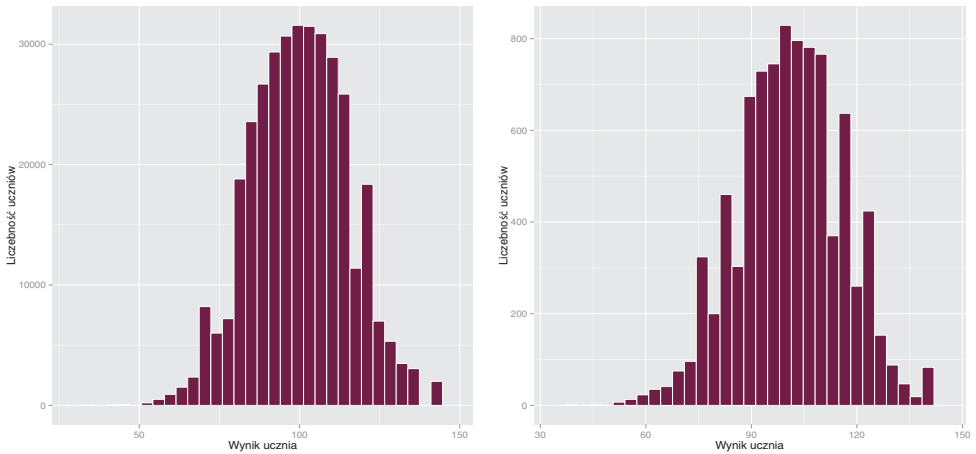
Wszystkie korelacje są znaczące statystycznie.

Pierwszy histogram przedstawia rozkład wyników ze sprawdzianu dla populacji  $N=355318$  uczniów. Wynik sprawdzianu zostały znormalizowane i wystandaryzowane do średniej  $\mu=100$  i odchylenia standardowego  $\sigma=15$ . Rozkład ten jest punktem odniesienia dla analiz reprezentatywności. Przedstawione statystyki dla próby wylosowanej dotyczą 179 ze 180 szkół, dla których udało się połączyć dane.

**Tabela 1.5. Parametry rozkładu wyników sprawdzianu w 2012 r. w populacji i wylosowanej próbie**

	populacja	wylosowana próba
N	355318	9069
Średnia	100,00	100,90
Mediana	100,22	100,22
Odchylenie standardowe	15,00	15,00
Rozstęp ćwiartkowy	20,98	20,98
Skośność	-0,01	-0,02
Kurtoza	-0,06	0,02

Źródło: opracowanie własne

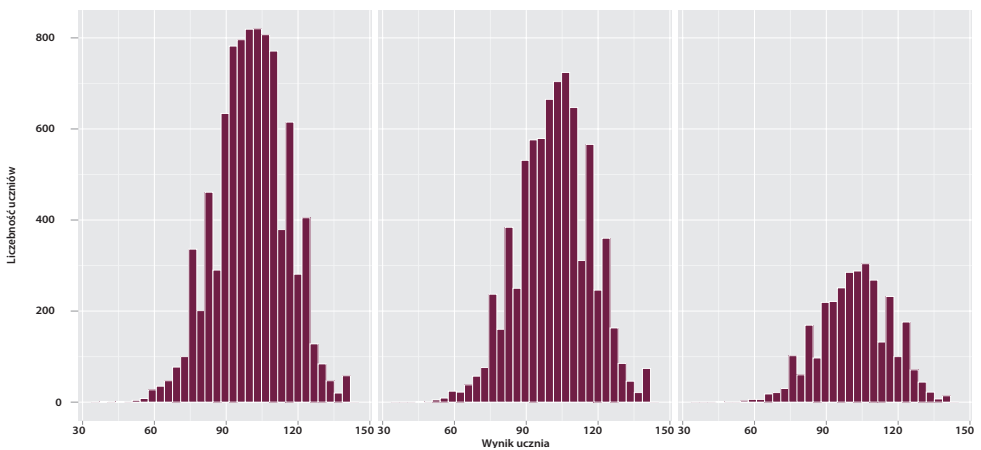


Rysunek 1.2. Rozkład wyników na sprawdzianie w roku 2012 w populacji oraz w wylosowanej próbie.

Zbadano również rozkład wyników dla populacji, z której zostały usunięte szkoły poniżej 11 uczniów – tak jak uczyniono to na etapie losowania próby. Jedyną zmianą jaką zaobserwowano, był nieznaczny wzrost średniej do  $\mu=100,09$  oraz zmiana odchylenia standardowego  $\sigma=15,01$ , przy liczbie uczniów  $N= 335731$ .

Gdyby udało się zrealizować badanie w szkołach wylosowanych w ramach próby głównej przy *response rate* równym 100%, wyniki wyglądałyby jak na histogramie i tabeli powyżej.

Kolejna analiza dotyczy rozkładu wyników sprawdzianu w próbie zrealizowanej.



Rysunek 1.3 Rozkład wyników sprawdzianu w próbach: zrealizowanej, z brakiem zgody dyrektora oraz z niską stopą realizacji badania.

Źródło: opracowanie własne

## CZY SZKOŁA MA ZNACZENIE?

**Tabela 1.6. Parametry rozkładu sprawdzianu 2012 dla próby zrealizowanej i szkół, w których nie zrealizowano badania**

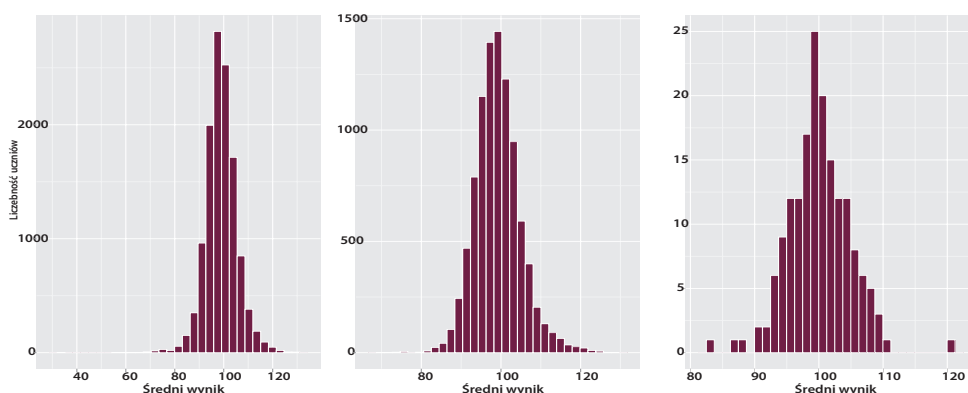
	próba zrealizowana	szkoły, w których odmówił dyrektor	szkoły z niską stopą realizacji badania
N	7947	6688	2563
Średnia	100,30	101,70	101,78
Mediana	100,22	101,91	101,91
Odchylenie standardowe	14,71	15,12	14,87
Rozstęp ćwiartkowy	20,98	19,18	19,18
Skośność	-0,06	-0,05	-0,08
Kurtoza	0,04	-0,01	-0,14

Źródło: opracowanie własne

Rozkład wyników sprawdzianu w próbie zrealizowanej jest prawie identyczny jak w próbie zakładanej, a tym samym w całej populacji uczniów szkół podstawowych w 2012 roku. Obserwujemy jedynie minimalną redukcję wariancji.

Przyglądając się rozkładom szkół, które nie wzięły udziału w badaniu, widzimy, że były to szkoły o wynikach średnio nieznacznie wyższych od średniej ogólnopolskiej. W przypadku szkół, w których dyrektor odmówił udziału w badaniu, średnia jest równa  $\mu=101,70$ , podobnie w szkołach o niskim *response rate*, które wykluczono z tego powodu z badania, średnia wynosi  $\mu=101,78$ .

Podsumowując, próba szkół, na której odbyło się badanie, ma na poziomie indywidualnych wyników uczniów statystyki bardzo zbliżone do populacji, a występujące różnice są nieznaczne.



**Rysunek 1.4. Rozkład średnich wyników sprawdzianu dla szkół w populacji i próbie zrealizowanej.**

Źródło: opracowanie własne

Podobnej procedurze weryfikacji poddano średnie wyników sprawdzianu w klasie VI dla szkół w populacji i próbie zrealizowanej szkół (w drugim etapie badania). Dla badań szkolnych

uwarunkowań efektywności nauczania ten aspekt reprezentatywności próby jest szczególnie ważny.

**Tabela 1.7. Statystyki rozkładu średnich wyników sprawdzianu 2012 dla szkół, wszystkie szkoły, szkoły ponad 10 uczniów i zrealizowana próba**

	wszystkie szkoły	szkoły ponad 10 uczniów	próba zrealizowana
N	12238	9434	172
Średnia	98,99	99,23	99,96
Mediana	98,80	98,87	99,80
Odchylenie standardowe	6,44	5,77	4,78
Rozstęp ćwiartkowy	7,48	6,93	5,65
Skośność	-0,10	0,49	0,27
Kurtoza	3,11	1,52	2,15

Źródło: opracowanie własne

W grupie szkół powyżej 10 uczniów wariancja międzyszkolna jest trochę mniejsza niż w całej populacji podstawówek. W próbie zrealizowanej zróżnicowanie międzyszkolne jest niższe niż w populacji. To niekorzystna cecha próby, na szczęście redukcja wariancji międzyszkolnej nie jest duża.

## 1.6. Grupy uczestniczące w badaniu i najważniejsze narzędzia badawcze<sup>2</sup>

W badaniu SUEK można wyróżnić cztery grupy respondentów, co przedstawia – wraz z wykazem narzędzi i liczebnościami – tabela 1.8.

Do dyrektorów badanych szkół skierowany był kwestionariusz ankiety, poruszający takie zagadnienia jak rekrutacja do szkoły czy sposoby zarządzania nią. Dyrektor proszony był również o opis swoich kwalifikacji z zakresu zarządzania oświatą oraz przekazanie informacji dotyczących swojej ścieżki zawodowej oraz szeregu charakterystyk niezbędnych np. do przygotowania takich wskaźników jak SES rodziny pochodzenia (*socioeconomic status*).

Nauczyciele badanych uczniów (z wyłączeniem nauczycieli wychowania fizycznego, etyki i religii) byli kolejną grupą respondentów. W kwestionariuszu ankiety dla nauczyciela znalazły się m.in.: skala satysfakcji/zadowolenia z życia, skala SDO (Social dominance orientation), skala

<sup>2</sup> Podrozdział autorstwa Leszka Słowickiego.



## CZY SZKOŁA MA ZNACZENIE?

satisfakcji z pracy, samopoczucia w pracy, arkusz oceny pracy (na podstawie testu Anny Zalewskiej) oraz szereg pytań o charakterze metryczkowym. Podobnie jak dyrektorzy szkół, nauczyciele odpowiadali na pytania dotyczące ich ścieżek zawodowych i SES rodziny pochodzenia.

Rodzice badanych uczniów odpowiadali na pytania mające na celu określenie zaplecza kulturowego i zasobności materialnej rodziny oraz mające na celu wyznaczenia wskaźników SES. Dodatkowo rodzice uzupełniali pytania składające się na skalę lęku oraz skalę motywacji (na podstawie „Arkusza zachowania się ucznia” Barbary Markowskiej).

Uczniowie podczas 1. etapu badania uzupełniali Test Matrycy Ravena w wersji STANDARD – forma Klasyczna (TMS-K), Test Inwentarz Stanu i Cechy Lęku dla Dzieci (STAI-C), Test integracji szkolnej KIU – polska adaptacja Grzegorza Szumskiego. W 2. etapie uczniowie rozwiązywali testy sprawdzające cztery obszary osiągnięć szkolnych: czytania i świadomości językowej, kompetencje matematyczne oraz posługiwanie się językiem obcym – angielskim.

Realizacja w terenie każdego etapu była kontrolowana przez pracowników IBE, którzy weryfikowali zgodność przeprowadzania badania przez ankietera bądź psychologa ze standardami opisanymi w procedurach realizacji badania. Również wyznaczony procent materiałów badawczych i protokołów realizacji badania oraz wszystkie bazy danych powstałe po fazie terenowej zostały skontrolowane.

### 1.7. Wagi zastosowane w analizach

W związku z zastosowaniem złożonego schematu doboru próby o zróżnicowanym prawdopodobieństwie wejścia elementów populacji do próby, aby uzyskać prawidłowe oszacowania parametrów w populacji, należy przy ich obliczaniu uwzględnić schemat doboru poprzez zastosowanie wag.

Dla wszystkich szkół, oddziałów i uczniów należących do pierwotnej próby badanych (tj. próby z pierwszego etapu badania), wyliczono warunkowe wagi próbkowania. Zastosowana procedura wyliczania wag była analogiczna jak ta dla dużych badań międzynarodowych (Kennedy, Martin, & Mullis, 2007). Wagi te stanowiły odwrotność prawdopodobieństwa dostania się danej jednostki doboru do próby. Oznacza to, że dla 176 szkół, które wzięły udział w badaniu na pierwszym etapie, wyliczono prawdopodobieństwo dostania się do próby, warunkując po sześciu jawnych warstwach. W ramach każdej ze szkół wyliczono prawdopodobieństwo dostania się oddziału klasowego do badania (waga ta wynosiła 1 dla klas w szkołach jedno i dwu oddziałowych). Każdemu zaś uczniowi w wylosowanej klasie przypisano wagę 1, jako że każdy z uczniów w klasie został zakwalifikowany do badania. Ostateczną wagę ucznia uzyskuje się poprzez przemnożenie przez siebie tych wag. Suma wag ostatecznych stanowi nieobciążony estymator liczebności populacji.

Tabela 1.8. Wykaz narzędzi użytych podczas dwóch pierwszych etapów badania SUEK

etap	grupa respondentów	narzędzie	liczba wszystkich respondentów	liczba respondentów przebadanych	%
1 etap (XI 2010–IV 2011) III klasa szkoły podstawowej	dyrektor	Ankieta dla dyrektora szkoły podstawowej	176	174	98,86%
	nauczyciel	Ankieta dla nauczyciela szkoły podstawowej	1871	1737	92,84%
		Test satysfakcji zawodowej	1871	1744	93,21%
		Ankieta dla nauczyciela języka angielskiego	271	259	95,57%
	uczeń	Test matrycy Ravena	5625	5429	96,52%
		Test Inwentarz Stanu i Cechy Lęku dla Dzieci STAI-C		5433	96,59%
		Badanie socjometryczne		5448	96,85%
		Test integracji szkolnej KIU		5450	96,89%
	rodzic	Ankieta dla rodzica ucznia szkoły podstawowej		5598	99,52%
	2 etap (VI–XI 2011) III i IV klasa szkoły podstawowej	dyrektor	Ankieta dla dyrektora dot. nauczania języków obcych	172	152
nauczyciel		Ankieta dla nauczyciela szkoły podstawowej	2533	2380	93,96%
uczeń		Test czytania	5292	5151	97,34%
		Test świadomości językowej		5151	97,34%
		Test matematyczny		5040	95,24%
		Test z języka angielskiego		4708	88,96%
		Wypowiedź ustna j. angielski	572	572	100,00%
rodzic	Ankieta dla rodzica ucznia szkoły podstawowej	5292	4931	93,18%	

Źródło: opracowanie własne

Niestety, nie wszystkie szkoły, klasy lub uczniowie wzięli udział we wszystkich etapach badania SUEK. Zarówno na pierwszym jak i na kolejnych etapach badania część badanych uczniów nie wzięła udziału w badaniu. Sprawia to, że konieczne staje się uwzględnienie poprawek

na niepełną realizację próby (*non-participation adjustment*). Poprawki te są uwzględniane na każdym etapie doboru – uzyskujemy więc wagi warunkowe z poprawkami dla szkół, klas i uczniów. Iloczyn tych wag to ostateczna waga ucznia z poprawką. Zastosowanie poprawek sprawia, że suma poprawionych wag stanowi wciąż nieobciążony estymator liczebności populacji. W zależności jednak od tego, jakie zmienne są analizowane, różne podzbiory uczniów można traktować jako te, które wzięły udział w badaniu. Dla każdej analizy wyliczano więc wagi z poprawkami, w zależności od zastosowanych w analizach zmiennych. Do automatycznego generowania wag z poprawkami zastosowano skrypt napisany w programie R.

## **Literatura cytowana**

- Biedrzycki, K., Cyngot, D., Federowicz, M., Grzęda, M., Haman, J., Karpiński, Z. L., i inni. (2009). Wyniki Badania OECD PISA 2009 w Polsce. Warszawa: Międzynarodowe konsorcjum realizujące badanie OECD PISA 2009.
- Kennedy, A. M., Martin, M. O., & Mullis, I. V. (2007). PIRLS 2006 Technical Report. Boston: TIMSS & PIRLS International Study Center, Boston College.
- Olson, J., Martin, M., & Mullis, I. (2008). TIMSS 2007 Technical Report. Boston: TIMSS & PIRLS International Study Center, Boston College.
- Schulz, W., & Sibberns, H. (2004). IEA Civic Education Study Technical Report. Amsterdam: IEA.
- Smith, P., Lynn, P., Elliot, D. (2009). Sample Design for Longitudinal Surveys. W: P. Lynn (red). Methodology of Longitudinal Surveys. Wiley.

## Rozdział 2

### Pomiar wyników nauczania w zakresie czytania, świadomości językowej i matematyki<sup>1</sup>

Podstawowym celem badania SUEK jest oszacowanie i następnie wyjaśnienie między-szkolnego i międzyoddziałowego zróżnicowania wyników nauczania. Znalezienie odpowiedzi na pytania o uwarunkowania wyników nauczania wymaga jednak dobrej miary uczniowskich osiągnięć, jakoś kluczowej zmiennej zależnej determinuje bowiem trafność uzyskanych na podstawie analiz wniosków. Oczywiście dotyczy to wszystkich zmiennych uwzględnionych w modelu, niemniej słabszy pomiar zmiennej wyjaśniającej może spowodować obciążenie wyników analiz, przede wszystkim z użyciem tej zmiennej, a niewystarczająca jakoś zmiennej zależnej sprawia, że wątpliwe stają się wszystkie wyniki. Ponieważ wyniki nauczania są kluczową zmienną zależną dla analiz przedstawionych w tym raporcie, jakości tego pomiaru poświęcono dużo miejsca.

Opracowanie dobrych narzędzi do pomiaru osiągnięć szkolnych jest poważnym wyzwaniem. Potwierdzają to doświadczenia międzynarodowych badań edukacyjnych skoncentrowanych na pomiarze umiejętności uczniów (np. PISA, PIRLS, TIMSS). W badaniach tych wiele czasu, uwagi i wysiłku poświęca się konstrukcji narzędzi. Oprócz prac nad definiowaniem mierzonych konstruktów i opracowywaniem zadań, szczególne znaczenie przypisuje się badaniom pilotażowym, które mają na celu zweryfikowanie jakości pozycji testowych i skonstruowanie skal pomiarowych spełniających przyjęte standardy. Badania te pokazują także, jak ważne i zarazem użyteczne jest stosowanie w pomiarze dydaktycznym zaawansowanej metodologii skalowania wyników.

Waga i złożoność procesu konstruowania testów osiągnięć znajduje także swoje odzwierciedlenie w standardach tworzenia i ewaluacji narzędzi do pomiaru dydaktycznego, opisanych w *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999)<sup>2</sup>. Standardy

---

<sup>1</sup> Fragmenty tego rozdziału zostały opublikowane w: Jasińska, A., i Modzelewski, M. (2012). *Można inaczej. Wykorzystanie IRT do konstrukcji testów osiągnięć szkolnych*. [W:] B. Niemierko i M. K. Szmigel (red.), Regionalne i lokalne diagnozy edukacyjne: XVIII Krajowa Konferencja Diagnostyki Edukacyjnej, Wrocław, 21–23 września 2012 r. (157–168). Grupa Tomami. Opisywane tu testy osiągnięć szkolnych zostały, po ich dostosowaniu do nowej podstawy programowej, wykorzystane w innym badaniu podłużnym. Na podstawie wyników obu badań przeanalizowano właściwości psychometryczne testów i opisano je w: Jasińska, A., i Modzelewski, M. (2014). Testy osiągnięć szkolnych TOS3: przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha. *Edukacja* 2(127), 85–107.

<sup>2</sup> Polskie wydanie: AERA, APA, NCME (2007). *Standardy dla testów stosowanych w psychologii i pedagogice*. Gdańsk: GWP.

## CZY SZKOŁA MA ZNACZENIE?

te są stosowane przez najbardziej renomowane i doświadczone instytucje konstruuje testy osiągnięć szkolnych.

W czasie, w którym zespół badawczy SUEK przystępował do realizacji badania podłużnego, nie istniały w Polsce standaryzowane, o udokumentowanej rzetelności i trafności, testy mierzące osiągnięcia szkolne uczniów szkół podstawowych kończących I etap edukacji, które mogłyby zostać wykorzystane do celów badawczych. Nie było także banku zadań, na podstawie którego można by takie testy skonstruować. Dlatego, mając świadomość trudności wyzwania, przystąpiono do realizacji tego zadania. Podczas prac korzystano przede wszystkim z doświadczeń zagranicznych oraz inspirowano się metodologią międzynarodowych badań edukacyjnych.

### 2.1. Podstawowe założenia i cel pomiaru

Opisana w tym rozdziale koncepcja pomiaru wyników nauczania po I etapie edukacyjnym oraz testy osiągnięć spełniające jej założenia zostały opracowane specjalnie na potrzeby *Badania Szkolnych Uwarunkowań Efektywności Kształcenia* realizowanego w Instytucie Badań Edukacyjnych. Badanie to objęło kohortę uczniów rozpoczynających naukę w III klasie szkół podstawowych w roku szkolnym 2010/11. Są to uczniowie nieobjęci jeszcze reformą programową wprowadzoną *Rozporządzeniem Ministra Edukacji Narodowej* z dnia 23 grudnia 2008 r. w sprawie podstawy programowej wychowania przedszkolnego oraz kształcenia ogólnego w poszczególnych typach szkół. Dlatego opracowane testy osiągnięć szkolnych powstały z myślą o uczniach kształconych według tzw. *starej podstawy programowej* opisanej w załączniku nr 2 do *Rozporządzenia Ministra Edukacji Narodowej i Sportu* z dnia 26 lutego 2002 r.

Charakter badania podłużnego SUEK wymagał, by skonstruowane na jego potrzeby testy były testami szerokiego zasięgu, czyli takimi, które pozwalają na zdobycie jak największej ilości informacji o całej badanej grupie<sup>3</sup>. Testy osiągnięć szkolnych musiały być więc tak przygotowane, by niezależnie od poziomu wiedzy uczniów, umożliwiały dostatecznie precyzyjny pomiar. Z uwagi na cele badania SUEK<sup>4</sup>, priorytetem było rzetelne określenie wyników nauczania na poziomie badanych oddziałów i szkół.

Spełnienie tak postawionych celów wymagało dobrania do testu zadań o zróżnicowanej trudności. W praktyce przekłada się to na potrzebę uwzględnienia w teście odpowiedniej liczby zarówno zadań bardzo łatwych (odwołujących się do prostszych umiejętności niż

<sup>3</sup> Innymi typami testów ze względu na stawiane im cele są testy przesiewowe, które pozwalają na precyzyjne stwierdzenie, czy uczeń osiągnął minimalny, ustalony przez ekspertów poziom umiejętności lub testy selekcyjne, które mają na celu jak najdokładniejsze zróżnicowanie wyników uczniów o wysokich umiejętnościach (Jakubowski, Pokropek, 2009).

<sup>4</sup> Efektywność nauczania przypisywana jest bowiem przede wszystkim oddziałowi klasowemu (to w klasie odbywa się nauczanie; wiele gromadzonych danych będą stanowiły zmienne z poziomu klasy), a w dalszej kolejności szkołom.

opanowanie zakłada podstawa programowa z danego poziomu kształcenia), zadań o średnim poziomie trudności, jak i bardzo trudnych (czyli takich, które wykraczają poza określony przez podstawę programową poziom). Należy podkreślić, że testy osiągnięć szkolnych przygotowywane na potrzeby badania SUEK nie mają na celu kryterialnego pomiaru opanowania przez uczniów wiadomości i umiejętności opisanych przez podstawę programową. Podstawa programowa określa cele kształcenia i podstawowy poziom wiadomości i umiejętności, które powinien posiadać każdy uczeń na danym etapie kształcenia. Uczniowie zdolni powinni mieć odpowiednio dostosowywany przez nauczycieli program nauczania. Aby możliwy był precyzyjny pomiar poziomu umiejętności tych uczniów, w testach muszą się znaleźć także zadania wykraczające mierzonymi umiejętnościami poza zakres podstawy programowej. Należy jednak zadbać o to, by nie odwoływały się one do treści, które zwyczajowo pojawiają się dopiero na kolejnych etapach kształcenia, ale by wymagały wykonania bardziej złożonych operacji z wykorzystaniem treści, które powinny być omawiane w szkole na badanym poziomie.

Testy osiągnięć szkolnych były konstruowane w zgodzie z metodologią teorii odpowiedzi na zadanie testowe (*item response theory*, IRT). Podejście IRT umożliwia oszacowanie parametrów psychometrycznych testu bez konieczności testowania całego narzędzia na próbie z populacji docelowej, co było niezbędne w planowanym badaniu pilotażowym. Pozwala także na precyzyjne porównanie wyników różnych wersji testów<sup>5</sup>. Przyjęta teoria pomiaru rzutowała na sposób konstrukcji zadań testowych i schematów oceniania, wybór zadań po badaniu pilotażowym oraz skalowanie ostatecznych wyników z testów.

## **2.2. Charakterystyka zawartości treściowej testów**

Testy osiągnięć szkolnych, przygotowywane na potrzeby badania, koncentrują się na pomiarze najważniejszych z punktu widzenia kształcenia w szkole podstawowej obszarach: umiejętności czytania, świadomości językowej i umiejętności matematycznych. W badaniu SUEK przeprowadzono także pomiar osiągnięć z dominującego w nauczaniu języka obcego: języka angielskiego. Pomiar ten został opisany w osobnym rozdziale.

Wybór skal pomiarowych do testów osiągnięć szkolnych został poprzedzony analizą podstawy programowej kształcenia ogólnego, raportów z krajowych, międzynarodowych i zagranicznych badań umiejętności<sup>6</sup> oraz wniosków płynących z projektu poświęconego nowej

<sup>5</sup> Co miało znaczenie dla przygotowywanego pomiaru z uwagi na planowane skonstruowanie dwóch zakotwiczonych wersji testów.

<sup>6</sup> TIMSS, PIRLS, ECLS, PISA, Badanie umiejętności podstawowych uczniów trzecich klas szkoły podstawowej.

## CZY SZKOŁA MA ZNACZENIE?

formule sprawdzianu dla klasy VI<sup>7</sup>. Opracowana koncepcja testów osiągnięć szkolnych podlegała konsultacjom merytorycznym z ekspertami i została poddana zewnętrznej recenzji. Zebrane uwagi pozwoliły na dopracowanie ostatecznego kształtu koncepcji.

Osiągnięcia szkolne uczniów biorących udział w badaniu SUEK zostały zbadane za pomocą czterech testów:

1. testu umiejętności czytania
2. testu świadomości językowej
3. testu umiejętności matematycznych
4. testu z języka angielskiego (opisanego w oddzielnym rozdziale)

a każdy z nich stanowił odrębną skalę. Ze względu na ograniczoną długość każdego testu, nie zakładano tworzenia podskal przedstawiających wyniki uczniów w ramach bardziej szczegółowych umiejętności.

Wybór wymienionych skal pomiarowych nie jest podyktowany przeświadczeniem, że wyniki kształcenia można ograniczyć do tych czterech obszarów umiejętności. Niemniej nie sposób objąć badaniem wszystkich efektów kształcenia, które osiągają lub osiągać powinny szkoły na określonym etapie edukacji. Jednak spośród poznawczych wyników nauczania konieczny jest wybór tych kluczowych, w zakresie których chce się wyciągać wnioski, gdyż każde badanie wymaga dokonania pewnej redukcji.

Poniżej opisano strukturę każdego z trzech testów. Na podstawie założonej struktury opracowane zostały plany testów dla każdej skali pomiarowej, które precyzowały, ile zadań, mierzących jakie szczegółowe umiejętności, powinno znaleźć się w teście. Opis struktury testów oraz wynikających z niej planów testów miał na celu zagwarantowanie różnorodności i reprezentatywności mierzonych treści i umiejętności, a tym samym za pewnienie trafności treściowej testu.

### 2.2.1. Test umiejętności czytania

Test ten miał na celu pomiar rozumienia znaczenia czytanych samodzielnie poleceń, zadań i tekstów różnego typu; nie obejmował techniki czytania.

Zadania składające się na test umiejętności czytania zostały opisane za pomocą dwóch wymiarów: **typ materiału testowego**, który należy przeczytać oraz **umiejętności**, które trzeba wykorzystać, rozwiązując dane zadanie.

<sup>7</sup> Nowa formuła sprawdzianu w klasie VI, projekt realizowany w latach 2007–2010 przez CKE, koordynator: Anna Pregler.

**Typ materiału testowego.** Test umiejętności czytania składa się z kilku tekstów oraz wiązek pytań do każdego z nich. Są to teksty różnorodnego typu: literackie (prozatorskie i poetyckie), popularnonaukowe i użytkowe (regulamin, ulotka, ogłoszenie). Niektóre z nich nie mają tradycyjnego charakteru tekstu ciągłego, ale zawierają dodatkowe ramki, przypisy.

**Umiejętności.** Zadania do każdego tekstu wymagają odwołania się do różnych umiejętności:

1. Wyszukiwania informacji – definiowanej jako umiejętność odszukania potrzebnej informacji w tekście głównym oraz uzupełniającym (tabeli, ramce, przypisie), podjęcia decyzji, które informacje są ważne, a które nie ze względu na ich związek z tematem lub pytaniem, a także ustalenia kolejności wydarzeń w tekście.
2. Interpretacji – jest to umiejętność zrozumienia tekstu jako całości, wydobycia i wyjaśnienia jego sensu. Od ucznia oczekuje się, że będzie potrafił określić temat i główną myśl utworu (rozumienie tekstu na poziomie podstawowym – o czym tekst jest), jak również, że będzie rozumiał przenośne (ukryte) znaczenie utworu. Umiejętność ta odnosi się także do rozumienia związków między zawartymi w tekście informacjami, porównywania informacji zawartych w tekście, dostrzegania i wyjaśniania przyczyn i skutków zdarzeń, sytuacji, zjawisk opisanych w tekście, podawania przypuszczalnych motywów działania, zachowania lub postawy bohaterów, rozróżniania informacji nacechowanych emocjonalnie w określonym celu (np. reklamowym) od informacji emocjonalnie obojętnych, a także określania zamiaru i intencji autora.
3. Refleksji i oceny – umiejętność odniesienia tekstu do własnego doświadczenia i wiedzy o świecie; odróżnienia wydarzeń rzeczywistych od fikcyjnych; oceny wiarygodności podanych informacji. Od ucznia oczekuje się, że będzie umiał dokonać oceny zdarzeń, postaci, stanowisk, poglądów w kontekście własnego doświadczenia czytelniczego i pozaszkolnego, jak również wyrazić opinię na temat myśli zawartych w utworze. Jest to także umiejętność oceny informacji o charakterze reklamowym, oceny szaty graficznej czytanych tekstów oraz zastosowanych zabiegów literackich, a także oceny kompletności i spójności tekstu.

### **2.2.2. Test świadomości językowej**

Opisując strukturę testu mierzącego umiejętności językowe uczniów, należy omówić problem wynikający z pomiaru umiejętności pisania za pomocą zadań polegających na samodzielnym tworzeniu tekstów na zadany temat, gdyż świadomość jego istnienia zadecydowała o kształcie tego testu.

Kształtowanie umiejętności pisania i świadomości językowej ma na celu rozwój umiejętności samodzielnego tworzenia tekstu w sposób czytelny, logiczny, zgodny z zasadami gramatyki i ortografii. Naturalne wydaje się więc przyjęcie założenia, że poziom opanowania



## CZY SZKOŁA MA ZNACZENIE?

tych umiejętności uczeń najlepiej zademonstruje, pisząc samodzielnie tekst na zadany temat. Niestety, mimo iż umiejętność pisania i świadomość językowa są rozwijane przez pisanie własnych tekstów, pomiar osiągnięć szkolnych w tym zakresie wymaga spojrzenia z innej perspektywy. Wykorzystanie w teście zadań polegających na samodzielnym napisaniu tekstu na zadany temat pociąga za sobą co najmniej dwa poważne problemy.

Pierwszy jest związany z rzetelnością oceniania tego typu zadań i polega na istnieniu znaczących różnic w ocenianiu uczniowskich wypowiedzi pisemnych między egzaminatorami (tzw. efekt egzaminatora). Wyniki badań poświęconych tej problematyce (Dolata, Putkiewicz, Wiłkomirska, 2004), a także doświadczenia własne nie pozwalają na zbagatelizowanie tego problemu<sup>8</sup>.

Drugi problem dotyczy zaś liczby zadań w teście (decydującej o dokładności pomiaru) i czasu potrzebnego na jego rozwiązanie. Zastosowanie w teście jednego zadania wymagającego napisania dłuższej wypowiedzi, daje mniejszą precyzję pomiaru niż kilka lub kilkanaście krótszych zadań, które uczeń może rozwiązać w tym samym czasie.

Z uwagi na konsekwencje przedstawionych powyżej problemów, zdecydowano nie włączać do testu zadań dłuższej wypowiedzi, polegających na napisaniu tekstu na zadanych temat. W zamian za to uwzględniono w teście zadania sprawdzające szczegółowe umiejętności związane z pisaniem tekstu i świadomością językową, mające formę zadań zamkniętych lub otwartych krótkiej odpowiedzi.

Test świadomości językowej mierzy zatem wiadomości i umiejętności umożliwiające refleksję nad językiem jako narzędziem komunikowania się: umiejętności tworzenia tekstu pisanego zgodnie z zasadami ortografii i gramatyki języka polskiego, o czytelnej strukturze oraz w zgodzie z zasadą jasnego przekazywania myśli i zasadami logiki, a także umiejętności w zakresie argumentowania oraz zasób słownikowy dziecka.

Test świadomości językowej składa się z trzech części (bloków zadań):

1. zadań sprawdzających umiejętności związane z pisaniem tekstów
2. zadań sprawdzających zasób słownikowy
3. zadań sprawdzających elementy wiedzy o języku.

Wśród zadań mierzących umiejętności związane z pisaniem tekstów znalazły się zadania sprawdzające umiejętność redagowania tekstu, dzielenia wypowiedzi na zdania, rozpoznawania i tworzenia czytelnej struktury tekstu, rozpoznawania i nadawania poprawnego stylu wypowiedzi oraz umiejętności argumentowania.

<sup>8</sup> Wyniki i badania pilotażowego testów osiągnięć szkolnych w szkole podstawowej prowadzonego w ramach projektu „Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD)” współfinansowanego ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego pokazały znaczące różnice w ocenie prac między parą egzaminatorów sprawdzających te same wypowiedzi pisemne uczniów.

Grupa zadań mająca na celu pomiar zasobu słownikowego uczniów objęła zadania polegające na utworzeniu lub wybraniu spośród podanych wyrazów wyrazu o podobnym lub przeciwstawnym znaczeniu, zadania na dobranie poprawnej definicji lub wyjaśnienia podanego wyrazu lub wpisanie odpowiedniego wyrazu do podanej definicji, zadania na tworzenie, rozpoznawanie lub wyjaśnianie powszechnie występujących porównań i związków frazeologicznych, a także zadania na rozpoznanie niepoprawnego użycia słowa w zdaniu ze względu na jego znaczenie.

Na część testu mającą na celu sprawdzenie elementów wiedzy o języku składały się zadania mierzące umiejętność tworzenia i uzupełniania zdań zgodnie z zasadami składni, wyróżniania w tekście i tworzenia zdań oznajmujących, pytających, rozkazujących, a także zadania z zakresu ortografii i interpunkcji oraz poprawności językowej.

### **2.2.3. Test umiejętności matematycznych**

Zadania składające się na test umiejętności matematycznych zostały opisane za pomocą dwóch głównych wymiarów: **treści matematycznych**, do których należy się odwołać, rozwiązując zadanie oraz **umiejętności**, którymi należy się posłużyć. Dodatkowym wymiarem jest **kontekst**, w którym umieszczony jest problem matematyczny.

**Treści matematyczne.** Skala umiejętności matematycznych obejmuje trzy obszary treści:

1. **Ilość** – obszar ten odnosił się do rozumienia przez uczniów pojęcia liczby, rozumienia i odkrywania relacji między liczbami, umiejętności wykonywania obliczeń oraz rozumienia znaczenia tych operacji, a także umiejętności wykorzystania opisanych kompetencji w sytuacjach praktycznych. W obszarze tym mieściły się także zagadnienia związane z pomiarem. Odwoływały się one do rozumienia problematyki długości, ciężaru, objętości, temperatury i czasu.
2. **Przestrzeń i kształt** – na obszar ten składały się sytuacje geometryczne oraz związki przestrzenne między obiektami. Obejmował on umiejętność rozpoznawania i rysowania figur geometrycznych, dostrzegania symetrii i regularności oraz wyobraźnię przestrzenną (zadania te nie wymagały zastosowania wiedzy formalnej z geometrii).
3. **Zmiana i związki** – zakres tego obszaru dobrze oddaje cytat z *Podstawy programowej kształcenia ogólnego*: „czynny udział w zdobywaniu wiedzy matematycznej przybliży dziecko do matematyki, rozwija kreatywność, umożliwia samodzielne odkrywanie związków i zależności; duże możliwości samodzielnych obserwacji i działań stwarza geometria, ale także w arytmetyce można znaleźć obszary, gdzie uczeń może czuć się odkrywcą”<sup>9</sup>. Na obszar ten składa się więc umiejętność dostrzegania

<sup>9</sup> Podstawa programowa z komentarzami. Tom 6. Edukacja matematyczna i techniczna w szkole podstawowej, gimnazjum i liceum, „Zalecane warunki i sposób realizacji”, s. 34.

## CZY SZKOŁA MA ZNACZENIE?

przez ucznia związków i zależności, reprezentowanych w sposób graficzny, słowny, tabelaryczny lub symboliczny.

**Umiejętności.** Zadania odwołujące się do każdego z trzech obszarów treści mierzyły różne grupy umiejętności:

1. **Odtwarzania wiadomości i dobrze wyćwiczonych schematów**, którą zdefiniowano jako umiejętność rozwiązywania zadań typowych, które wymagają użycia wyćwiczonych, prostych technik i posłużenia się dobrze znanymi procedurami, inaczej mówiąc przywołania z pamięci pewnych pojęć lub algorytmów.
2. **Powiązania wiadomości i dobrze wyćwiczonych schematów**, potrzebnej do rozwiązania zadań mniej rutynowych, ale niezbyt odległych od zadań typowych. Uczeń musi zwykle wykonać większą liczbę kroków, aby rozwiązać zadanie. Musi wybrać pojęcia (modele, wzory, procedury) matematyczne odpowiednie dla rozwiązania danego problemu. Od ucznia oczekuje się, że będzie potrafił wykorzystać posiadane wiadomości do rozwiązania zadań, z których nie wynika wprost, jakie pojęcia czy procedury powinno się zastosować, że będzie umiał porównać ze sobą informacje (zarówno podane w zadaniu, jak i przywołane z pamięci) i zinterpretować je w celu znalezienia rozwiązania.
3. **Rozumowania** matematycznego składającego się z kilku kroków. Od ucznia oczekuje się, że będzie umiał ustalić kolejność czynności prowadzących do rozwiązania problemu (sytuacji nowej, nieoczywistej dla osoby rozwiązującej test), że będzie potrafił wyciągać wnioski z kilku informacji podanych w różnej postaci.

**Kontekst.** Uwzględnienie różnych kontekstów problemów matematycznych umożliwia zachowanie właściwej proporcji między zadaniami typowo szkolnymi, zadaniami odwołującymi się do osobistych doświadczeń ucznia oraz tych osadzonych w kontekście bardziej odległym dla dziecka, czyli przestrzeni publicznej.

Zadania z testu umiejętności matematycznych odwołują się do trzech kontekstów:

1. **Osobisty** – jest to kontekst najbliższy uczniom. Obejmuje sytuacje związane z życiem codziennym ucznia lub jego rodziny.
2. **Szkolny** – kontekst ten dotyczy sytuacji często spotykanych w szkole na lekcjach.
3. **Publiczny** – kontekst ten jest związany m.in. z życiem społeczności lokalnej (np. sąsiedzi, komunikacja, bank, urząd, punkty usługowe) i społeczeństwa (wybory, zabytki, ochrona środowiska, praca zawodowa).

### **2.3. Konstrukcja testów osiągnięć szkolnych**

Skonstruowanie trafnych i rzetelnych testów osiągnięć szkolnych jest procesem wieloetapowym (Downing, 2006a). Prace nad testami osiągnięć szkolnych wykorzystanymi w badaniu SUEK trwały półtora roku. Rozpoczęły się od opracowania przedstawionej wyżej koncepcji pomiaru formułującej ogólne założenia i definiującej skale pomiarowe. Przyjęta koncepcja pozwoliła na zaprojektowanie specyfikacji zawartości testów, czyli tzw. planów testów, które precyzowały ile zadań, mierzących jakie szczegółowe umiejętności, powinno znaleźć się w teście.

Zadania do testów były tworzone przez osoby wyłonione w drodze konkursu. Autorzy byli zobowiązani do układania zadań zgodnych z zapotrzebowaniem wynikającym z planów testów oraz sformułowanymi uprzednio zasadami konstrukcji zadań testowych i schematów oceniania (por. Haladyna, Downing, Rodriguez, 2002; Downing, 2006b). Praca autorów była na bieżąco recenzowana, a pula ostatecznie przyjętych zadań podlegała dodatkowej ocenie i korekcie podczas warsztatów z udziałem matematyków, polonistów, pedagogów wczesnoszkolnych, dydaktyków praktyków, koderów oraz członków zespołu badawczego. Etap ten zakończyło opracowanie graficzne 823 przygotowanych w ten sposób zadań.

Kolejnym bardzo ważnym etapem było zaplanowanie i zrealizowanie badania pilotażowego przygotowanych zadań, które miały utworzyć trzy skale pomiarowe. Badanie pilotażowe musiało zostać zaprojektowane tak, żeby parametry psychometryczne zadań składających się na jedną skalę pomiarową zostały przedstawione na wspólnej skali, dzięki czemu możliwy był trafny wybór zadań do ostatecznych wersji testów. Należało także zadbać o jak najdokładniejsze oszacowanie parametrów psychometrycznych zadań i każdego z trzech testów. Spełnienie tych założeń osiągnięto przez opracowanie adekwatnego do nich planu testowania, precyzującego, które zadania będą rozwiązywane przez którego ucznia. Zadania testowe podzielono na 44 zeszyty testowe, a plan testowania został ostatecznie przygotowany tak, by każdy zeszyt testowy występował z każdym innym dla ustalonej grupy uczniów. Równoważył także prawdopodobieństwo rozwiązywania w próbie każdego zeszytu testowego oraz uwzględniał różną kolejność rozwiązywania zeszytów testowych w poszczególnych dniach testowania. Badanie zostało zrealizowane na ogólnopolskiej losowej próbie 80 szkół podstawowych w klasach III i V (łącznie przebadano 281 klas, co dało 5454 uczniów) w roku szkolnym 2010/11. Okienko testowe zostało wyznaczone na początek II semestru<sup>10</sup>.

---

<sup>10</sup> Harmonogram pracy nad testami, niestety, nie pozwolił na realizację badania pilotażowego w czasie najbardziej wskazanym – na dokładnie rok przed planowanym pomiarem zasadniczym (czyli na początku roku szkolnego 2010/11). Dlatego zdecydowano się objąć badaniem uczniów trochę młodszych oraz trochę starszych, ze świadomością tego, że nie jest to rozwiązanie idealne.

## CZY SZKOŁA MA ZNACZENIE?

Analiza danych z badania pilotażowego miała na celu wybór zadań o najlepszych właściwościach pomiarowych. W analizach przyjęto model Rascha. Analizy wykonano za pomocą oprogramowania ConQuest 2.0 (Wu, Adams, Wilson & Haldane, 2007). Decyzje o wartości psychometrycznej zadań podejmowano na podstawie miar dopasowania zadania do modelu (INFIT – *weighted fit*, OUTFIT – *unweighted fit*), trudności zadania, wniosków płynących z analizy dystraktorów, wniosków płynących z analiz efektów zróżnicowanego funkcjonowania zadania DIF (*differential item functioning*). Na podstawie banku zadań o wystarczająco dobrych właściwościach pomiarowych, czyli dopasowanych do założonego modelu oraz przyjętych założeń skal pomiarowych, skonstruowano testy do badania zasadniczego. Etap ten polegał na wyborze zadań o najlepszych właściwościach psychometrycznych, które mierzą założone przez plany testów wiadomości i umiejętności oraz są jak najlepiej dostosowane do prognozowanego rozkładu umiejętności uczniów w populacji docelowej, a także mają odpowiednio zróżnicowaną trudność<sup>11</sup>. Ustalenie ostatecznej puli zadań tworzących test osiągnięć, pozwoliło na skonstruowanie dwóch równoległych wersji testów z pulą zadań kotwiczących. Wersje te tworzone w taki sposób, by zarówno zadania kotwiczące, jak i zadania w każdej wersji testu stanowiły reprezentatywną próbkę planu testu, a także, by obie wersje miały taką samą trudność<sup>12</sup>.

### 2.4. Opis zawartości treściowej ostatecznej wersji testów

Testy osiągnięć szkolnych przygotowane na potrzeby badania SUEK są testami papierowymi, dostosowanymi do badania audytoryjnego. Testy mają dwie równoległe wersje, z pulą 15–16 zadań wspólnych (kotwiczących) dla obu wersji. Zastosowanie dwóch wersji testów wynika z dwóch przyczyn. Pierwszą jest zwiększenie długości testu<sup>13</sup>, co pozwala na zwiększenie trafności pomiaru przez lepsze pokrycie zadaniami szerokiego spektrum treści i umiejętności nauczanych na danym etapie kształcenia, czyli lepsze oszacowanie wyników nauczania na poziomie klasy. Drugą, utrudnienie ściągania wśród uczniów podczas badania testowego, co podnosi wiarygodność uzyskanych wyników. Zarówno zadania kotwiczące, jak i zadania w każdej wersji testu są reprezentatywną próbką planu testu pod względem mierzonych treści i umiejętności, a obie wersje dla każdego testu mają porównywalną trudność.

<sup>11</sup> Aby krzywa informacyjna testu była jak najwyższa w jak najszerszym obszarze poziomu umiejętności uczniów.

<sup>12</sup> Procedura konstrukcji testów osiągnięć szkolnych wykorzystanych w badaniu SUEK została bardziej szczegółowo opisana w: Jasińska, Modzelewski (2012).

<sup>13</sup> W raporcie tym termin test stosuje się do wszystkich zadań składających się na obie wersje testu, a nie do jednej wersji testu. Dlatego na długość testu składają się wszystkie zadania wchodzące w skład obu wersji testu (nie licząc podwójnie zadań wspólnych), czyli wszystkie zadania wchodzące w skład skali pomiarowej.

## POMIAR WYNIKÓW NAUCZANIA W ZAKRESIE CZYTANIA, ŚWIADOMOŚCI JĘZYKOWEJ I MATEMATYKI

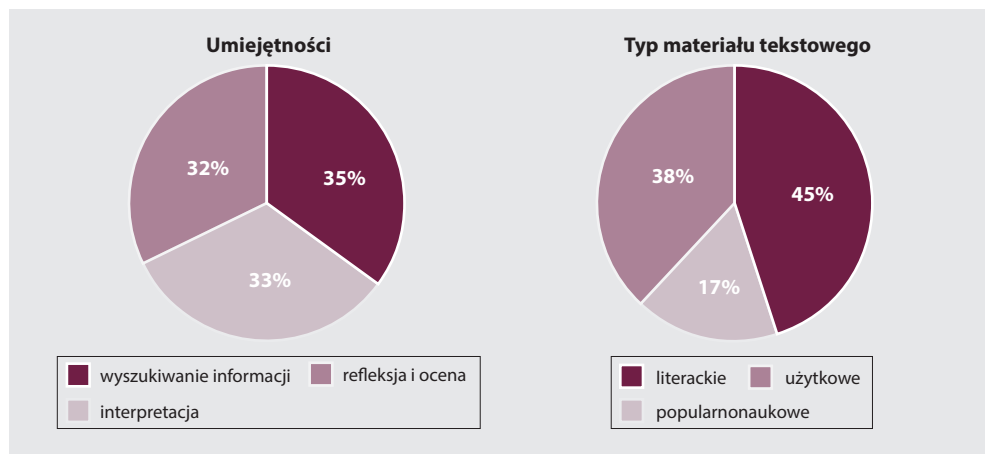
**Tabela 2.1. Liczba zadań w testach osiągnięć szkolnych**

	Liczba zadań w teście	Liczba zadań w każdej wersji testu	Liczba zadań kotwiczących
Test umiejętności czytania	51	33	15
Test świadomości językowej	43	29 i 30	16
Test umiejętności matematycznych	53	34 i 35	16

Źródło: opracowanie własne

Na każdy z trzech testów składa się od 43 do 53 zadań (od 29 do 35 zadań na każdą wersję). Zadania są pogrupowane w 12 zeszytów testowych (po 6 na każdą wersję). Zadania z testu umiejętności matematycznych zostały wydzielone do odrębnych zeszytów (po 2 na każdą wersję), natomiast zadania z testu umiejętności czytania i testu świadomości językowej umieszczono we wspólnych zeszytach (po 4 na każdą wersję) z uwagi dłuższy czas potrzebny do przeczytania tekstów z testu umiejętności czytania. Każdy badany uczeń rozwiązywał więc 6 zeszytów testowych<sup>14</sup>, po 2 jednego dnia testowania, a na rozwiązanie jednego zeszytu miał 35 minut.

Zadania wchodzące w skład poszczególnych testów mierzyły wiadomości i umiejętności zgodne z koncepcją skal pomiarowych. Poniżej przedstawiono podsumowanie struktury każdego testu, prezentując procentowy rozkład zadań mierzących określone w koncepcji grupy umiejętności.



**Rysunek 2.1. Podsumowanie struktury testu umiejętności czytania.**

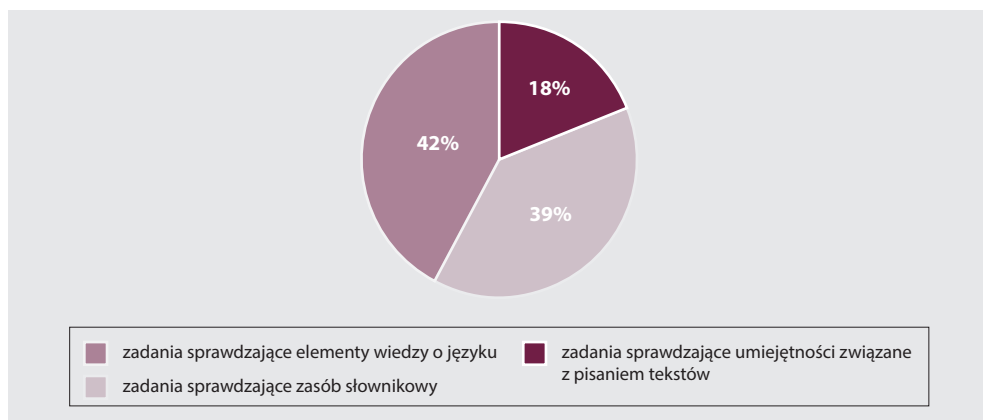
Źródło: opracowanie własne

W teście umiejętności czytania ok. jedna trzecia zadań mierzy każdą z opisanych w koncepcji umiejętność: wyszukania informacji, interpretacji oraz refleksji i oceny. W teście tym

<sup>14</sup> Jeśli był nieobecny podczas którejś sesji testowej, mógł rozwiązać ich mniej.

## CZY SZKOŁA MA ZNACZENIE?

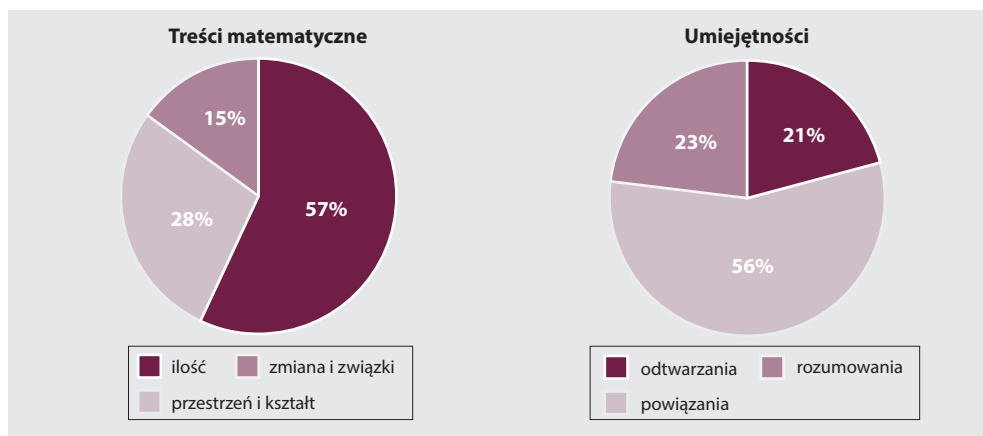
dominują zadania wymagające przeczytania tekstów literackich i użytkowych. Natomiast 17% zadań sprawdza umiejętność czytania tekstów popularnonaukowych.



Rysunek 2.2. Podsumowanie struktury testu świadomości językowej.

Źródło: opracowanie własne

Na test świadomości językowej składa się prawie 20% zadań sprawdzających umiejętności związane z pisanem tekstów oraz po ok. 40% zadań mierzących zasób słownikowy i elementy wiedzy o języku.



Rysunek 2.3. Podsumowanie struktury testu umiejętności matematycznych.

Źródło: opracowanie własne

Najwięcej zadań z testu umiejętności matematycznych odwołuje się do treści z kategorii „ilość”. Jest to umotywowane tym, że na I etapie edukacyjnym treści te są najliczniej

reprezentowane w programach nauczania. W zakresie mierzonych umiejętności najwięcej zadań w teście matematycznym wymaga powiązania wiadomości i dobrze wyćwiczonych schematów. Więcej niż jedna piąta zadań sprawdza umiejętność rozumowania. Podobny odsetek mierzy umiejętność odtwarzania wiadomości i dobrze wyćwiczonych schematów.

W zakresie pomocniczego wymiaru testu matematycznego, jakim był kontekst, nie udało się zrealizować w pełni założeń koncepcji dotyczących pożądanych proporcji zadań umieszczonych w opisanych kontekstach. Zakładano, że najwięcej zadań powinno odwoływać się do osobistych doświadczeń ucznia (kontekst osobisty), a najmniej do zdarzeń najodleglejszych (kontekst publiczny). Autorom zadań testowych trudno było jednak wyjść poza kontekst szkolny, stąd wiele dobrych zadań było osadzonych właśnie w tym kontekście. Ostatecznie 28% zadań testu matematycznego umieszczonych jest w kontekście osobistym, 59% zadań odwołuje się do kontekstu szkolnego, a 13% zadań – do kontekstu publicznego.

## **2.5. Model skalowania testów**

Zarówno na etapie pilotażu jak i badania zasadniczego, do analizy wyników testów zastosowano metodologię skalowania *Item Response Theory* (pol. teoria odpowiedzi na zadanie testowe). Metodologia ta jest stosowana powszechnie w międzynarodowych badaniach osiągnięć uczniów (takich jak PIRLS, TIMSS czy PISA) do skalowania wyników testów osiągnięć szkolnych. Jedną z głównych jej zalet jest możliwość ulokowania zarówno poziomu umiejętności uczniów, jak i trudności zadań na jednej skali, a przez to modelowanie prawdopodobieństwa udzielenia przez ucznia prawidłowej odpowiedzi na zadanie w zależności od jego/jej poziomu umiejętności oraz parametrów zadania. Modele konstruowane w ramach IRT są więc modelami probabilistycznymi, a nie deterministycznymi.

Zdecydowano się na najbardziej restrykcyjny model IRT, czyli model Rascha (Rasch, 1980), a dokładniej, na aplikację tego modelu dla zadań ocenianych dychotomicznie (tj. zadań wielokrotnego wyboru z jedną poprawną odpowiedzią oraz zadań kodowanych 0–1) oraz dla zadań z trzema lub więcej kategoriami odpowiedzi (tzw. model *Partial Credit*). Szczegółowa specyfikacja tych modeli przedstawiona jest na przykład w: van der Linden i Hambleton (1997). Model Rascha zakłada, że wszystkie zadania mają tak samo dobre właściwości pomiarowe (de Ayala, 2009, s. 33–34), a różnić się między sobą mogą jedynie parametrem trudności. Na prawdopodobieństwo udzielenia przez ucznia poprawnej odpowiedzi na dane zadanie ma wpływ tylko różnica poziomu umiejętności ucznia oraz trudności zadania.

Model *Partial Credit* pozwala na zastosowanie tej samej logiki do zadań z wieloma uporządkowanymi kategoriami odpowiedzi (np. de Ayala, 2009, s. 163–169, Masters i Wright, 1997).



## CZY SZKOŁA MA ZNACZENIE?

W tym przypadku poszczególnym kategoriom odpowiedzi przypisane są parametry trudności, które warunkują prawidłową odpowiedź na zadania.

Stworzenie testów z zadań spełniających założenie modelu Rascha jest niezwykle istotne z punktu widzenia komunikowania wyników w postaci informacji o osiągnięciu pewnego poziomu wiadomości i umiejętności<sup>15</sup>, a także tworzenia tablic przeliczeniowych wyniku surowego na wynik przeskalowany<sup>16</sup>, co umożliwi wykorzystywanie testów w diagnozie nauczycielskiej i porównanie otrzymanych wyników z określonymi normami.

Każdy z trzech testów skalowany był osobno, co oznacza, że oszacowania trudności dla zadań, jak również oszacowania poziomów umiejętności dla uczniów, w ramach poszczególnych skal pomiarowych liczone były niezależnie od siebie. Za metodę estymacji parametrów zadań przyjęto *Marginal Maximum Likelihood*. Metoda ta pozwala na oszacowanie parametrów zadań niezależnie od parametrów uczniów oraz opiera się na założeniu, że cecha latentna przyjmuje rozkład normalny w populacji (de Ayala, 2009). Logika, która stoi za takim rozwiązaniem, opiera się na założeniu, że dla oszacowania parametrów dla zadań jest więcej informacji (w przypadku testów stworzonych dla badania SUEK, każde zadanie zostało rozwiązane przez ponad 2500 uczniów) niż dla uczniów (patrz tabela 1.). W związku z tą własnością przyjętej metody, estymacja parametrów dla zadań i uczniów odbywała się w dwóch krokach: najpierw oszacowano parametry dla zadań, a następnie, z wykorzystaniem tak otrzymanych parametrów zadań, wyestymowano parametry dla uczniów. Skalowanie wykonano za pomocą oprogramowania ConQuest 2.0 (Wu i in., 2007).

### 2.5.1. Kalibracja zadań

Podczas estymowania parametrów zadań wykorzystano informacje o wykonaniach pochodzące od wszystkich uczniów, którzy podeszli do testów. Należy jednakże zwrócić uwagę na fakt, że niektóre zadanie zostało rozwiązane tylko przez podpróbkę uczniów, a więc informacja o danym zadaniu jest dostępna tylko dla części uczniów. Taki stan rzeczy wynika zarówno z konstrukcji testów i planu testowania – każdy uczeń rozwiązywał tylko zadania z jednej z dwóch wersji testu, ale także ze względów związanych z realizacją badania: część uczniów mogła nie być obecna podczas któregoś z dni testowania w szkole. Istnieje także jeszcze jedno źródło braku informacji o rozwiązaniu zadania, z którego należy zdać sobie sprawę – uczeń podczas rozwiązywania testu mógł pominąć konkretne zadanie, nie zostawiwszy tym samym informacji o tym, czy jest w stanie je rozwiązać, czy też nie. Ten ostatni przypadek może wynikać z różnych przyczyn (np. z niskiej

<sup>15</sup> Jest to możliwe dzięki temu, że w modelu Rascha uporządkowanie pytań ze względu na trudność jest takie samo dla każdego ucznia (każdego poziomu umiejętności), ponieważ krzywe charakterystyczne zadań się nie przecinają, a także umiejętności uczniów i trudności pytań są wyrażone na tej samej skali.

<sup>16</sup> Surowa liczba punktów jest statystyką dostateczną (de Ayala, 2009, s. 25).

motywacji do rozwiązania testów), jednakże dwie możliwości wydają się być najbardziej prawdopodobne: uczeń mógł pominąć zadanie, gdyż wydało mu się za trudne do rozwiązania lub mógł nie zdążyć go rozwiązać w czasie przewidzianym na sesję testową. Obie te przyczyny wiążą się z poziomem umiejętności ucznia, jednakże pierwsza dopuszcza możliwość, że uczeń po zapoznaniu się z danym zadaniem sam oszacował swój poziom umiejętności jako niewystarczający do jego rozwiązania, druga z kolei pozwala nam domniemywać, że uczeń nie miał możliwości zmierzenia się z zadaniem. W związku z tym, że podczas kalibracji zadań odsetek poprawnych i niepoprawnych odpowiedzi ma duże znaczenie dla estymowanych parametrów, rozróżnienie tych dwóch okoliczności na etapie kalibracji zadań, wydaje się zasadne.

Idąc śladem wytyczonym przez takie międzynarodowe badania edukacyjne jak TIMSS, PIRLS czy PISA, w badaniu SUEK wyróżniono następujące kategorie braków odpowiedzi na zadanie testowe:

- a. Zadanie niezaadministrowane: w tej kategorii mieszczą się te przypadki braków odpowiedzi, które wynikają z fizycznego niezaadministrowania zadania uczniowi (uczeń był nieobecny podczas sesji testowej lub zadanie znajdowało się w zeszytcie testowym nie rozwiązywanym zgodnie z planem przez ucznia). Brak odpowiedzi na ten typ zadania traktowany jest zawsze jako brak danych.
- b. Zadanie pominięte: w tej kategorii znajdują się przypadki braków odpowiedzi na te zadania, na które uczeń mógł odpowiedzieć, ale tego nie uczynił.
- c. Zadania nie rozwiązywane: jest to ciąg zadań pominiętych, który znajduje się na końcu zeszytu testowego, z zastrzeżeniem, że pierwsze zadanie z tego ciągu traktowane jest jako zadanie pominięte.

Podczas kalibracji zadań, braki odpowiedzi z kategorii c traktowane były jako zadania niezaadministrowane (przekodowywane były na braki danych), natomiast braki odpowiedzi z kategorii b traktowane były jako odpowiedzi błędne.

### **2.5.2. Szacowanie wyników uczniów**

Posiadając oszacowania parametrów dla zadań, przystąpiono do estymacji wyników dla uczniów. W tym kroku zadania pominięte i nierozwiązywane traktowane były jako odpowiedzi błędne.

W badaniu SUEK zdecydowano się na oszacowanie dwóch rodzajów estymatorów umiejętności uczniów: punktowego estymatora EAP (*expected a posteriori*) oraz zestaw pięciu imputacyjnych estymatorów, tzw. *plausible values* (PV). Opis tych estymatorów można znaleźć np. w: Adams (2005), Wu (2005). Oba rodzaje estymatorów związane są z bayesowskimi metodami statystycznymi i pojęciem rozkładu *a posteriori* dla oszacowania umiejętności danego ucznia. Wartość oczekiwana tego rozkładu jest estymatorem EAP dla danego ucznia. *Plausible*

## CZY SZKOŁA MA ZNACZENIE?

*values* z kolei stanowią losową próbkę  $n$  wartości z tego rozkładu. Estymatory EAP posiadają tę własność w ramach bezwarunkowego modelu Rascha, że dla każdego ucznia o określonej sumarycznej liczbie punktów na teście, oszacowanie to przyjmuje tę samą wartość, gdyż rozkład *a posteriori* jest ten sam dla tych uczniów. Inaczej jest w przypadku PV: w związku z tym, że oszacowanie to stanowi losowa próbka z rozkładu *a posteriori*, uczniowie o tej samej liczbie punktów z testu mogą mieć różne oszacowania, w związku z czym nie wykorzystuje się ich zazwyczaj do komunikowania wyników uczniom, ale do celów badawczych. Oszacowania PV pozwalają bowiem na nieobciążone oszacowanie wariancji (wariancja estymatora EAP jest niedoszacowana), a w przypadku analiz wtórnych, np. w analizach stosowanych na potrzeby niniejszego raportu, umożliwiają wzięcie pod uwagę faktu, że testy osiągnięć szkolnych użyte do pomiaru umiejętności konkretnych uczniów charakteryzują się pewnym określonym poziomem precyzji, pozwalają zatem oddać niepewność pomiaru zmiennej zależnej w analizach.

Podobnie jak we wspomnianych wcześniej międzynarodowych badaniach osiągnięć szkolnych, tak i w badaniu SUEK oba rodzaje estymatorów wyników uczniów liczone były w ramach tzw. modeli warunkowych lub mówiąc inaczej, przy użyciu regresji latentnych. Podczas estymacji parametrów uczniowskich (zarówno EAP jak i PV), do skonstruowania rozkładu *a posteriori* potrzebny jest pewien model, który opisywałby rozkład danej cechy w populacji. W przypadku modelu bezwarunkowego najczęściej wykorzystywany jest rozkład normalny. Model z regresją latentną pozwala na bezpośrednie ujęcie zależności (lub ich braku) zachodzących w populacji pomiędzy umiejętnością a szeregiem innych zmiennych, zgodnie ze specyfikacją modelu. Uzyskane w ten sposób oszacowania wyników uczniów, pozwalają na ujęcie tych zależności w dalszych analizach, w których oszacowania te są używane. Więcej o zastosowaniu modelu z regresją latentną do uzyskiwania oszacowań poziomu umiejętności można znaleźć u Wu (2005).

Do modelu warunkowego dla danej umiejętności, jako zmienne niezależne wykorzystano zmienne znajdujące się w modelach kontrolnych dla tej umiejętności. Modele te opisane są w rozdziale 3<sup>17</sup>.

### 2.5.3. Ocena dopasowania danych do modelu

Zastosowany do skalowania danych model IRT, model Rascha, opiera się na trzech głównych założeniach. Pierwsze założenie mówi o tym, że zadania mierzą jeden konstrukt, tzn. że dane dotyczące odpowiedzi uczniów można wytłumaczyć, odwołując się wyłącznie do mierzonej umiejętności. Drugie założenie mówi o tym, że dla danego poziomu

<sup>17</sup> Należy zwrócić uwagę, że oszacowania dla uczniów nie uwzględniają podziału uczniów na klasy i szkoły, może powodować zaniżone oszacowania wariancji międzygrupowej liczonej za pomocą tak uzyskanych wyników. Obecnie jednak, według wiedzy autorów raportu, nie istnieje oprogramowanie, które umożliwiłoby uwzględnienie w sposób poprawny złożonego schematu doboru próby podczas szacowania wyników uczniów.

umiejętności ucznia odpowiedzi na poszczególne zadania są niezależne od siebie. Zgodnie z trzecim założeniem, dane dotyczące odpowiedzi uczniów na zadania wykazują określone prawidłowości, które można ująć w postaci funkcji.

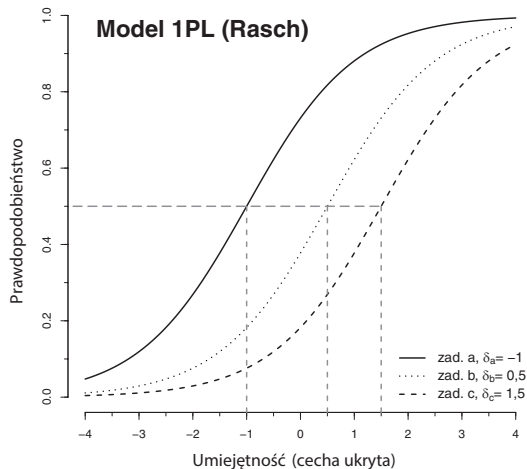
I tak, dla zadania ocenianego dychotomicznie, gdzie 1 oznacza odpowiedź poprawną, a 0 odpowiedź niepoprawną, prawdopodobieństwo udzielenia poprawnej lub błędnej odpowiedzi na zadanie  $j$  przez ucznia  $i$  (tj.  $x_{ij}$ ) zależy wyłącznie od różnicy poziomu umiejętności ucznia  $\theta_i$  oraz trudności zadania  $\delta_j$ . Różnica ta wyrażona jest na skali logitowej i odpowiada logarytmowi naturalnemu ilorazu szans, tj. prawdopodobieństwa udzielenia poprawnej odpowiedzi na zadanie  $j$  przez ucznia  $i$  oraz prawdopodobieństwa udzielenia odpowiedzi niepoprawnej:

$$\theta_i - \delta_j = \ln \left( \frac{P(x_{ij} = 1)}{P(x_{ij} = 0)} \right)$$

Z powyższego wzoru łatwo można wyprowadzić formułę na uzyskanie prawdopodobieństwa udzielenia odpowiedzi poprawnej:

$$P(x_{ij} = 1) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}$$

Trudność zadania jest więc definiowana jako punkt na skali umiejętności, dla którego prawdopodobieństwo udzielenia przez ucznia odpowiedzi poprawnej wynosi 50%. Reprezentacją graficzną tej funkcji jest krzywa charakterystyczna dla zadania. Poniżej zaprezentowano krzywe charakterystyczne dla trzech hipotetycznych zadań ocenianych dychotomicznie:



**Rysunek 2.4. Krzywe charakterystyczne dla trzech zadań a, b i c.**

Źródło: opracowanie własne

## CZY SZKOŁA MA ZNACZENIE?

W związku z tym, że zadania różnią się tylko parametrem trudności, posiadają one takie same własności różnicujące uczniów (tj. ten sam parametr nachylenia, zdefiniowany jako parametr nachylenia stycznej do krzywej charakterystycznej dla zadania  $j$  w punkcie  $[\delta_j; 0,5]$ ). Założenie to implikuje, że krzywe charakterystyczne dla zadań nie przecinają się. Oznacza to, że model stawia bardzo restrykcyjne wymagania co do zależności obserwowanych w danych. Ocena tego, czy wymagania te są spełnione, pozwala na określenie, czy dany zestaw zadań tworzy dobry, zgodny z modelem pomiarowym test.

Weryfikacja hipotezy o spełnieniu założeń modelu opiera się na analizie dopasowania empirycznych krzywych charakterystycznych dla zadań do odpowiednich krzywych teoretycznych. Miary dopasowań zdają sprawę ze zgodności przewidywań modelu i danych empirycznych. Do oceny tego dopasowania służą dwie miary – INFIT i OUTFIT. Miara OUTFIT jest równie wrażliwa na odstępstwa od przewidywań modelu dla przypadków w całym zakresie skali. Z kolei miara INFIT, większą wagę przykładą do tych przypadków, które znajdują się na skali w pobliżu punktu wyznaczonego jako trudność zadania (czyli tam, gdzie zadanie teoretycznie powinno najlepiej różnicować uczniów). Gdy miary te osiągną jeden, przyjmuje się, że zadanie jest dobrze dopasowane do modelu. W praktyce dopuszcza się jednak pewne rozchwianie wartości tych statystyk (de Ayala, 2009, s. 55–57). W badaniu SUEK na etapie pilotażu i konstrukcji docelowych narzędzi przyjęto, że zadania, dla których miary dopasowania wahały się między 0,8 i 1,2 były wystarczająco dobrze dopasowane. Dokładny opis porównania miar dopasowania dla zadań, które weszły do testów osiągnięć szkolnych użytych w badaniu SUEK, pomiędzy badaniem pilotażowym i zasadniczym znajduje się w: Jasińska, Modzelewski (2012), warto jednak nadmienić, że dla modeli bezwarunkowych (bez zastosowania regresji latentnej), skonstruowane testy zachowały swoje dobre własności pomiarowe na etapie badania zasadniczego.

### 2.5.4. Standaryzacja

Otrzymane oszacowania wyników uczniów – zarówno oszacowania EAP jak i pięć *plausible values* – wyjściowo znajdują się na skali logitowej. Dla jasności prezentowanych wyników analiz umieszczonych w dalszych częściach raportu, oszacowane wyniki uczniów wystandaryzowano i przeniesiono na skalę o średniej 100 i odchyleniu standardowym 15, według następującego wzoru:

$$X_i^{st} = \frac{X_i - \bar{U}}{\sigma_{rr}} \times 15 + 100$$

gdzie:

$X_i^{st}$ , oznacza wystandaryzowany wynik  $i$ -tego ucznia w teście mierzącym umiejętność U

$X_p$ , oznacza surowy wynik  $i$ -tego ucznia w teście mierzącym umiejętność U

$\bar{U}$ , oznacza estymator średniej populacyjnej dla umiejętności U

$\sigma_U$ , oznacza estymator odchylenia standardowego umiejętności U w populacji.

Procedura standaryzacji wymagała oszacowania średniej oraz odchylenia standardowego dla każdej ze zmiennej poddanej standaryzacji. Zarówno oszacowania EAP, jak i PV dla danej umiejętności zostały wystandaryzowane z użyciem tej samej średniej i odchylenia standardowego. Wykorzystano najlepsze dostępne estymatory tych parametrów dla populacji. Dla każdej skali umiejętności wyliczono średnią oraz wariancję obliczoną jako średnią z wariancji każdej z *plausible values* dla danej umiejętności. Pierwiastek z tak uzyskanej wariancji posłużył jako estymator odchylenia standardowego w populacji dla danej umiejętności. Zarówno oszacowanie średniej jak i oszacowania wariancji dla każdej z PV zostały obliczone z wykorzystaniem wag próbkowania.

Wartości estymatorów średniej i odchylenia standardowego dla każdej z umiejętności zostały przedstawione w tabeli 2.2.

**Tabela 2.2. Wartości estymatorów średniej i odchylenia standardowego dla poszczególnych umiejętności (skala logitowa)**

Umiejętność	Estymator średniej	Estymator odchylenia standardowego
Czytanie	0,517	1,06
Matematyka	0,302	1,05
Świadomość językowa	0,148	1,07

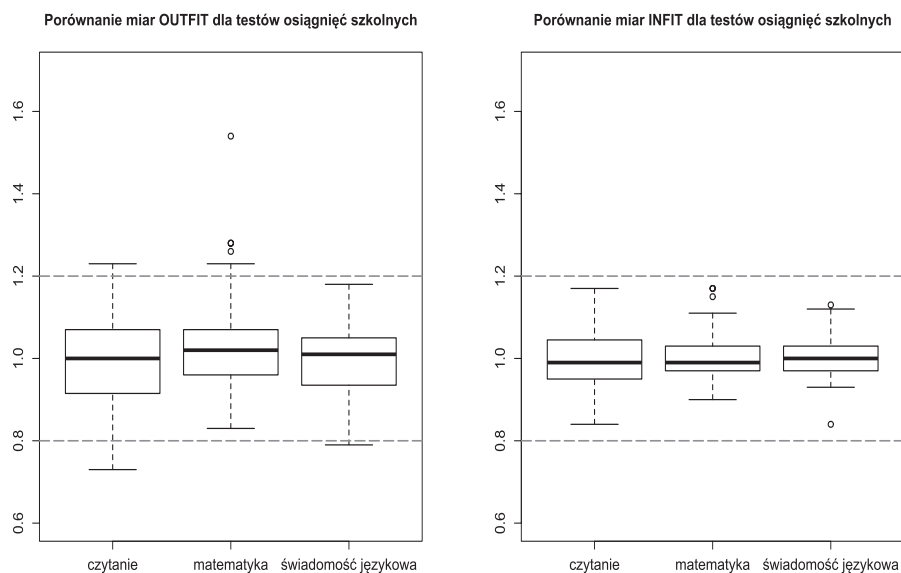
Źródło: opracowanie własne

## 2.6. Właściwości psychometryczne stworzonych testów

### 2.6.1. Właściwości zadań

W poprzednim podrozdziale wspomniano o procedurze służącej do ewaluacji hipotezy o dopasowaniu modelu do danych. Ze względu na restrykcyjne założenia obecne w modelu Rascha, często mówi się, że próba dopasowania tego modelu do danych ma charakter analizy confirmacyjnej, w której hipoteza badawcza poddawana próbie brzmi: „użyte narzędzie pozwala na dobry pomiar badanej umiejętności”. Czy zatem testy osiągnięć szkolnych skonstruowane na potrzeby badania SUEK pozwalają na przyjęcie wspomnianej hipotezy? Przyjrzyjmy się rozkładowi miar dopasowania dla zadań w trzech stworzonych testach.

## CZY SZKOŁA MA ZNACZENIE?



**Rysunek 2.5. Wykresy skrzynkowe miar dopasowania zadań dla testów osiągnięć szkolnych.**

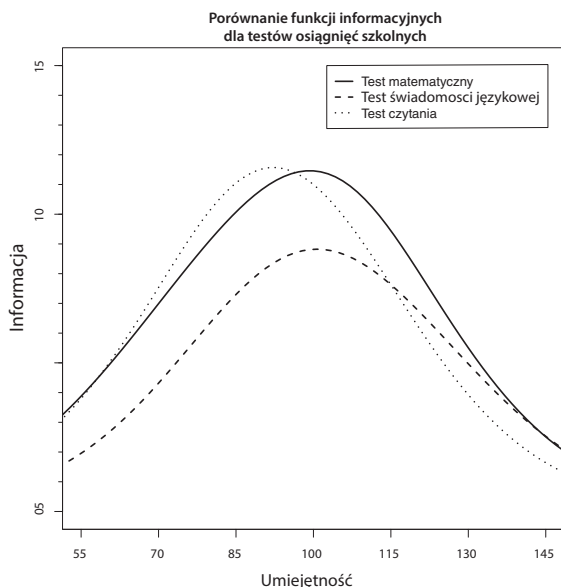
*Źródło: opracowanie własne*

Wykresy skrzynkowe przedstawiają rozkłady miar dopasowania zadań (INFIT i OUTFIT) wyliczone dla parametrów zadań oszacowanych w kroku kalibracji zadań, a następnie zadanych jako stałe w modelu z regresją latentną. Jak widać, dla zdecydowanej większości zadań miary dopasowania znajdują się w wyznaczonych granicach, tj. pomiędzy 0,8 i 1,2. Kilka zadań, które nie wykazują się dobrym dopasowaniem, zachowano ze względu na zapisy planów testów. Można zatem powiedzieć, że testy osiągnięć szkolnych skonstruowane na potrzeby badania SUEK, w bardzo wysokim stopniu spełniają wymogi stawiane przez model Rascha, a tym samym, że są narzędziem, skonstruowanym z zadań, które w badanych przez siebie zakresach dostarczają rzetelnej informacji odnośnie umiejętności uczniów.

### 2.6.2. Właściwości skal pomiarowych

Powyższe analizy pozwalają stwierdzić, że testy osiągnięć szkolnych skonstruowane na potrzeby badania SUEK pozwalają na pomiar umiejętności uczniów w zgodzie z logiką modelu Rascha. Narzędzia zostały stworzone jako testy szerokiego zasięgu, tzn. mają na celu precyzyjny pomiar dla jak najszerszego zakresu badanej umiejętności w populacji uczniów po klasie trzeciej. Zarówno na etapie konstrukcji testów, jak i podczas ich zastosowania w badaniu zasadniczym, precyzję pomiaru badanej umiejętności w różnych zakresach analizowano na podstawie krzywej informacyjnej testu.

## POMIAR WYNIKÓW NAUCZANIA W ZAKRESIE CZYTANIA, ŚWIADOMOŚCI JĘZYKOWEJ I MATEMATYKI



**Rysunek 2.6. Porównanie funkcji informacyjnych dla trzech testów osiągnięć szkolnych w odniesieniu do skali z wystandaryzowanymi wynikami.**

*Źródło: opracowanie własne*

Krzywa informacyjna testu jest przydatnym narzędziem pozwalającym graficznie przedstawić zdolność testu do szacowania poziomu umiejętności w różnych jej zakresach (de Ayala, 2009, s. 31–33). Im większa wartość funkcji informacyjnej, tym większa precyzja pomiaru badanej cechy. W zależności od uwzględnionych w teście zadań, przyjętego modelu estymacji, krzywa informacyjna dla testu może przybierać różne kształty. Na etapie konstrukcji testu, narzędzie to pozwala więc dosłownie na projektowanie własności pomiarowych testu w zależności od potrzeb – wystarczy określić pożądany kształt krzywej. Krzywa informacyjna dla idealnego testu osiągałaby wysoką wartość maksimum dla średniego wyniku w populacji i utrzymywałaby wysoki wynik aż po skraje skali. Znaczyłoby to – przy zachowaniu założeń modelu – że test pozwala na bardzo dokładny pomiar umiejętności wśród wszystkich członków badanej populacji, niezależnie od ich poziomu umiejętności. Narzędzie takie musiałoby składać się z bardzo wielu zadań, o bardzo zróżnicowanej trudności<sup>18</sup>.

<sup>18</sup> Należy jednak pamiętać, że dokładność pomiaru, wraz z czasem potrzebnym na jego przeprowadzenie, z zasady znajdują się w relacji monotonicznie rosnącej: im więcej zadań w teście, tym dokładniejszy pomiar. Jednym ze sposobów na zwiększenie precyzji pomiaru, przy zachowaniu rozsądnego czasu badania, stanowi testowanie adaptatywne: poziom umiejętności badanego szacowany jest na bieżąco, a każde kolejne zadanie dobierane jest ze względu na poprzednie odpowiedzi badanego. W ten sposób narzędzie zatacza coraz węższe okrężenia wokół najpewniejszego oszacowania poziomu umiejętności badanego, aż do osiągnięcia pożądanego stopnia precyzji.



## CZY SZKOŁA MA ZNACZENIE?

Ulokowanie krzywej informacyjnej testu nad skalą umiejętności pozwala na określenie czy dany test dobrze „wpasował się” w zakres umiejętności uczniów. Poniżej przedstawione zostały krzywe informacyjne dla trzech testów skonstruowanych na potrzeby badania SUEK.

Krzywe dla funkcji informacyjnych dla każdego z testów zostały ułożone nad wystandaryzowaną skalą, na której wyrażone są wyniki uczniów dla poszczególnych umiejętności. Na tej skali punkt 100 odpowiada wartości estymatora średniej populacyjnej, a odcinek o długości 15 odpowiada estymatorowi odchylenia standardowego danej umiejętności w populacji.

Funkcja informacyjna dla testu stanowi sumę wartości funkcji informacyjnych dla poszczególnych zadań, obliczaną warunkowo dla danego poziomu cechy. W celu ułatwienia interpretacji wykresu, można dodać, że dla danego zadania  $j$  ocenianego dychotomicznie, funkcja informacyjna dla modelu Rascha wyraża się wzorem:

$$I_j(\theta) = P(x_j = 1 | \theta, \delta_j) * P(x_j = 0 | \theta, \delta_j)$$

i jej maksimum wynosi 0,25.

Analizując wykres, można powiedzieć, że testy są dobrze dopasowane do rozkładu umiejętności uczniów w populacji – w zakresie  $\pm 2$  odchylenia standardowe (gdzie mieści się około 95% populacji) skonstruowane testy osiągnięć szkolnych pozwalają na pomiar umiejętności z zadowalającym stopniem precyzji. Stopień dopasowania testów waha się jednak dla poszczególnych umiejętności. Test czytania okazał się trochę za łatwy w stosunku do poziomu uczniów badanych, wynika to jednak z trudności związanych z pozyskaniem zadań o wyższej trudności na etapie konstrukcji testu. Test świadomości językowej z kolei posiada zdecydowanie mniej zadań niż test matematyczny czy wspomniany test czytania – stąd maksimum jego funkcji informacyjnej jest niższe niż dwóch pozostałych narzędzi, jednakże posiada porównywalną do testu z matematyki jakość do precyzyjnego pomiaru uczniów o wysokim poziomie umiejętności.

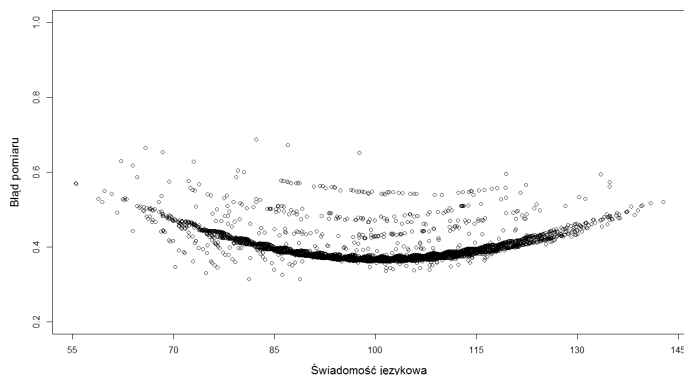
### 2.6.3. Rzetelność i błąd pomiaru

Niepewność pomiaru dla danego narzędzia przyjęło się określać w klasycznej teorii testu (KTT) jako jego rzetelność, czyli związek pomiędzy wynikiem testu a tzw. wynikiem prawdziwym. Związane jest to bezpośrednio z założeniami dotyczącymi pomiaru, zgodnie z którymi wynik konkretnego badanego w teście równy jest sumie wyniku prawdziwego i nieskorelowanego z tym wynikiem błędu pomiaru. W ramach KTT, stworzono grupę statystyk, które pozwalają na globalne oszacowanie tej własności testu (jedną z najpopularniejszych jest miara wewnętrznej zgodności narzędzia, alfa Cronbacha). Można wskazać na co najmniej dwa problemy związane z pojęciem rzetelności w KTT. Po pierwsze, bardzo często podając oszacowanie rzetelności dla testu, przyjmuje się, że jest to jego immanentna własność, zapominając o tym, że oszacowanie to zależy od próby badanych, którzy test rozwiązywali (narzędzie może okazać się rzetelne w jednej próbie, a w innej

nie). Po drugie, za tak rozumianą rzetelnością stoi założenie, że precyzja pomiaru jest identyczna dla każdego ucznia – to znaczy, że zarówno uczniowie o niskich jak i wysokich umiejętnościach badani są z tą samą precyzją. Założenie to jest jednak bardzo wątpliwe.

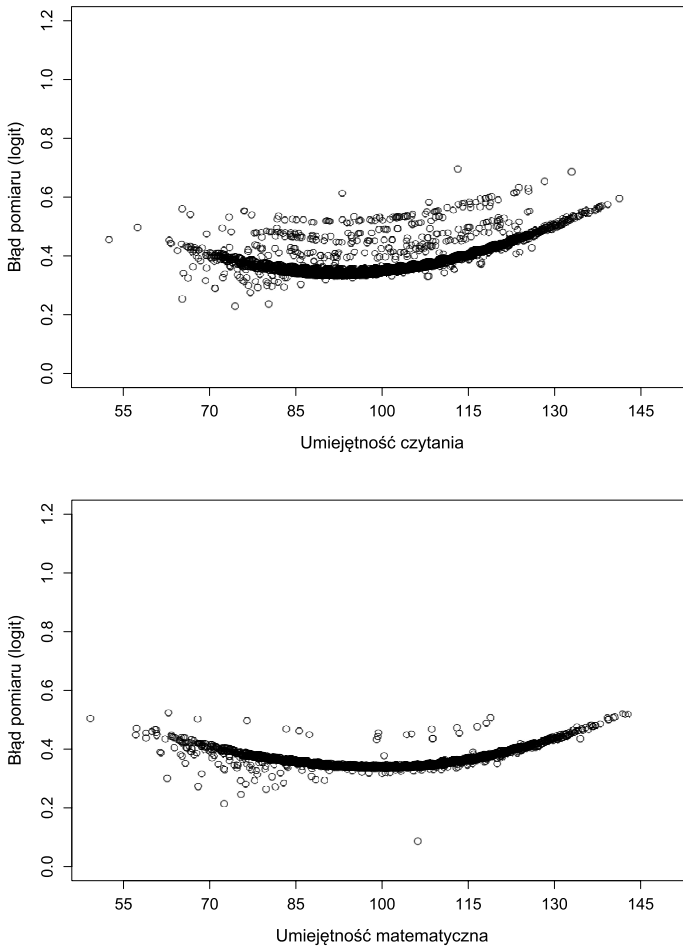
IRT przyjmuje zupełnie inne założenia odnośnie pomiaru umiejętności. W teorii tej modelowane jest prawdopodobieństwo prawidłowej odpowiedzi ucznia na określone zadanie w zależności od jego parametrów i poziomu umiejętności ucznia. Nic dziwnego zatem, że podejście do precyzji pomiaru również jest odmienne. W IRT, bazując na informacji o tym, jak uczeń poradził sobie z danymi zadaniami (a przy tym znając ich parametry oraz wiedząc jaki rozkład przyjmuje umiejętność w populacji), możemy określić najbardziej prawdopodobne ułożenie ucznia na skali, np. oszacowanie EAP. Ułożenie to obarczone jest pewnym błędem, który można przedstawić jako odchylenie standardowe rozkładu umiejętności *a posteriori* dla ucznia. W zależności od tego jak dobrze dobraliśmy zadania, które przed uczniem postawiliśmy (liczba informacji dostarczanej przez zadanie zależy od jego trudności) oraz jak wiele ich było, tym precyzyjniej jesteśmy w stanie określić jego poziom umiejętności. W związku z tym, w IRT możemy powiedzieć o wiele więcej o precyzji pomiaru konkretnego ucznia niż w ramach KTT – jesteśmy bowiem w stanie oszacować indywidualny błąd pomiaru. Znika tym samym problem braków danych dla poszczególnych zadań, który często paraliżuje możliwość oceny rzetelności pomiaru w KTT – w IRT, brak odpowiedzi ucznia na dane zadanie zmniejsza po prostu precyzję pomiaru jego poziomu umiejętności.

Poniżej dla każdego z trzech testów prezentujemy wykresy indywidualnego błędu pomiaru, rozumianego jako odchylenie standardowe rozkładu *a posteriori* umiejętności uczniów w zależności od oszacowanego dla nich wyniku. Błąd pomiaru wyrażony jest na skali logitowej.



**Rysunek 2.7. Wykres rozrzutu dla błędu pomiaru wyrażonego na skali logitowej w zależności od wystandaryzowanego oszacowanego wyniku dla trzech umiejętności.**

## CZY SZKOŁA MA ZNACZENIE?



**Rysunek 2.7. Wykres rozrzutu dla błędu pomiaru wyrażonego na skali logitowej w zależności od wystandaryzowanego oszacowanego wyniku dla trzech umiejętności (c.d.).**

*Źródło: opracowanie własne*

Na wszystkich trzech wykresach możemy dostrzec, że błędy pomiaru są najmniejsze dla tych oszacowań, które znajdują się blisko średniej testu (a dokładniej, bliżej tego poziomu umiejętności, dla którego funkcja informacyjna testu osiąga maksimum). Widzimy jednak, że nie dla wszystkich uczniów o określonym poziomie umiejętności, wynik został oszacowany z podobną precyzją. Dla każdego testu można określić główną zależność pomiędzy błędem pomiaru a poziomem umiejętności – wyznaczoną przez krzywą złożoną z dużego zagęszczenia punktów. Są to ci uczniowie, którzy rozwiązywali wszystkie zadania w przeznaczony im

## POMIAR WYNIKÓW NAUCZANIA W ZAKRESIE CZYTANIA, ŚWIADOMOŚCI JĘZYKOWEJ I MATEMATYKI

wersji testu. Dla testów świadomości językowej i czytania możemy także wyróżnić trzy inne formy zależności, tj. krzywe znajdujące się powyżej głównego trendu. Tworzą je uczniowie, którzy rozwiązyali tylko część przeznaczonych dla nich zadań (uczniowie ci zapewne nie byli obecni na wszystkich sesjach testowych). Dla testu matematycznego poza głównym trendem możemy dostrzec jeszcze jedną formę zależności. Wyznaczona jest ona przez uczniów, którzy rozwiązyali tylko jeden zeszyt testowy – ich liczba jest jednak niewielka, gdyż obie sesje testowe z matematyki odbywały się jednego dnia.

Dla każdego z trzech testów wielkość błędów w zależności od poziomu umiejętności jest zadowalająca – w granicach  $\pm 2$  odchyłeń standardowych oscyluje ona w okolicach 0,4 logita, co w przełożeniu na standaryzowaną skalę wynosi mniej niż 40% odchylenia standardowego – należy jednakże pamiętać, że rozkłady umiejętności *a posteriori* nie mają ściśle rzecz ujmując rozkładów normalnych. Precyzja pomiaru stopniowo maleje dla uczniów oddalonych od średniej – ich rozkłady *a posteriori* są zazwyczaj skośne w stronę „od średniej”, co zwiększa ich wariancję. Jest to zgodne ze zdroworozsądkowym podejściem do pomiaru dydaktycznego – poziom umiejętności ucznia, który rozwiązał wszystkie lub prawie wszystkie zadania, jest prawdopodobnie wysoki – jednakże jak wysoki? Podobnie ma się sprawa w przypadku uczniów, którzy nie poradzili sobie z dużą liczbą zadań. Widzimy jednak, że skonstruowane testy pozwalają na pomiar umiejętności dla uczniów ze skrajów skali ze stosunkowo dużą precyzją – dla całego zakresu umiejętności wielkości błędów dla wszystkich testów nie przekraczają poziomu 0,6 logita (w przypadku głównej zależności) – test czytania relatywnie słabiej mierzy poziom umiejętności uczniów zdolnych, a test świadomości językowej uczniów słabszych. Test matematyczny zachowuje się w miarę spójnie w całym zakresie skali. Wielkość błędów ma także związek z dostępnymi informacjami uwzględnionymi w regresji latentnej, na podstawie której wyznaczano wyniki uczniów (Wu, 2005).

Choć dla metodologii IRT liczenie jednego globalnego współczynnika, który zdawałby sprawę z rzetelności nie do końca jest zgodne z jej logiką, podjęto działania mające na celu stworzenie wskaźnika precyzji pomiaru dla testu w danej próbie badanych. Można wyróżnić kilka takich metod (por. de Ayala, 2009; Kim i Feldt, 2010; Adams, 2005), w rozdziale tym przedstawimy jedną z nich – tzw. współczynnik rzetelności EAP/PV (Adams, 2005). Jest to jedna ze statystyk dla testów, którą wylicza program ConQuest 2.0.

Współczynnik ten jest stosunkiem wariancji oszacowań EAP dla uczniów i wariancji umiejętności w populacji. Nieobciążonym estymatorem tej ostatniej jest wariancja wyliczona dla PV. Współczynnik ten osiąga wartości z przedziału (0;1). Im wyższa jest jego wartość, tym średnio błąd pomiaru w próbie jest mniejszy. Wynika to z faktu, że im precyzyjniej pomierzylibyśmy poziom umiejętności uczniów, tym mniejsza byłaby wariancja rozkładów *a posteriori* ich umiejętności, a w związku z tym wylosowane z tego rozkładu PV byłyby bliskie jego wartości

## CZY SZKOŁA MA ZNACZENIE?

oczekiwanej, EAP. Wariancja oszacowań EAP dla uczniów byłaby więc bardzo zbliżona do wariancji dla PV.

W tabeli poniżej przedstawiono oszacowania tego współczynnika dla testów osiągnięć szkolnych.

**Tabela 2.3. Wartość współczynnika rzetelności EAP/PV dla testów osiągnięć szkolnych**

Test	Wartość współczynnika rzetelności EAP/PV
Test czytania	0,866
Test matematyczny	0,876
Test świadomości językowej	0,864

*Źródło: opracowanie własne*

Wartości współczynników są wysokie dla wszystkich trzech testów, co potwierdza wyniki analiz rozkładów błędów dla każdego z nich. Widzimy także, że większa precyzja pomiaru dla testu matematycznego dla całego zakresu skali jest odwzorowana w relatywnie wyższej wartości współczynnika rzetelności dla tego testu. Należy jednakże pamiętać, że współczynnik ten stanowi uśrednienie precyzji pomiaru dla wszystkich uczniów – informacja o indywidualnych różnicach w precyzji pomiaru zostaje utracona za cenę wygody posługiwania się jedną wartością liczbową dla całego testu.

Podsumowując rozważania o rzetelności testów osiągnięć szkolnych, można powiedzieć, że złożony i przemyślany proces ich tworzenia a także zastosowana metodologia pomiaru, zaowocowały uzyskaniem narzędzi o bardzo dobrych własnościach psychometrycznych. Pomiar umiejętności przeprowadzony za pomocą tych narzędzi zapewnia wyniki obarczone niepewnością o rozsądnych rozmiarach dla całego zakresu skali.

### 2.6.4. Trafność pomiaru osiągnięć szkolnych

Zagadnienie trafności pomiaru związane jest z tym, jak dobrze test mierzy to, co zakłada się, że mierzyć powinien. W psychologicznym ujęciu problematyki trafności wiele miejsca poświęca się na wyszczególnianie i opisywanie różnych rodzajów trafności, konsekwencji ich braku oraz sposobów ich badania (Anastasi i Urbina, 1999). W kontekście pomiaru dydaktycznego najważniejsze wydają się pytania o to, czy test obejmuje reprezentatywną próbę mierzonych wiadomości i umiejętności oraz czy wywołuje u uczniów żądane procesy myślowe i pozwala na zarejestrowanie dowodów, że takie procesy zaistniały (Jakubowski i Pokropek, 2009).

## POMIAR WYNIKÓW NAUCZANIA W ZAKRESIE CZYTANIA, ŚWIADOMOŚCI JĘZYKOWEJ I MATEMATYKI

Trafność treściowa narzędzia pomiarowego mówi o tym, czy pozycje testowe stanowią reprezentatywną próbę dziedziny, która ma być przedmiotem pomiaru. Ocena testu osiągnięć szkolnych ze względu na to kryterium zasadniczo polega na analizie jego treści w celu stwierdzenia, czy wchodzące w jego skład zadania mierzą wszystkie ważne z punktu widzenia badanego konstruktowi wiadomości i umiejętności oraz że zadania odwołujące się do szczegółowych umiejętności, znajdują się w teście we właściwych proporcjach. Trafność treściowa testu jest w dużym stopniu wynikiem przyjętych procedur konstrukcji narzędzia, a nie oceny *post factum*. Aby ustrzec się przed przypadkowością pozycji składających się na test lub też nadreprezentacją zadań mierzących takie umiejętności, dla których łatwiej ułożyć dobre psychometrycznie zadanie, należy badaną dziedzinę z góry dobrze opisać, a także opracować dokładną specyfikację zawartości testu. Ważna jest także ekspercka ocena zadań oraz weryfikacja empiryczna, które mogą ustrzec przed włączeniem do testu zadania, które tylko z pozoru mierzą założoną umiejętność.

Na procedury, które miały zapewnić trafność treściową testów osiągnięć wykorzystanych w badaniu SUEK, składały się na etapie planowania narzędzia: zdefiniowanie i opisanie skal pomiarowych, analiza podstawy programowej i obowiązujących programów nauczania oraz opracowanie na ich podstawie szczegółowych planów testów precyzujących, ile zadań mierzących jakie umiejętności powinno się znaleźć w teście. Następnie koncepcja skal pomiarowych, plany testów oraz opracowane zgodnie z nimi zadania testowe zostały poddane recenzji zewnętrznej i drobiazgowej ocenie eksperckiej. Uwagi przekazane przez recenzentów, pozwoliły na udoskonalenie części zadań, które tego wymagały i wykluczeniu zadań najbardziej problematycznych. Tak wyłoniona grupa zadań została poddana weryfikacji empirycznej w badaniu pilotażowym, a do testów zasadniczych zostały wybrane zadania nie tylko najlepsze pomiarowo, ale także zgodne, poza opisanymi wcześniej nielicznymi wyjątkami, z przyjętymi planami testów. W ten sposób udało się zapewnić różnorodność i reprezentatywność treściową opracowanych testów.

Specyfika badania pilotażowego (realizowanego z udziałem uczniów klas III i V szkół podstawowych) pozwoliła także na weryfikację dodatkowych informacji na temat trafności treściowej testu. Analizy objęły bowiem także problematykę poziomu wykonania zadań przez uczniów klas III i V dla pozycji testowych, które były wykorzystywane w obu populacjach. Na podstawie tych analiz wykluczano pozycje, dla których nie obserwowano większego prawdopodobieństwa porażenia sobie z zadaniem wśród uczniów klas V niż wśród uczniów klas III, biorąc to za dowód tego, że zadanie mierzy taką specyficzną umiejętność, która nie jest rozwijana przez szkołę, więc nie może być traktowana jako wskaźnik osiągnięć szkolnych.

Innym ważnym rodzajem trafności, któremu poświęca się wiele miejsca w literaturze (zob. przegląd tekstów: Brzeziński, 2005), jest trafność teoretyczna, która mówi o tym, w jakim

## CZY SZKOŁA MA ZNACZENIE?

stopniu test mierzy teoretyczny konstrukt, do którego pomiaru został stworzony. W przypadku testów osiągnięć szkolnych zbudowanych na potrzeby badania SUEK należy postawić pytanie o to, czy faktycznie mierzą one założone w koncepcji osobne (nawet jeśli skorelowane ze sobą), wewnętrznie spójne grupy umiejętności, czy też raczej ogólne umiejętności związane z funkcjonowaniem szkolnym. W języku psychometrii pytamy więc o to, czy znajdujemy w danych potwierdzenie trójczynnikowej struktury testu (Konarski, 2009). Przyjęta w koncepcji definicja mierzonych konstruktów pozwala założyć, że związek między umiejętnością czytania i świadomością językową będzie silniejszy (jako że są to kompetencje językowe) niż związek tych wymiarów z umiejętnościami matematycznymi.

Aby odpowiedzieć na to pytanie, przetestowano dwa konkurencyjne modele: jednowymiarowy model Rascha, w którym wszystkie pozycje testowe są wskaźnikami jednego wymiaru ogólnych osiągnięć szkolnych oraz trójwymiarowego modelu Rascha, zgodnego z przyjętą koncepcją. Do obliczenia estymatorów największej wiarygodności wykorzystano metodę Monte Carlo. Analizy wykonano za pomocą oprogramowania ConQuest 2.0 (Wu i in., 2007). Wyniki tych analiz pokazują, że model trójwymiarowy dla tych danych jest lepiej dopasowany (deviance = 580477,21; liczba estymowanych parametrów = 159; N = 5536) niż model jednowymiarowy (deviance = 583839,91; liczba estymowanych parametrów = 154; N = 5536). Wartość statystyki deviance opartej na wartości funkcji wiarygodności jest znacząco niższa dla modelu trójwymiarowego ( $\Delta = 3362,7$ )<sup>19</sup>.

Za pomocą modelu trójwymiarowego oszacowano także korelacje między trzema ukrytymi wymiarami, czy inaczej umiejętnościami latentnymi, odpowiadającymi trzem skalom pomiarowym. Na oszacowania tych korelacji nie ma wpływu rzetelność testów, dlatego też są one zwane „korelacjami prawdziwymi” lub „latentnymi”. Korelacje te są z reguły wyższe niż odpowiadające im korelacje dla estymatorów punktowych.

**Tabela 2.4. Korelacje latentne między trzema wymiarami umiejętności**

	Skala umiejętności matematycznych	Skala umiejętności czytania
Skala umiejętności czytania	0,820	-
Skala świadomości językowej	0,824	0,916

*Źródło: opracowanie własne*

Otrzymane korelacje są zgodne z oczekiwaniami: związek między umiejętnością czytania i świadomością językową jest silniejszy niż związek kompetencji językowych z umiejętnościami matematycznymi, co także świadczy na korzyść trafności teoretycznej skonstruowanych

<sup>19</sup> Interpretując te dane, należy zwrócić uwagę na to, że testowane modele nie są modelami zagnieżdżonymi.

## POMIAR WYNIKÓW NAUCZANIA W ZAKRESIE CZYTANIA, ŚWIADOMOŚCI JĘZYKOWEJ I MATEMATYKI

testów. Korelacje te są także bardzo wysokie, co pokazuje, że na początku edukacji umysły szkolne uczniów nie są raczej silnie wyspecjalizowane. W świetle tych danych nieuzasadnione wydają się twierdzenia o uczniach o umysłach matematycznych lub humanistycznych w kontekście edukacji wczesnoszkolnej. Jeśli uczniowie mają wysokie osiągnięcia w zakresie matematyki, to w większości przypadków dobrze też radzą sobie z treściami językowymi i *vice versa*. Tak wysokie korelacje mogą być także pochodną wpływu jednego nauczyciela, który na tym etapie edukacji jest przewodnikiem w uczeniu się wszystkich treści objętych omawianymi tu testami osiągnięć. Natomiast korelacja między umiejętnością czytania i świadomością językową może być troszkę zawyżona ze względu na samą konstrukcję zeszytów testowych, w których zadania z obu tych testów zostały połączone we wspólne zeszyty z języka polskiego. Dlatego wyniki obu tych testów mogą być obciążone wspólnymi dla nich źródłami błędu pomiaru, jak np. dyspozycja ucznia w danym dniu, motywacja testowa, czy wpływ czynników zewnętrznych w trakcie badania testowego.

Pytanie dotyczące trafności teoretycznej testów obejmowało także aspekt zgodności wewnętrznej w ramach trzech wymiarów umiejętności, czyli tego, jak dobrze poszczególne pozycje testowe są związane z wynikiem ogólnym danej skali. Zgodność wewnętrzna została zapewniona już na etapie wyboru zadań z badania pilotażowego do wersji finalnych testów, dzięki przyjętemu modelowi analizy danych. Do testów zostały bowiem włączone zadania o odpowiedniej wartości parametru dyskryminacji będącego odzwierciedleniem korelacji pozycji z wynikiem ogólnym. Przyjęty model analizy danych (model Rascha) zakłada, że wszystkie zadania są w takim samym stopniu związane z wynikiem ogólnym, dzięki czemu z jednej strony w teście mogą znaleźć się tylko zadania, które istotnie korelują z wynikiem ogólnym, a ponadto wykluczona jest możliwość włączenia do testu zadań zbyt silnie determinujących wynik. Sprawia to, że wszystkie zadania składające się na test, wybrane zgodnie z planem testu, są tak samo ważne przy wyznaczaniu wyniku.

Konkludując, analizy potwierdzają trójwymiarową strukturę testu, czyli przemawiają za tym, że możemy mówić o pomiarze trzech, wewnętrznie spójnych, grup umiejętności. Wymiary te są wprawdzie silnie ze sobą skorelowane (w szczególności umiejętność czytania i świadomość językowa), ale pokazanie w tym raporcie specyficznych uwarunkowań poszczególnych grup osiągnięć, może przynieść więcej korzyści niż posługiwanie się jednym wymiarem osiągnięć szkolnych. Zdecydowano się zatem na odrębne analizy dla każdej skali, tym bardziej, że mają one wystarczającą rzetelność.

Trafność testów bada się także, odnosząc ich wyniki do zewnętrznego kryterium (jest to tzw. badanie trafności kryterialnej). Jeśli badacz jest zainteresowany tym, czy testy dobrze mierzą daną cechę, a ma inne wiarygodne źródło informacji o obecnym stanie tej cechy (lub cech, o których wiadomo, że pozostają w związku z mierzoną), może zweryfikować związek pomiędzy wynikami



## CZY SZKOŁA MA ZNACZENIE?

testu, a tym zewnętrznym kryterium (będzie to badanie trafności diagnostycznej). Jeśli jest zainteresowany tym, jak dobrze wyniki testu przewidują osiągnięcie pewnych właściwości w przyszłości, powinien odnieść wyniki testu do zewnętrznego kryterium zmierzonego po upływie określonego czasu (w ten sposób zweryfikuje trafność prognostyczną). Trafności prognostycznej omawianych tu testów osiągnięć nie sposób jeszcze zweryfikować ze względu na brak odpowiednich danych. Będzie to jednak możliwe po zakończeniu kolejnych etapów badania podłużnego SUEK. Możemy jednak w pewnym stopniu zweryfikować trafność diagnostyczną.

W przypadku testów osiągnięć szkolnych jako zewnętrzne kryterium trafności często wykorzystuje się oceny nauczycielskie lub wyniki innych testów osiągnięć (Anastasi i Urbina, 1999). Mimo że oceny nauczycielskie są miarą, ze statystycznego punktu widzenia, często mało rzetelną, zasadniczo nieporównywalną między grupami uczniów ocenianymi przez różnych nauczycieli (por. Jasińska, 2010), 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), a także w mniejszym stopniu uzależnionych od tymczasowej dyspozycji ucznia.

W badaniu podłużnym SUEK poproszono nauczycieli nauczania zintegrowanego badanych klas, by pod koniec nauczania w III klasie ocenili wszystkich swoich uczniów na cztero-stopniowej skali opisowej osobno dla umiejętności językowych, osobno dla matematycznych (ocenom na etapie kodowania przypisano następujące wartości liczbowe: 1 – ocena najniższa; 4 – ocena najwyższa). Aby zbadać związek ocen nauczycielskich z wynikami testów osiągnięć, policzono trzypoziomowy model mieszanych efektów (model z losową stałą), wykorzystując oprogramowanie HLM 6.06. Wykorzystano metodę estymacji *full maximum likelihood*. Z analiz usunięto wszystkie obserwacje z brakami danych. Analizy wykonano z wykorzystaniem wag uwzględniających schemat doboru próby do badania, odpowiednio przeskalowanych za pomocą programu HLM. Zastosowanie modelu wielopoziomowego, umożliwiło poradzenie sobie z problemem nieporównywalności ocen nauczycielskich między klasami. Estymatorami wyników testowych uczniów są oszacowania EAP. Oceny nauczycielskie zrekodowano na zmienne pomocnicze w taki sposób, by wartości współczynników regresji mówiły o przyroście średnich warunkowych w odniesieniu do kategorii o jeden niższej ze względu na uzyskaną ocenę. Kategorią odniesienia dla zmiennej V2 są uczniowie, którym została przypisana przez nauczycieli najniższa ocena<sup>20</sup>. W modelach, gdzie zmienną zależną jest wynik testu umiejętności czytania lub świadomości językowej, wykorzystano oceny dotyczące obszaru edukacji językowej, a w modelu gdzie zmienną wyjaśniają jest wynik testu umiejęt-

<sup>20</sup> Rozkłady ocen nauczycielskich są skośne – częściej występują oceny wyższe, jednak kategoria odniesienia stanowi ok. 12% próby.

## POMIAR WYNIKÓW NAUCZANIA W ZAKRESIE CZYTANIA, ŚWIADOMOŚCI JĘZYKOWEJ I MATEMATYKI

ności matematycznych, oceny z obszaru edukacji matematycznej. Tak więc np. współczynnik regresji przy zmiennej V2\_mat mówi o przyroście średniego wyniku testu umiejętności matematycznych dla grupy osób, które otrzymały ocenę 2 z matematyki. Współczynnik regresji przy zmiennej V3\_mat oznacza przyrost średniego wyniku testu dla grupy osób, które otrzymały ocenę 3 itd. Poniższa tabela prezentuje wyniki analiz.

**Tabela 2.5. Związek wyników testów z ocenami wystawionymi przez nauczycieli. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych**

Zamienna zależna:			
	wynik testu czytania	wynik testu świadomości językowej	wynik testu. matematycznego
<b>Oszacowanie efektów stałych</b>			
V2_pol <sup>a</sup>	8,37 (0,7)	9,41 (0,61)	
V3_pol <sup>b</sup>	7,86 (0,47)	8,01 (0,43)	
V4_pol <sup>c</sup>	10,05 (0,45)	10,06 (0,4)	
V2_mat <sup>a</sup>			8,86 (0,69)
V3_mat <sup>b</sup>			7,90 (0,49)
V4_mat <sup>c</sup>			11,38 (0,44)
Stała	83,26 (0,55)	82,0 (0,61)	81,82 (0,76)
<b>Oszacowanie efektów losowych</b>			
Wariancja efektów szkół	7,57	10,3	7,38
Wariancja efektów klas	8,56	8,36	11,12
Wariancja na poziomie ucznia	98,84	89,74	92,18
Podsumowanie			
Log likelihood	-18392,38	-18178,38	-17793,30
Liczba uczniów	4898	4898	4776
Liczba klas	288	288	288
Liczba szkół	167	167	167

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

a – grupa odniesienia: uczniowie, którym nauczyciele przyznali ocenę najniższą

b – grupa odniesienia: uczniowie, którym nauczyciele przyznali ocenę 2 na skali od 1 do 4

c – grupa odniesienia: uczniowie, którym nauczyciele przyznali ocenę 3 na skali od 1 do 4

*Źródło: opracowanie własne*

Związki wyników testowych z ocenami przyznanymi przez nauczycieli są istotne i bardzo podobne dla każdego z trzech testów. Uczniowie, którzy otrzymali najniższe oceny, mają średnie wyniki z testów poniżej jednego odchylenia standardowego wyników w populacji.

## CZY SZKOŁA MA ZNACZENIE?

Uczniowie, którzy otrzymali ocenę o jeden wyższą, mają średnio wyniki o ok. 8–9 punktów wyższe na skali o odchyleniu standardowym równym 15. Uczniowie, którym została przypisana 3 kategoria oceny, mają także średnio o ok. 8 punktów wyższe wyniki w testach od uczniów z oceną 2. Natomiast grupa uczniów najwyższej oceniona przez nauczycieli, ma średnie wyniki o kolejne 10–11 punktów wyższe, co daje średni wynik na skali standardowej ok. 110 punktów. Analizy te potwierdzają oczekiwany związek wyników testów osiągnięć z ocenami przyznanymi przez nauczycieli, co jest pozytywnym wynikiem w badaniu trafności diagnostycznej testów.

Drugim kryterium zastosowanym do oceny trafności diagnostycznej są wyniki uczniów z testu z języka polskiego i matematyki uzyskane w *Ogólnopolskim Badaniu Umiejętności Trzecioklasistów (OBUT)*<sup>21</sup> realizowanym w roku szkolnym 2010/11. Wykorzystując wyniki z tego pomiaru, należy zwrócić uwagę na kilka jego cech, które z punktu widzenia celów prezentowanych analiz, są jego mankamentami. Diagnoza OBUT jest przeprowadzana i oceniana przez nauczycieli uczących uczniów wypełniających testy. Szkoła otrzymuje opracowane przez zespół badawczy testy oraz instrukcje przeprowadzenia badania, jednak nie ma pewności, czy procedury te są przestrzegane. Narzędzia pomiarowe są krótkie, co sprawia, że wyniki pomiaru są mało dokładne (skala umiejętności matematycznych ma 17 rozróżnialnych kategorii, a skala testu języka polskiego 24), a w rozkładach wyników dostrzega się silny efekt sufitowy – testy są za łatwe, by dobrze różnicować uczniów najzdolniejszych (skośność wynosi odpowiednio: -0,964 dla testu z języka polskiego i -0,384 dla testu z matematyki). Mankamenty te mogą powodować zaniżenie korelacji między wynikami testów. Ponadto wyniki z badania OBUT udało się przyłączyć tylko dla ok. 63% uczniów objętych badaniem testowym SUEK.

W tabeli 2.6. zaprezentowano korelacje Pearsona między punktowymi oszacowaniami wyników testowych. W analizach wykorzystano sumę punktów z testów OBUT oraz estymatory EAP dla testów osiągnięć z badania SUEK. Z analiz wyłączono parami obserwacje z brakami danych.

<sup>21</sup> Badanie OBUT jest częścią projektu „Badanie umiejętności podstawowych uczniów trzeciej klasy szkoły podstawowej” współfinansowanego ze środków Unii Europejskiej. Projekt od 2005 roku był realizowany przy Centralnej Komisji Egzaminacyjnej, a od 09.2012 roku decyzją MEN został przeniesiony do Instytutu Badań Edukacyjnych.

**POMIAR WYNIKÓW NAUCZANIA W ZAKRESIE CZYTANIA,  
ŚWIADOMOŚCI JĘZYKOWEJ I MATEMATYKI**

**Tabela 2.6. Korelacje Pearsona między punktowymi oszacowaniami wyników testowych. Kursywą podano liczbę obserwacji, dla których wyznaczono współczynnik korelacji**

	umiejętność czytania (SUEK)	świadomość językowa (SUEK)	umiejętności matematyczne (SUEK)	test z języka polskiego (OBUT)	test z matematyki (OBUT)
umiejętność czytania (SUEK)	1	0,784 <i>N=5148</i>	0,703 <i>N=5033</i>	0,613 <i>N=3236</i>	0,548 <i>N=3236</i>
świadomość językowa (SUEK)	0,784 <i>N=5148</i>	1	0,702 <i>N=5033</i>	0,651 <i>N=3236</i>	0,550 <i>N=3236</i>
umiejętności matematyczne (SUEK)	0,703 <i>N=5033</i>	0,702 <i>N=5033</i>	1	0,565 <i>N=3171</i>	0,636 <i>N=3171</i>
test z języka polskiego (OBUT)	0,613 <i>N=3236</i>	0,651 <i>N=3236</i>	0,565 <i>N=3171</i>	1	0,624 <i>N=3483</i>
test z matematyki (OBUT)	0,548 <i>N=3236</i>	0,550 <i>N=3236</i>	0,636 <i>N=3171</i>	0,624 <i>N=3483</i>	1

*Źródło: opracowanie własne*

Uzyskany wzorzec korelacji jest zgodny z założeniami. Obserwujemy silniejszy związek wyników testu z języka polskiego z badania OBUT z wynikami testu umiejętności czytania i świadomości językowej wykorzystanymi w badaniu SUEK niż z wynikami z testu umiejętności matematycznych. Test z matematyki z badania OBUT także silniej koreluje ze skalą umiejętności matematycznych z testów z badania SUEK niż ze skalami umiejętności językowych. Niższa rzetelność pomiaru testami OBUT przyczyniła się także do uzyskania niższych korelacji z udziałem zmiennych z tego badania.

Badając trafność diagnostyczną testów można także sięgać po inne kryteria zewnętrzne, co do których wiemy, że powinny pozostawać w związku z mierzonym za pomocą testu konstruktem. W naszym przypadku byłyby to wskaźniki ogólnych zdolności poznawczych, statusu rodziny pochodzenia, czy wykształcenia rodziców. Analizy takie, choć wykonane na potrzeby innych interpretacji, zostały przedstawione w rozdziale poświęconym modelowi kontrolnemu dla analiz uwarunkowań efektywności kształcenia. Aby niepotrzebnie ich tu nie powtarzać, zachęcamy czytelnika do lektury tego rozdziału.

### 2.7. Podsumowanie

Podsumowując rozdział o pomiarze wyników nauczania, warto zastanowić się nad praktycznymi wnioskami płynącymi z zamieszczonych tu rozważań.

Doświadczenia z międzynarodowych badań edukacyjnych takich jak TIMSS, PIRLS czy PISA uczą nas, że rzetelny pomiar wyników kształcenia nie jest sprawą prostą, którą można by sprowadzić do zadania kilku arbitralnie wybranych pytań uczniom. Proces selekcji zadań do testów musi być poprzedzony namysłem nad celem pomiaru oraz gruntownym zdefiniowaniem mierzonego konstruktów – w efekcie okazuje się, że standardy, jakim muszą sprostać dobre narzędzia służące do pomiaru osiągnięć, są z pewnością nie niższe niż te, których wymagamy od testów psychologicznych.

Wykorzystanie standaryzowanych testów osiągnięć szkolnych w badaniach edukacyjnych często ma na celu oszacowanie zależności pomiędzy wynikami kształcenia a innymi charakterystykami uczniów, czy szkół. Poszukiwane zależności mają zazwyczaj charakter populacyjny, tzn. na podstawie pewnej reprezentatywnej próbki jednostek badania, próbuje się odnaleźć pewne prawidłowości statystyczne zachodzące dla pewnych większych grup społecznych, z których próby te pochodzą, takich jak np. uczniowie klas trzecich, czy ogół szkół podstawowych. Jakkolwiek rzetelny pomiar osiągnięć szkolnych na poziomie indywidualnym jest w tym kontekście bardzo ważny, to jednak zastosowane analizy dotyczą zazwyczaj większych grup i w związku z tym nie są aż tak czułe na niepewność pomiaru związaną z pomiarem indywidualnym.

Zupełnie inaczej jednak ma się sprawa w przypadku, gdy głównym przedmiotem naszego zainteresowania jest indywidualny wynik ucznia, tak jak ma to miejsce w przypadku np. egzaminów zewnętrznych w Polsce. Egzamin gimnazjalny oraz egzamin maturalny są testami, na podstawie których decydują się losy poszczególnych uczniów. Niestety, jakość psychometryczna stosowanych w nich testów pozostawia wiele do życzenia (Pokropek, 2011). Wydaje się, że wiele grzechów naszego już 12-letniego systemu egzaminacyjnego wynika zarówno ze słabo sprecyzowanego celu jakiego egzaminy te mają służyć oraz z niskich standardów i niedostatecznych procedur związanych z zachowaniem jakości narzędzi pomiarowych (Dolata i in., 2004). Pracownia SUEK IBE, inspirując się międzynarodowymi badaniami edukacyjnymi, pokazała, że skonstruowanie narzędzi służących pomiarowi dydaktycznemu w zgodzie ze światowymi standardami, jest w polskich warunkach możliwe. Zastosowanie podobnej metodologii konstrukcji testów w systemie egzaminacyjnym może więc wpłynąć pozytywnie na jakość stosowanych w nim narzędzi.

Należy zwrócić uwagę, że przyjęta w badaniu SUEK metodologia konstrukcji testu skupia się na funkcjonowaniu poszczególnych zadań. Pozwala na tworzenie testów za pomocą

poszczególnych zadań o określonych parametrach. Tak tworzone testy mogą posiadać własności psychometryczne dopasowane do potrzeb i celów ich konstruktora. Teoria odpowiedzi na zadanie testowe dostarcza również narzędzi pozwalających na ocenę stopnia zrealizowania tych celów.

Należy jednak jeszcze raz podkreślić, że wysoka jakość wszelkich testów, zarówno diagnostycznych jak i tych, które służą selekcji, jest możliwa do osiągnięcia tylko, gdy zadania, z których są skonstruowane, również są dobrej jakości. Doświadczenie pracowni SUEK uczy, że prace nad testami osiągnięć szkolnych są bardzo czasochłonne. Wydaje się, że efektywny proces konstrukcji narzędzi o odpowiedniej rzetelności w metodologii IRT powinien zacząć się od budowy obszernych banków skalibrowanych zadań, które mogłyby służyć zarówno do celów badawczych jak i diagnostycznych. Należy jednak podkreślić, że takie banki zadań musiałyby być utajnione oraz systematycznie uzupełniane, by zapewnić jakość psychometryczną zadań w banku. Nasz system egzaminacyjny, wzbogacony o taki element, uzyskałby dodatkowe możliwości diagnostyczne (np. porównywanie egzaminów między latami), co pozytywnie wpłynęłoby na możliwości prowadzenia polityki oświatowej opartej na faktach.

Na zakończenie warto więc podzielić się doświadczeniami związanymi z samą konstrukcją testów osiągnięć szkolnych. Po pierwsze należy zauważyć, że choć konstrukcja testów zaczyna się od pewnych założeń teoretycznych, to weryfikacja tych założeń zawsze powinna odbywać w empirii. Ostatecznie bowiem, przekonanie o tym, że dany test (bądź pojedyncze zadanie) jest w stanie mierzyć założony konstrukt, jest wiedzą bez większej wartości bez informacji o tym, jak dobrze (precyzyjnie) test (zadanie) to czyni w danej populacji badanych.

Po drugie trzeba podkreślić, że tworzenie testów osiągnięć powinno być procesem przemyślanym i starannie zaplanowanym. Jasna koncepcja i świadomość celu jakiemu ma służyć narzędzie, pozwalają na dobór odpowiednich środków do pomyślnego stworzenia testów. Jedną z pierwszych i ważniejszych kwestii jest wybór teorii pomiaru, która wiąże się z pewnymi założeniami co do konstrukcji narzędzia. W rozdziale tym przedstawiono teorię odpowiedzi na zadanie testowe, której specyficzny przypadek, tzw. model Rascha posiada bardzo dobre własności, pozwalające na konstrukcję testów dostosowanych do potrzeb, oraz ocenę tego jak dobrze dany test sprawdza się w różnych populacjach badanych. Kwestia trafności tak konstruowanych testów jest wynikiem głównie przyjętych metod konstrukcji narzędzia i jakości planu testu, służącego dokładnemu zdefiniowaniu badanego konstrukt.

Przyjęta przez pracownię SUEK metodologia konstrukcji testu posiada jeszcze jedną zaletę, którą dobrze obrazuje zastosowany podczas konstrukcji testów schemat badania pilotażowego. Podczas badania pilotażowego zadania rozwiązywali uczniowie trochę młodsi i trochę starsi niż docelowa populacja. Ponadto zadania pilotowane były w otoczeniu innych zadań niż występowały w badaniu zasadniczym. Zadania wyselekcjonowane do ostatecznych testów

## CZY SZKOŁA MA ZNACZENIE?

na podstawie wyników pilotażu, w przeważającej większości nie straciły swoich dobrych właściwości psychometrycznych podczas badania zasadniczego, a test z nich skonstruowany mierzył zakładany zakres na skali umiejętności z projektowaną dokładnością. Przykład ten pokazuje, że zastosowanie modelu Rascha do konstrukcji testów osiągnięć, umożliwia poluzowanie założenia, że narzędzia do pomiaru efektów kształcenia należy pilotować dokładnie na docelowej populacji, a także dostarcza dowodów, że usytuowanie zadań w różnych kontekstach, co do zasady, nie zmienia znacząco ich właściwości pomiarowych. Stanowi to silny argument praktyczny za szerszym stosowaniem tej metodologii do konstrukcji testów osiągnięć szkolnych w systemie egzaminacyjnym.

### Literatura cytowana

- Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31, 162–172.
- AERA, APA, NCME (1999). *Standards for Educational and Psychological Testing*.
- Anastasi, A., Urbina, S. (1999). *Testy psychologiczne*. Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Brzeziński J., (red.). (2005). *Trafność i rzetelność testów psychologicznych. Wybór tekstów*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York London: The Guilford Press.
- Dolata, R., Putkiewicz, E., Wilkomirska, A. (2004). *Reforma egzaminu maturalnego – oceny i rekomendacje*, Warszawa: Wydawnictwo Instytutu Spraw Publicznych.
- Downing, S. M. (2006a). Twelve Steps for Effective Test Development. W: S. M. Downing, T. M. Haladyna (red.), *Handbook of Test Development*. (s. 3–26). New York London: Routledge.
- Downing, S. M. (2006b). Selected-Response Item Formats in Test Development. W: S. M. Downing, T. M. Haladyna (red.), *Handbook of Test Development*. (s. 287–302). New York London: Routledge.
- Haladyna, T., Downing, S., Rodriguez, M. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment Applied Measurement in Education. *Applied Measurement in Education*, 15(3), 309–334.
- Jakubowski, M., Pokropek, A. (2009). *Badając egzaminy. Podejście ilościowe w badaniach edukacyjnych*. Warszawa: Centralna Komisja Egzaminacyjna.

- Jasińska, A. (2010). Pomiar gotowości szkolnej za pomocą skali quasi-obszaryjnej. W: B. Niemierko, M. K. Szmigiel (red.), *Teraźniejszość i przyszłość oceniania szkolnego* (s. 415–424), Kraków: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- Jasińska, A., Modzelewski, M. (2012). Można inaczej. Wykorzystanie IRT do konstrukcji testów osiągnięć szkolnych. W: B. Niemierko, M. K. Szmigiel (red.), *Regionalne i lokalne diagnozy edukacyjne* (s. 157–168). Wrocław: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- Jasińska, A., Modzelewski, M. (2014). Testy osiągnięć szkolnych TOS3: przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha. *Edukacja*, 2(127), 85–107.
- Kim, S., Feldt, L. S. (2010). The estimation of the IRT reliability coefficient and its lower and upper bounds, with comparisons to CTT reliability statistics. *Asia Pacific Education Review*, 11, 179–188.
- Konarski, R., (2009). *Modele równań strukturalnych. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe PWN.
- Masters, G. N., Wright, B. D., (1997). The Partial Credit Model. W: W. J. van der Linden, R. K. Hambleton (red.), *Handbook of Modern Item Response Theory* (s. 101–122). New York: Springer.
- Pollack, J. M., Rock, D. A., Weiss, M. J., Atkins-Burnett, A., Tourangeau, K., West, J., Hausken, E. G. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). Psychometric Report for the Third Grade*. Washington: National Center for Education Statistics.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago: University of Chicago Press.
- Sprawozdanie z realizacji projektu. Pilotaż nowych egzaminów gimnazjalnych oraz sprawdzianu dla klasy VI. Nowa formuła sprawdzianu.* (2010). Warszawa: Centralna Komisja Egzaminacyjna.
- van der Linden, W. J., Hambleton, R. K. (red.). (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31, 114–128.
- Wu, M., Adams, R. J., Wilson, M. R. i Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised Item Response Modelling Software*. ACER Press.





## Rozdział 3

### Modele analizy zróżnicowania wyników nauczania

#### 3.1. Problem przyczynowości w badaniach edukacyjnych

Badania w nurcie *School Effectiveness Research* są badaniami nomotetycznymi, tzn. ich celem jest odkrywanie zależności przyczynowych między badanymi zjawiskami i na ich podstawie formułowanie w miarę ogólnych praw, które mówią o wpływie różnych czynników na efektywność pracy szkół. W badaniach nad efektywnością pracy szkół zainteresowanie badaczy skoncentrowane jest na określeniu związku pomiędzy charakterystykami szkoły, klasy a wynikami uczniów. W idealnej sytuacji eksperymentu badacze mogliby losowo przydzielić uczniów do szkół (klas), których charakterystyki byłyby zgodnie z hipotezami badawczymi ustalane, zmierzyć początkowy poziom umiejętności, a następnie, po jakimś czasie, np. po zakończeniu roku szkolnego, ponowić badanie i stwierdzić, jakie cechy szkoły wpływają na wyniki uczniów. Niestety, w rzeczywistości możliwości badawcze są dużo skromniejsze. Po pierwsze, uczniowie nie są przypisani do szkoły w sposób losowy, po drugie, badacze nie kontrolują cech szkoły, a po trzecie bardzo często dysponujemy tylko jednym pomiarem interesujących nas zmiennych (są to tzw. badania przekrojowe). W takich warunkach, możliwości formułowania twierdzeń o przyczynowo-skutkowych związkach pomiędzy cechami szkoły a wynikami uczniów są bardzo ograniczone, nie są jednak całkowicie wykluczone.

W badaniach społecznych, by móc mówić o zależnościach przyczynowo-skutkowych między zmiennymi, relacja pomiędzy tymi zmiennymi musi spełniać wszystkie kryteria nomotetycznej przyczynowości: zmienne muszą być skorelowane, przyczyna musi nie później niż skutek, a obserwowana zależność nie może być pozorna (Babbie, 2008, s.110–111). W kontekście badań nad efektywnością kształcenia, a zwłaszcza badań przekrojowych, oznacza to restrykcyjny dobór zmiennych niezależnych, które spełniają powyższe założenia. W kontekście badań przekrojowych zmienne niezależne muszą charakteryzować się dużą stałością w czasie, by uniknąć sytuacji, w której rzeczywisty kierunek zależności mógłby być odwrotny od postulowanego (skutek poprzedza przyczynę) oraz muszą umożliwiać w miarę kompletną charakterystykę badanych zjawisk, aby nie dopuścić do formułowania wniosków, które mogłyby okazać się korelacjami pozornymi.

W praktyce badań nad efektywnością pracy szkoły ten ostatni wymóg sprowadza się do potrzeby szczegółowego opisu cech szkoły, ale także, a może w szczególności, do w miarę

## CZY SZKOŁA MA ZNACZENIE?

wyczerpującego opisu tych cech badanych uczniów, które od szkoły nie są zależne, a są powiązane z osiągnięciami uczniów. Potrzebny jest więc pomiar biopsychicznych i społecznych cech ucznia, które mogą mieć wpływ na osiągnięcia szkolne, a nie są kształtowane przez szkołę. Zbiór takich zmiennych określa się mianem zmiennych kontrolnych. Tylko gdy w analizach kontrolujemy takie cechy ucznia jak inteligencja płynna, płeć, czy status ekonomiczno-społeczny rodziny pochodzenia, możemy zastanawiać się nad wpływem na wyniki uczniów poszczególnych czynników zależnych od szkoły.

### 3.2. Wielopoziomowe modele analizy a struktura danych edukacyjnych

Empiryczne studia edukacyjne są przykładem badań społecznych, w których występuje swoista hierarchia poziomów obserwacji: uczniowie pogrupowani są w oddziały klasowe, klasy zaś istnieją wewnątrz szkół. Hierarchia ta mogłaby iść dalej: moglibyśmy powiedzieć, że szkoły odpowiadają przed jakimś organem zarządzającym, lub że są zlokalizowane w jakiejś miejscowości. Każdy ze szczebli tej hierarchii może być uznany przez nas jako interesujący poziom analizy: moglibyśmy zadać pytanie, czy to, jakie uczniowie osiągają wyniki na egzaminie zależy od cech nauczyciela, albo czy zadowolenie z pracy nauczycieli zależy od stylu zarządzania szkołą przez dyrektora. A może wpływ na obie te rzeczy ma także wielkość środków przeznaczanych na edukację przez jednostkę samorządu terytorialnego? W pierwszym przypadku jednostką analizy są uczniowie i ich wyniki na egzaminie, w drugim nauczyciele i poziom satysfakcji z pracy. W obu jednak przypadkach chcemy zweryfikować hipotezę o istnieniu zależności pomiędzy interesującymi nas charakterystykami jednostek z jednego poziomu a cechami jednostek analizy z wyższego poziomu, do których one przynależą.

Nawet w przypadku, gdy zależności pomiędzy poszczególnymi poziomami tej hierarchii nie są przedmiotem naszych badań – powiedzmy, że jesteśmy zainteresowani tylko indywidualnymi uwarunkowaniami osiągnięć uczniów lub satysfakcji z pracy nauczycieli – to i tak musimy w analizach uwzględniać pogrupowanie jednostek związane z schematem losowania. Zazwyczaj bowiem o wiele łatwiej i z ekonomicznego punktu widzenia rozsądniej jest przebadать wszystkich uczniów lub nauczycieli w nawet kilkudziesięciu (bądź stu kilkudziesięciu) szkołach, niż wylosować i dotrzeć do reprezentatywnej dla Polski prostej próby powiedzmy tysiąca uczniów czy nauczycieli. Hierarchiczne pogrupowanie danych pojawia się w badaniach edukacyjnych z powodu charakteru pytań badawczych (czy czynnik z wyższego poziomu wpływa na cechy jednostek z niższego poziomu) i/lub z powodu zespołowego schematu doboru próby. W obu tych przypadkach musimy wziąć pod uwagę fakt, że jednostki na poszczególnych poziomach są pogrupowane w ramach jednostek na poziomach wyższych.

Pogrupowanie to może prowadzić do tego, że jednostki obserwacji z jednej grupy mogą być do siebie bardziej podobne niż do członków innych grup. Uwzględnienie tego potencjalnego podobieństwa jest bardzo ważne na poziomie analizy danych i wymaga wykorzystania odpowiednich metod statystycznych. Użycie niepoprawnych metod analizy do danych o charakterze hierarchicznym, może prowadzić do błędnych wniosków.

W praktyce analitycznej, dosyć często można się jeszcze dziś spotkać z dwiema niepoprawnymi strategiami radzenia sobie z danymi o charakterze wielopoziomowym (Snijders i Bosker, 2011, s. 14–17). Pierwszą z nich jest tzw. agregacja, która polega na tym, że dane z niższego poziomu agreguje się na wyższy poziom (np. w postaci średnich arytmetycznych), a następnie tak uzyskane agregaty traktuje się jako charakterystyki jednostek wyższego poziomu i poszukuje między nimi zależności na tym właśnie poziomie analizy, a wnioski generalizuje się na jednostki niższego poziomu. Drugą strategią jest procedura odwrotna (z ang. *disaggregation*), polegająca na przypisaniu charakterystyki opisującej pewną zbiorowość poszczególnym jej członkom. W pierwszym przypadku tracimy informacje o zależnościach zachodzących wewnątrz grup, które równie dobrze mogą mieć kierunek odwrotny do tego z wyższego poziomu; w drugim przypadku standardowe procedury statystyczne mogą prowadzić do przyjmowania błędnych wniosków co do testowanych hipotez, zarówno jeśli chodzi o analizę zależności wewnątrzgrupowych jak i międzygrupowych.

Istnieją jednakże metody statystyczne umożliwiające skuteczne analizowanie danych wielopoziomowych. Analizy przedstawione w niniejszym raporcie zostały wykonane z użyciem trypoziomowych hierarchicznych modeli liniowych opisanych szczegółowo np. w Raudenbush i Bryk (2002) oraz Snijders i Bosker (2012) i powszechnie stosowanych w badaniach w nurcie *School Effectiveness Research*. Główną różnicą pomiędzy tradycyjnymi metodami analizy danych, takimi jak analiza wariancji, czy regresja liniowa, a hierarchicznymi modelami liniowymi jest kwestia modelowania składnika błędu. W przypadku zastosowanych w niniejszym raporcie metod, możliwe jest uwzględnienie wielopoziomowej struktury danych oraz faktu, że na każdym poziomie analizy mamy do czynienia z próbą większej populacji.

### 3.3. Zastosowane modele analizy

Hierarchiczne modele liniowe pozwalają na formułowanie i testowanie różnorodnych hipotez. Jak wspomniano już wcześniej ich główną zaletą jest możliwość modelowania zależności na wielu poziomach analizy jednocześnie, a także uwzględnianie wpływu zmiennych z wyższego poziomu na zmienne z poziomu niższego. Jest to możliwe dzięki odmiennemu niż w przypadku, np. regresji liniowej czy analizy wariancji, modelowaniu składnika błędu.

## CZY SZKOŁA MA ZNACZENIE?

Żeby dobrze zrozumieć różnicę pomiędzy klasycznymi metodami analizy, a tymi zastosowanymi na potrzeby niniejszego raportu, warto wyobrazić sobie prosty przykład analizy dwupoziomowej, gdzie na pierwszym poziomie analizy znajdują się uczniowie, a na drugim szkoły. Zmienną zależną są wyniki uczniów na teście z matematyki,  $Y$ , przewidywane za pomocą wskaźnika statusu ekonomiczno społecznego (SES) oznaczonego jako  $X$ . W pojedynczej szkole relacja ta może zostać przedstawiona za pomocą regresji liniowej,

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

gdzie:

$Y_i$ , oznacza wynik w teście dla  $i$ -tego ucznia w szkole,

$X_i$ , oznacza wartość wskaźnika SES dla  $i$ -tego ucznia w szkole,

$\beta_0$  oznacza stałą równania regresji,

$\beta_1$ , oznacza współczynnik nachylenia dla zmiennej  $X$ , a

$e_i$ , to składnik błędu przewidywania dla  $i$ -tego ucznia, po kontroli wpływu SES.

Jeżeli teraz zwiększymy liczbę szkół, to teoretycznie dla każdej z nich moglibyśmy wyliczyć podobne równanie regresji. Prawdopodobnie współczynniki tej regresji,  $\beta_0$  i  $\beta_1$ , będą różnić się pomiędzy szkołami. Można by więc powiedzieć, że dla populacji szkół istnieje pewien rozkład współczynników regresji, z pewną wartością oczekiwaną (średnią) i wariancją. Odchylenia od tej średniej dla każdego ze współczynników regresji są określane jako efekty losowe, natomiast średnie tych rozkładów, stanowiłyby stałą (lub zafiksowaną) część równania. Co ciekawe, dla tak wyliczonych współczynników regresji możemy wyliczyć ich kowariancję (a co za tym idzie również korelację), która zdaje nam sprawę ze związku pomiędzy przewidywanym wynikiem dla ucznia, którego wartość wskaźnika SES równa się 0, a parametrem nachylenia dla szkół. Pozwala to uzyskać pełniejszy obraz zależności pomiędzy tymi zmiennymi na poziomie szkół.

Metoda estymacji nie polega jednak na wyliczaniu poszczególnych linii regresji dla każdej ze szkół, a raczej na oszacowaniu za pomocą iteracyjnej metody najwyższej wiarygodności efektów stałych oraz wariancji i kowariancji efektów losowych. Powyższy przykład pozwala jednak uzmysłowić sobie logikę, która stoi za hierarchicznymi modelami liniowymi.

Na potrzeby analiz przygotowanych do niniejszego raportu, skupiono się głównie na trzech typach modeli trzypoziomowych oraz ich dwupoziomowych odpowiednikach.

Pierwszym typem modeli jest tak zwany model pusty lub bezwarunkowy. Poza zmienną zależną i hierarchicznym pogrupowaniem jednostek analizy nie uwzględnia on żadnych innych

predyktorów. Model ten pozwala na dekompozycję wariancji zmiennej zależnej pomiędzy trzy poziomy analizy. Model ten jest równoważny analizie wariancji z losowymi efektami.

Dla pierwszego poziomu analizy (poziomu uczniów) model ten przybiera następującą postać:

$$Y_{ijk} = \beta_{0jk} + e_{ijk}$$

gdzie:

$Y_{ijk}$  oznacza wartość zmiennej zależnej dla i-tego ucznia w j-tej klasie i k-tej szkole,

$\beta_{0jk}$  oznacza stałą równania regresji, czyli średnią dla zmiennej zależnej w j-tej klasie i k-tej szkole,

$e_{ijk}$  to wyraz błędu przewidywania dla i-tego ucznia w j-tej klasie i k-tej szkole.

Zakłada się, że wyraz błędu  $e_{ijk}$  ma rozkład normalny o średniej 0 i wariancji  $\sigma^2$ .

Dla drugiego poziomu analizy (poziomu klas) model wygląda następująco:

$$\beta_{0jk} = \pi_{00k} + r_{0jk}$$

gdzie:

$\pi_{00k}$  oznacza średnią zmiennej zależnej dla k-tej szkoły, a

$r_{0jk}$  oznacza błąd przewidywania dla j-klasy w k-tej szkole.

Zakłada się, że wyraz błędu  $r_{0jk}$  ma rozkład normalny o średniej 0 i wariancji  $\tau^2$ , oraz że jest nieskorelowany z błędami z pierwszego poziomu.

Wreszcie dla poziomu trzeciego (poziomu szkół) model ten przyjmuje postać:

$$\pi_{00k} = \gamma_{000} + u_{00k}$$

gdzie:

$\gamma_{000}$  oznacza średnią w próbie dla zmiennej zależnej, a

$u_{00k}$  oznacza błąd przewidywania dla stałej dla k-tej szkoły.

Podobnie jak w przypadku błędów dla pozostałych poziomów,  $u_{00k}$  ma rozkład normalny o średniej 0 i wariancji  $\varphi^2$  oraz jest nieskorelowany z błędami z pozostałych poziomów.

Poszczególne wariancje stanowią estymatory komponentów całkowitej wariancji zmiennej zależnej dla każdego poziomu.

Drugim typem modeli jest tak zwany model kontrolny. W modelu tym dołączono szereg predyktorów na poziomie indywidualnym, jednakże współczynniki stojące przy tych zmiennych traktowane są jako stałe na każdym z poziomów analizy. Modele te określa się czasami nazwą modeli z losowymi stałymi. Modele te są równoważne analizie kowariancji z losowymi efektami.

## CZY SZKOŁA MA ZNACZENIE?

Dla pierwszego poziomu analizy, model przyjmuje następującą postać:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} X_{1ijk} + \dots + \beta_{pjk} X_{pijk} + e_{ijk}$$

gdzie:

$\beta_{0jk}$  oznacza stałą równania regresji, czyli przewidywany wynik dla zmiennej zależnej w j-tej klasie i k-tej szkole, dla uczniów, dla których zmienne niezależne przyjmują wartość 0,

$\beta_{1jk} \dots \beta_{pjk}$  to parametry nachylenia dla poszczególnych od 1 do P zmiennych niezależnych dla j-tej klasy w k-tej szkole

$X_{1ijk} \dots X_{pijk}$  to poszczególne zmienne niezależne z pierwszego poziomu, a

$e_{ijk}$  to wyraz błędu przewidywania dla i-tego ucznia w j-tej klasie i k-tej szkole, po uwzględnieniu wpływu zmiennych  $X_{1ijk} \dots X_{pijk}$ .

Poziom drugi uwzględnia poszczególne współczynniki regresji, jednakże tylko stała regresji posiada efekty losowe:

$$\beta_{0jk} = \pi_{00k} + r_{0jk},$$

$$\beta_{1jk} = \pi_{10k},$$

$$\beta_{pjk} = \pi_{p0k},$$

gdzie:

$\pi_{10k} \dots \pi_{p0k}$  oznacza współczynniki nachylenia dla zmiennych niezależnych w k-tej szkole,

Podobnie na poziomie trzecim, współczynnik dla stałej jest uwolniony, natomiast współczynniki nachylenia są zafiksowane:

$$\pi_{00k} = \gamma_{000} + u_{00k},$$

$$\pi_{10k} = \gamma_{100},$$

$$\pi_{p0k} = \gamma_{p00},$$

gdzie:

$\gamma_{100} \dots \gamma_{p00}$  oznacza zagregowane współczynniki nachylenia dla zmiennych niezależnych.

Założenia dotyczące błędów są identyczne jak dla modelu pustego.

Trzecim typem modeli stosowanych na potrzeby niniejszego raportu są modele, w których dodano kolejne zmienne niezależne z różnych poziomów analizy lub uwolniono parametry nachylenia dla poszczególnych predyktorów. Modele te są opisywane w kolejnych

rozdziałach tego raportu. Ich główną zaletą jest możliwość testowania różnorodnych hipotez dotyczących związków pomiędzy cechami szkoły bądź klasy a wynikami kształcenia przy kontroli zmiennych, na które szkoła nie ma wpływu. Uwolnienie nachyleń dla zmiennych niezależnych na poszczególnych poziomach ma doniosłe znaczenie dla estymacji wyrazu błędu na tym poziomie, poza wariancjami dla poszczególnych efektów losowych, szacowane są również kowariancje pomiędzy nimi.

W dalszej części niniejszego rozdziału omawiane są modele kontrolne dla poszczególnych umiejętności kształconych przez szkołę.

### 3.4. Problem wag w modelach wielopoziomowych

W badaniach społecznych, w których użyto złożonego schematu doboru próby, jednostki doboru mogły mieć różne prawdopodobieństwo dostania się do próby. Do uzyskania nieobciążonych estymatorów parametrów populacyjnych (takich jak średnia, czy wariancja zmiennej) konieczne jest użycie odpowiednich wag. Użycie wag pozwala również na poprawne wyliczenie błędów standardowych dla oszacowanych parametrów. Wagi, które pozwalają na wyliczenie tych estymatorów opisane są w rozdziale 1. Przyjmuje się, że wagi te powinny być również uwzględnione w hierarchicznych modelach liniowych (Snijders i Bosker, 2012, s. 216).

Problem wykorzystywania wag próbkowania w modelach wielopoziomowych jest obecnie ciągle dyskutowany i różne rozwiązania są rozwijane. Nie ma jednak ustaleń co do tego, które podejście jest najlepsze, czyli zapewnia wyliczenie najmniej obciążonych estymatorów (Snijders i Bosker, 2012, s. 225). Pociągające jest jednak to, że różne z przyjętych sposobów wykorzystywania wag w modelach dwupoziomowych dają bardzo zbliżone rezultaty (Pfeffermann, Skinner, Holmes, Goldstein, Rasbach, 1998). Nadal brakuje jednak w literaturze jasnych wskazań odnośnie wyliczania wag na potrzeby modeli trypoziomowych. Podaje się kilka propozycji przeskalowania wag, wszystkie one dotyczą natomiast modeli dwupoziomowych. Istnieje jednak przeświadczenie, że dla poprawnego szacowania parametrów i błędów standardowych w przypadku modeli dwupoziomowych dla jednostek z pierwszego poziomu analizy wagi powinny być przeskalowane. Pfeffermann i współpracownicy (1998) opisali metody przeskalowania wag warunkowych dla pierwszego poziomu w przypadku w pełni informatywnych i częściowo informatywnych (informatywnych tylko na poziomie drugim) schematów doboru próby.

Analizy przedstawione w niniejszym raporcie zostały przeprowadzone z wykorzystaniem wag warunkowych na wszystkich trzech poziomach z użyciem programu HLM 6.06. Zgodnie



## CZY SZKOŁA MA ZNACZENIE?

z opisem zawartym w pomocy użytkownika, program ten wykorzystuje metody przeskalowania zgodne z metodą Pfeiffermanna i aplikuje je zarówno dla dwu- i trzypoziomowych modeli (Scientific Software International, 2012).

### 3.5. Opis modeli kontrolnych opracowanych na potrzeby dalszych analiz

Proces budowania modelu kontrolnego na potrzeby analiz uwarunkowań efektywności kształcenia składał się z kilku etapów. Najpierw obliczono trzypoziomowe modele puste stanowiące punkt odniesienia dla modeli ze zmiennymi niezależnymi. Następnie kolejno dodawano zmienne wyjaśniające z poziomu uczniów, badając istotność oszacowanych efektów stałych i poprawę dopasowania modelu. Testowano znaczenie uwzględnienia w modelach różnego sposobu rekodowania wykształcenia rodziców oraz wykorzystania wskaźników dotyczących statusu społecznego rodziny, dla każdego z rodziców osobno i jako miary łącznej. W dalszej części rozdziału skoncentrowano się na wynikach analiz z zastosowaniem tych wskaźników, które mają lepszą moc predykcyjną i czytelniejszą interpretację. Badano także istotność efektów interakcyjnych, w szczególności znaczenie płci dla siły związku z innymi zmiennymi. W celu zwiększenia czytelności wyników w tabelach 3.2– 3.4 przedstawiono tylko wybrane modele.

Zanim przejdziemy do prezentacji wyników analiz, opiszemy wykorzystane w nich zmienne niezależne. Za biopsychiczne korelaty osiągnięć szkolnych, których znaczenie badano w modelu kontrolnym, można uznać płeć (kodowaną 0 dla chłopców i 1 dla dziewcząt), inteligencję płynną i wiek biologiczny. Poziom inteligencji płynnej uczniów mierzono za pomocą Testu Matryc Ravena. Wyniki tego testu zostały wyskalowane wielogrupowym, dwuparametrycznym modelem IRT. W analizach wykorzystano estymatory EAP (*expected a posteriori*), przeliczone na standardową skalę o średniej 100 i odchyleniu standardowym 15 w populacji uczniów. Szczegóły dotyczące charakterystyki tej zmiennej wyjaśniającej można znaleźć w rozdziale poświęconym związkom inteligencji z wynikami nauczania.

Wiek uczniów wyrażono w tygodniach w czasie pisania testów osiągnięć szkolnych. Dodatkowo w modelach uwzględniono dwie zmienne kategoryjne opisujące uczniów o opóźnionym lub przyspieszonym toku kształcenia. Zmienna reprezentująca uczniów o opóźnionym toku kształcenia przyjmuje wartość 1 dla uczniów starszych niż uczniowie z głównej kohorty oraz 0 dla pozostałych. Przyczyny opóźnionego toku kształcenia mogą być związane zarówno z wydłużeniem toku kształcenia (powtarzanie któregoś roku nauki) lub z późniejszym posłaniem dziecka do szkoły. Podobnie zmienna opisująca przyspieszony tok kształcenia przyjmuje wartość 1 dla uczniów młodszych niż główna

kohorta wiekowa oraz 0 dla pozostałych. Uczniów o opóźnionym i przyspieszonym toku kształcenia było niewiele w wylosowanej próbie: odpowiednio 1,75% i 0,88%, dlatego nie rozdrabniano już tych kategorii w zależności od liczby lat, które dzieliły tych uczniów od głównej kohorty wiekowej.

Kolejną grupą czynników, których znaczenie było badane w modelu kontrolnym, były zmienne opisujące status społeczno-ekonomiczny rodziny pochodzenia ucznia. Wzięto tu pod uwagę trzy wskaźniki: najwyższy status zawodowy rodziców (HISEI), indeks zasobów materialnych rodziny dziecka oraz wykształcenie rodziców. Dane na temat tych charakterystyk zebrano za pomocą ankiet wypełnianych przez rodziców.

Seria standardowych pytań pozwoliła zebrać dane potrzebne do zakodowania wykonywanych przez rodziców zawodów zgodnie z klasyfikacją ISCO-08 (International Standard Classification of Occupations). Poszczególnym wartościom ISCO-08 przypisano odpowiadające im wartości indeksu statusu społeczno-ekonomicznego ISEI-08 (International Socio-Economic Index of Occupational Status). Wartość tego indeksu jest ważoną średnią wykształcenia, mierzonego liczbą lat nauki i zarobków wśród reprezentantów poszczególnych grup zawodowych, przebadanych w ramach międzynarodowego badania standaryzacyjnego (metodologia badania została opisana w: Ganzeboom, Treiman, 1996, 2003). W badaniu SUEK zebrano dane dotyczące zarówno ojca jak i matki, a informacje te podawał ten rodzic, który wypełniał ankietę. Wskaźnik HISEI zbudowano w taki sposób, że przyjmuje on wyższą wartość ze wskaźników ISEI obojga rodziców lub jedyną dostępną, jeśli dla jednego z rodziców nie posiadano danych (por. OECD, 2009).

Indeks zasobów materialnych rodziny dziecka został opracowany w oparciu o odpowiedzi na pytania ułożone na podstawie analogicznych pytań w kwestionariuszach uczniowskich i rodzicielskich badania PISA (OECD, 2009, 2006). Dostosowanie koncepcji tego wskaźnika do potrzeb badania SUEK polegało na uwzględnieniu w nim większej liczby pozycji odnoszących się do dóbr mogących wspierać dziecko w nauce, a zrezygnowaniu z kilku powszechnie posiadanych w Polsce ogólnych dóbr materialnych (odtwarzacz DVD, telewizor, zmywarka, telefon komórkowy, samochód). W ten sposób wskaźnik zasobów gospodarstw domowych wykorzystany w badaniu SUEK w większym stopniu koncentruje się na pomocach i dobrach związanych z edukacją i tym samym jest on silniej związany z kapitałem kulturowym rodziny. W celu zbudowania wskaźnika zbadano strukturę danych za pomocą eksploracyjnej analizy czynnikowej przeprowadzonej w oparciu o macierz korelacji polichorycznych, z rotacją ortogonalną varimax, z normalizacją Kaisera. Do indeksu włączono te pozycje, które istotnie były związane z wymiarem posiadania ogólnych zasobów (wyłączono pozycję posiadania przez dziecko telefonu komórkowego, która nie wykazała istotnego związku ze skalą).

## CZY SZKOŁA MA ZNACZENIE?

Ostatecznie indeks ten został zbudowany na podstawie pytań o posiadaną liczbę książek w domu (zarówno dla dzieci, jak i pozostałych), posiadanie oddzielnego pokoju dla dziecka, biurka do nauki wyłącznie dla dziecka, spokojnego miejsca do nauki dla dziecka, komputera, którego dziecko może używać do nauki, dostępu do internetu, edukacyjnych programów komputerowych, encyklopedii, słowników, atlasów, globusa, modelu Układu Słonecznego, lunety, mikroskopu, modelu anatomicznego człowieka, instrumentów muzycznych, sztalug, palet, akcesoriów do malowania, gier edukacyjnych, puzzli edukacyjnych, szachów, kalkulatora oraz aparatu fotograficznego. Do obliczeń wykorzystano pakiety polycor, stats, psych oraz GPA rotation programu R. Następnie wartości indeksu wyliczane były za pomocą estymatorów EAP w skalowaniu IRT modelem 2PL/GRM, metodą MML (*Marginal Maximum Likelihood*). Do obliczeń wykorzystano pakiet ltm programu R. Na koniec wyniki przeliczono na skalę standardową o średniej 100 i odchyleniu standardowym 15 w populacji uczniów. Bardziej szczegółowe informacje na temat tego indeksu znajdują się w rozdziale 7.

Wykształcenie rodziców było mierzone za pomocą pytań ankietowych z dziesięciokategoryjną skalą. Dane te posłużyły do zbudowania wskaźnika najwyższego wykształcenia rodziców, który przyjmował wyższą wartość ze wskaźników wykształcenia obojga rodziców lub jedyną dostępną, jeśli dla jednego z rodziców nie posiadano danych. Na potrzeby analiz zdecydowano się zrekodować tę zmienną na mniejszą liczbę zmiennych zerojedynkowych, które będą mogły być wykorzystane w modelach regresji. Rekodowania tego dokonano w taki sposób, by rozkład nowych zmiennych był jak najbardziej równomierny oraz by scalono ze sobą kategorie najbardziej podobne ze względu na wartości zmiennych zależnych (wyniki testów osiągnięć szkolnych dzieci). W modelach testowano dwa sposoby rekodowania: z pięcioma i czterema kategoriami wykształcenia, jednak większa liczba kategorii nie dawała istotnej poprawy dopasowania modelu, dlatego zdecydowano się przyjąć model prostszy z czterema kategoriami wykształcenia rodziców. W tabeli 3.1. zaprezentowano sposób rekodowania i rozkłady procentowe dla wykorzystywanego w analizach wskaźnika najwyższego wykształcenia rodziców oraz dla porównania osobno dane dla ojców i matek. Podstawą procentowania jest liczebność próby badawczej. Odpowiedzi „nie wiem” przekodowano na braki danych. Zastosowanie w analizach łącznego wskaźnika wykształcenia pozwoliło zmniejszyć liczbę przypadków, które muszą zostać usunięte z analiz z powodu braków danych, a także zmniejszyć liczbę estymowanych parametrów modelu. Ponadto uwzględnienie w analizach osobnych wskaźników dla ojców i matek nie prowadziło do sprzecznych wniosków.

**Tabela 3.1. Rozkłady procentowe wykształcenia rodziców w próbie badawczej**

	Kategoria wykształcenia	najwyższe wykształcenie rodziców		wykształcenie ojca	wykształcenie matki
1	nieukończone podstawowe, podstawowe lub zasadnicze zawodowe	N=1477	26,26%	N=2691 47,84%	N=1787 31,77%
2	średnie – technikum, liceum zawodowe, średnie – liceum profilowane lub średnie – liceum ogólnokształcące	N=1828	32,50%	N=1677 29,81%	N=1712 30,44%
3	policealne lub pomaturalne, licencjackie lub inżynierskie	N=872	15,50%	N=416 7,40%	N=863 15,34%
4	wyższe studia magisterskie lub lekarskie, posiada stopień doktora, doktora habilitowanego lub tytuł profesora	N=1405	24,98%	N=717 12,75%	N=1199 21,32%
5	braki danych lub odpowiedź „nie wiem”	N=43	0,76%	N=124 2,20%	N=64 1,14%
	Razem	N = 5625		N = 5625	N = 5625

Źródło: opracowanie własne

Zmienne zależne wykorzystane w analizach zostały szczegółowo opisane w rozdziale 2. W tym miejscu wspomnimy tylko, że w analizach wykorzystano pięć wartości PV (*plausible value*) wyrażonych na standardowej skali o średniej równej 100 i odchyleniu standardowym równym 15 w populacji uczniów.

### 3.6. Prezentacja i omówienie wyników

Tabele 3.2–3.4 przedstawiają wyniki trzypoziomowych analiz regresji prowadzących do zbudowania modeli kontrolnych dla analiz efektywności nauczania. Zaprezentowano w nich najpierw tzw. modele puste z losową stałą, wyjaśniające wyniki testów tylko przez pogrupowanie uczniów w oddziałach i szkołach, następnie modele z dodanymi efektami stałymi na poziomie ucznia – zmiennymi wyjaśniającymi będącymi biopsychologicznymi korelatami osiągnięć oraz modele uwzględniające kolejną grupę zmiennych niezależnych z poziomu ucznia opisującą status społeczno-ekonomiczny rodziny pochodzenia. Pierwsza

## CZY SZKOŁA MA ZNACZENIE?

z tabel podsumowuje modele dla testu umiejętności czytania, kolejna dla testu świadomości językowej, a w ostatniej przedstawiono wyniki dla testu matematycznego.

Analizy wykonano za pomocą oprogramowania HLM 6.06, wykorzystując metodę estymacji *full maximum likelihood*. Do oszacowania istotności efektów stałych wykorzystano odporne (*robust*) błędy standardowe<sup>1</sup>. Wartości istotne na poziomie istotności  $p < 0,05$  zaznaczono pogrubionym drukiem. Wszystkie konkurencyjne modele dla danego testu osiągnięć szkolnych policzono dla tej samej liczby obserwacji, by zapewnić porównywalność między modelami. Dlatego z analiz usunięto wszystkie obserwacje z brakami danych dla zestawu zmiennych wykorzystanych w porównywanych modelach dla danej zmiennej zależnej. Dla obserwacji bez braków danych policzono wagi dla każdego poziomu analizy uwzględniające sposób doboru próby. Analizy wykonano z uwzględnieniem wag odpowiednio przeskalowanych za pomocą programu HLM.

Model (1) w każdej tabeli to „model pusty”, uwzględniający tylko informację o pogrupowaniu uczniów w oddziałach i szkołach. W modelu tym uwzględniamy trzy efekty losowe: efekt losowy z trzeciego poziomu, będący odchyleniem średnich szkół od średniej w populacji, efekt losowy z drugiego poziomu, opisujący odchylenie średnich klas od średnich szkół, do których należą oraz efekt losowy z pierwszego poziomu, czyli odchylenie wyników indywidualnych od średnich klas (tzw. reszty z pierwszego poziomu). Model ten dostarcza oszacowań dla czterech podstawowych parametrów: stałej regresji (będącej średnią wyniku testu osiągnięć w populacji szkół), wariancji efektów na poziomie szkół ( $\varphi^2$ ), wariancji efektów na poziomie klas ( $\tau^2$ ) i wariancji na poziomie uczniów ( $\sigma^2$ ), czyli wariancji wewnątrzgrupowej. Wariancje charakteryzują zróżnicowanie międzyszkolne, zróżnicowanie na poziomie oddziałów klasowych i wewnątrzoddziałowe. Pozwalają one na wyznaczenie trzech współczynników korelacji wewnątrzgrupowych (*intraclass correlation*). Współczynniki te pokazują, jaki jest stosunek wariancji efektu losowego z danego poziomu do całkowitej zmienności danej cechy. Są więc miarą homogeniczności kategorii z wyższego poziomu analizy.

Współczynnik korelacji wewnątrzgrupowej dla trzeciego poziomu (poziomu szkół) informuje, jaki odsetek zmienności wyników testu osiągnięć można przypisać podziałowi na szkoły i jest on wyznaczany wzorem:

$$\rho_3 = \frac{\varphi^2}{\varphi^2 + \tau^2 + \sigma^2}$$

<sup>1</sup> Są to oszacowania błędów standardowych odporne na naruszenie założeń dotyczących normalności rozkładu reszt (patrz np. Konarski, 2009, s. 318–324).

Współczynnik korelacji wewnątrzgrupowej dla poziomu klas zdaje sprawę z tego, jaki odsetek zmienności wyników testu jest wyjaśniany przez pogrupowanie uczniów na klasy i jest on opisany wzorem:

$$\rho_2 = \frac{\tau^2}{\varphi^2 + \tau^2 + \sigma^2}$$

Trzeci współczynnik korelacji wewnątrzgrupowej informuje, jaki odsetek wariancji wyników testu przypada na szkoły i klasy łącznie:

$$\rho_{3+2} = \frac{\varphi^2 + \tau^2}{\varphi^2 + \tau^2 + \sigma^2}$$

### 3.6.1. Wyniki analiz dla testu umiejętności czytania

Wyniki analiz dla modelu (1) przedstawione w tabeli 3.2 wskazują na istnienie istotnej wariancji wyników testu umiejętności czytania zarówno na poziomie uczniów, oddziałów i szkół. Znacznie większe zróżnicowanie wyników obserwuje się na poziomie uczniów, jednak wielopoziomowy charakter badanego zjawiska nie budzi wątpliwości. Wartości współczynników korelacji wewnątrzgrupowej wskazują na to, że podział uczniów na szkoły wyjaśnia ok. 4% zmienności wyników testu, podział na klasy – podobnie ok. 4%, czyli ok. 8% zmienności wyników testu umiejętności czytania daje się wyjaśnić za pomocą pogrupowania na klasy i szkoły rozpatrywanego łącznie.

Uważny czytelnik mógł zwrócić uwagę na to, że stała regresji nie jest równa 100, mimo że zmienna zależna jest wyrażona na skali o średniej równej 100 i odchyleniu standardowym 15 w populacji uczniów. Te niewielkie różnice dla każdego testu osiągnięć są spowodowane tym, że wyniki testów były standaryzowane dla wszystkich dostępnych danych, a do analiz regresji włączono tylko te obserwacje, które nie miały braków danych dla zmiennych uwzględnionych w testowanych modelach. Ponadto stała regresji jest średnią średnich szkół wyznaczaną na podstawie przeskalowanych na potrzeby analiz wielopoziomowych wag, natomiast średnia zmiennej wyjaśnianej jest średnią wyników indywidualnych wyznaczoną z wykorzystaniem nieprzeskalowanych wag próbkowania, czyli jest najlepszym z możliwych estymatorem średniego wyniku w populacji uczniów.

## CZY SZKOŁA MA ZNACZENIE?

**Tabela 3.2. Uwarunkowania osiągnięć w zakresie umiejętności czytania: model kontrolny. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych**

Zmienna zależna: wynik testu umiejętności czytania	Model (1)	Model (2)	Model (3)
<i>oszacowanie efektów stałych</i>			
<b>poziom ucznia</b>			
test matryc ravena		<b>0,486</b> (0,018)	<b>0,405</b> (0,018)
pleć dziecka <sup>a</sup>		<b>2,86</b> (0,54)	<b>3,12</b> (0,53)
wiek dziecka w tygodniach		<b>0,048</b> (0,014)	<b>0,060</b> (0,014)
opóźniony tok nauki		<b>-10,73</b> (2,03)	<b>-9,14</b> (1,96)
przyspieszony tok nauki		3,75 (3,10)	3,32 (2,73)
HISEI			<b>0,043</b> (0,018)
indeks zasobów materialnych rodziny dziecka			<b>0,126</b> (0,018)
wykształcenie rodziców: średnie <sup>b</sup>			<b>2,21</b> (0,73)
wykształcenie rodziców: policealne lub studia i stopnia <sup>b</sup>			<b>3,76</b> (0,92)
wykształcenie rodziców: wyższe magisterskie, dr lub prof. <sup>b</sup>			<b>5,93</b> (1,09)
stała	<b>99,1</b> (0,44)	<b>23,53</b> (7,92)	8,08 (7,73)
<b>oszacowanie efektów losowych</b>			
wariancja efektów szkół	<b>9,08</b>	<b>10,04</b>	<b>6,50</b>
wariancja efektów oddziałów klasowych	<b>8,97</b>	2,93	1,96
wariancja na poziomie ucznia	<b>205,48</b>	<b>153,78</b>	<b>141,71</b>
<i>współczynniki korelacji wewnątrzgrupowej:</i>			
poziom szkół	0,041	--	--
poziom klas	0,040	--	--
poziom klas i szkół łącznie	0,081	--	--
<b>podsumowanie</b>			
pseudo R <sup>2</sup>	--	0,254	0,328
pseudo R <sup>2</sup> (p3)	--	-0,105	0,284
pseudo R <sup>2</sup> (p2)	--	0,673	0,781
pseudo R <sup>2</sup> (p1)	--	0,252	0,310
liczba uczniów	4738	4738	4738
liczba klas	300	300	300
liczba szkół	172	172	172

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

a – grupa odniesienia: chłopcy

b – grupa odniesienia: wykształcenie nieukończone podstawowe, podstawowe lub zasadnicze zawodowe

Źródło: opracowanie własne

Model (2), oprócz informacji o pogrupowaniu uczniów, zawiera pięć zmiennych wyjaśniających z pierwszego poziomu: wynik Testu Matryc Ravena, płeć, wiek dziecka wyrażony w tygodniach i dwie zmienne zerowejedynkowe pokazujące efekty związane z opóźnionym lub przyspieszonym tokiem kształcenia. Zmienne te zostały zdefiniowane w modelu tylko jako efekty stałe, czyli zakłada się, że zależność między nimi a zmienną zależną jest taka sama w każdej klasie. Zależności dla wszystkich zmiennych, oprócz tej opisującej efekt dla przyspieszonego toku nauki, okazały się istotne statystycznie. Zostaną one opisane przy omawianiu modelu (3). Dodanie tych zmiennych do modelu zredukowało wariancję efektów na poziomie uczniów i klas. Wariancja efektów z drugiego poziomu okazała się nieistotnie różna od zera, co oznacza, że jeśli kontrolujemy biopsychiczne cechy uczniów, podział na klasy nie wyjaśnia zmienności w wynikach testu umiejętności czytania.

Dodanie zmiennych do modelu w naturalny sposób prowadzi do pytania, na ile poprawiło się przewidywanie zmiennej zależnej po uwzględnieniu w modelu zmiennych wyjaśniających. W klasycznej analizie regresji do oceny efektywności danego modelu stosuje się współczynnik  $R^2$ , który informuje o tym, jaką część wariancji zmiennej zależnej wyjaśniają zmienne niezależne uwzględnione w modelu. Analogiczne rozumowanie przenosi się na grunt analiz wielopoziomowych, dekomponując  $R^2$  na tyle składników, ile jest analizowanych poziomów (Raudenbush i Bryk, 2002, s. 73–79). Aby odróżnić tę miarę od wskaźnika  $R^2$  definiowanego klasycznie, przyjęło się nazywać ją *pseudo  $R^2$* , a opisuje się ją wzorem:

$$pseudo R^2 = \frac{(\varphi^2 (M0) + \tau^2 (M0) + \sigma^2 (M0)) - (\varphi^2 (M1) + \tau^2 (M1) + \sigma^2 (M1))}{(\varphi^2 (M0) + \tau^2 (M0) + \sigma^2 (M0))}$$

Pseudo  $R^2$  mierzy stopień redukcji niewyjaśnionej wariancji w modelu ze zmiennymi wyjaśniającym modelem pustym w stosunku do i. We wzorze tym wariancję efektów na poziomie szkół oznaczono  $\varphi^2$ , wariancję efektów na poziomie klas  $\tau^2$ , a wariancję na poziomie uczniów  $\sigma^2$ . Model pusty został oznaczony symbolem M0, a model ze zmiennymi wyjaśniającymi symbolem M1.

Współczynnik ten można zdekomponować na poszczególne poziomy. I tak dla przykładu dla poziomu szkół mamy:

$$pseudo R^2_{(p3)} = \frac{(\varphi^2 (M0)) - (\varphi^2 (M1))}{(\varphi^2 (M0))}$$

A dla poziomu klas:

$$pseudo R^2_{(p2)} = \frac{(\tau^2 (M0)) - (\tau^2 (M1))}{(\tau^2 (M0))}$$

Analogicznie wyznaczono pseudo  $R^2$  dla poziomu pierwszego.



## CZY SZKOŁA MA ZNACZENIE?

W raporcie tym zdecydowano się wykorzystać te miary z uwagi na to, że są one dość powszechnie stosowane w takich analizach (Domański i Pokropek, 2011), jednak należy zdawać sobie sprawę z tego, że w kontekście hierarchicznych modeli liniowych pojęcie odsetka wyjaśnionej wariancji jest problematyczne i stosowanie tak zdefiniowanej miary może prowadzić np. do uzyskania ujemnych wartości pseudo  $R^2$  (Snijders i Bosker, 2012), co chociażby z interpretacyjnego punktu widzenia jest kłopotliwe.

Z sytuacją taką mamy do czynienia dla pseudo  $R^2_{(p3)}$  dla modelu (2), co związane jest ze wzrostem wariancji na poziomie szkół po dodaniu zmiennych wyjaśniających. Jednak różnica wariancji efektów na poziomie szkół między modelem pustym i modelem (2) nie jest istotna statystycznie. Zgodnie z przyjętą interpretacją wskaźnika pseudo  $R^2$  można powiedzieć, że dodanie do modelu zmiennych wyjaśniających opisujących biopsychologiczne korelaty osiągnięć, zredukowało odsetek niewyjaśnionej wariancji o ok. 25%. Redukcja odsetka niewyjaśnionej wariancji była największa na poziomie klas.

Model (3) poza zmiennymi uwzględnionymi w modelu (2) zawiera także zmienne z pierwszego poziomu będące wskaźnikami statusu społeczno-ekonomicznego rodziny ucznia. Zmienne te zostały zdefiniowane w modelu tylko jako efekty stałe. Model w takim kształcie został uznany za najlepszy na potrzeby dalszych analiz model kontrolujący podstawowe, niezależne od szkoły czynniki warunkujące osiągnięcia szkolne. Badano także interakcje między zmiennymi z tego modelu, które znajdowały uzasadnienie w literaturze, jednak okazały się one nieistotne. Skupmy się jednak na zmiennych, które weszły w skład modelu kontrolnego.

Wynik testu Ravena okazał się, zgodnie z przewidywaniami, istotnie pozytywnie związany z wynikiem testu umiejętności czytania. Przeciętna różnica w wynikach w teście umiejętności czytania uczniów różniących się poziomem inteligencji o jedno odchylenie standardowe (15 punktów na skali wyniku Testu Matryc Ravena), wynosi ok. 6 punktów (niecałe pół odchylenia standardowego) na korzyść uczniów o wyższej inteligencji. Różnica między chłopcami i dziewczynkami także okazała się istotna statystycznie. Dziewczynki mają średnio o ok. 3 punkty wyższe wyniki w teście czytania niż chłopcy. Kierunek tej zależności jest spójny z wynikami innych badań (np. Konarzewski, 2007).

Wiek dziecka również okazał się znaczący, choć zależność nie jest silna. Biorąc pod uwagę główną kohortę wiekową, uczniowie rozwiązujący testy osiągnięć w badaniu SUEK mogli różnić się ze względu na wiek maksymalnie o rok (52 tygodnie). Średnia różnica w wynikach testu między takimi uczniami wynosi ok. 3 punkty (jedna piąta odchylenia standardowego). Zależność ta jest trochę silniejsza niż stwierdzona dla wyników na sprawdzianie w 6 klasie szkoły podstawowej i w związku z tym wpisuje się w obserwowany trend osłabiania się z czasem znaczenia wieku dla wyników testów osiągnięć (Dolata i Pokropek, 2012).

Uczniowie o opóźnionym toku nauki, czyli albo uczniowie, którzy później rozpoczęli formalną edukację albo ci, którzy raz lub więcej powtarzali naukę w tej samej klasie, mają znacznie niższy poziom umiejętności czytania. Otrzymali oni średnio o ponad 9 punktów w teście czytania mniej. Nie dziwi to o tyle, że w grupie tej (choć nielicznej: w próbie badawczej było 97 takich uczniów, co stanowi 1,75%) znaleźli się uczniowie, którzy przede wszystkim ze względu na trudności w nauce lub powody z tym ściśle związane mieli opóźniony tok kształcenia.

Efekt dla uczniów, którzy szybciej niż główna kohorta wiekowa rozpoczęli formalną naukę, okazał się nieistotny statystycznie. Należy mieć jednak na uwadze to, że uczniów tych było tylko 49 (0,88%) w badanej próbie, co znalazło także odzwierciedlenie w dużym błędzie standardowym. Zmienną tę zdecydowano się jednak zostawić w modelu kontrolnym, by móc w sposób bardziej poprawny szacować związek wieku z osiągnięciami szkolnymi dla głównej kohorty. Wiadomo bowiem, że dzieci wcześniej posłane do szkoły nie są losową próbką danej kohorty.

Czynniki związane ze statusem społeczno-ekonomicznym rodziny są, zgodnie z założeniami i wynikami wcześniejszych badań, pozytywnie związane z wynikami testu. Dla umiejętności czytania wszystkie wskaźniki statusu rodziny okazały się miarami dopełniającymi, z których każda wyjaśnia część zmienności wyników testu. Im wyższy status społeczno-ekonomiczny rodziny mierzony wskaźnikiem HISEI, opartym na wykonywanym przez rodziców zawodzie, tym wyższy poziom umiejętności czytania. Zależność ta jest raczej słaba, choć istotna statystycznie. Maksymalna różnica w wynikach testu między dziećmi z rodzin o najniższej i najwyższej wartości wskaźnika HISEI, pod kontrolą pozostałych zmiennych, to 3,32 punktu na standardowej skali testu umiejętności czytania.

Zasobność gospodarstw domowych, z których pochodzą uczniowie, jest również pozytywnie i istotnie związana z ich poziomem umiejętności czytania. Wraz ze wzrostem wartości tego indeksu o jeden, wyniki testu rosną przeciętnie o 0,126, czyli wzrost o jedno odchylenie standardowe (15 punktów na skali, w której jest wyrażone), daje przeciętny przyrost o niecałe 2 punkty w teście umiejętności czytania.

Wykształcenie rodziców zostało uwzględnione w modelu za pomocą zmiennych zerojedynkowych, dla których kategorią odniesienia są dzieci pochodzące z rodzin, w których najwyższym poziomem wykształcenia wśród rodziców jest niepełne podstawowe, podstawowe lub zasadnicze zawodowe. Uczniowie, których co najmniej jeden rodzic ma wykształcenie średnie, a drugi średnie lub niższe, uzyskują na teście umiejętności czytania średnio o ponad 2 punkty wyższy wynik od kolegów z kategorii odniesienia. Uczniowie, których rodzice mają wykształcenie pomaturalne lub licencjackie, otrzymali średnio o prawie 4 punkty więcej, a uczniowie, których co najmniej jeden z rodziców ma co najmniej wykształcenie wyższe magisterskie, zyskują prawie 6 punktów w stosunku do rówieśników z kategorii odniesienia.

## CZY SZKOŁA MA ZNACZENIE?

Problematyka związku cech rodziny pochodzenia z wynikami osiągnięć uczniów została bardziej szczegółowo zanalizowana i omówiona w oddzielnym rozdziale.

Dodanie do modelu opisanych wyżej zmiennych zredukowało odsetek niewyjaśnionej wariacji o ok. 33% w stosunku do modelu pustego. Największy odsetek redukcji niewyjaśnionej wariacji obserwujemy na poziomie klas (patrz pseudo  $R^2_{(p,2)}$ ), jednak uwzględnienie kolejnych zmiennych wyjaśniających, zmniejszyło także wariację na poziomie szkół i uczniów. Wariancja efektów losowych na poziomie klas jest dla modelu kontrolnego nieistotna statystycznie. Oznacza to, że jeśli kontrolujemy zmienne uwzględnione w modelu (3), nie możemy stwierdzić, że podział uczniów na klasy wyjaśnia jakąś część zmienności wyników testowych. Innymi słowy: jeśli skład oddziałów klasowych nie różniłby się ze względu na wartości zmiennych z modelu (3), nie moglibyśmy odrzucić hipotezy mówiącej o tym, że poszczególne oddziały wewnątrz szkół nie różnią się istotnie poziomem umiejętności czytania. Część zróżnicowania wyników moglibyśmy tłumaczyć podziałem na szkoły, a zdecydowana większość obserwowanej wariacji jest wariacją wewnątrzoddziałową. Ponieważ bez kontroli zmiennych uwzględnionych w modelu (3) wariancja efektów na poziomie klas była istotna, można przypuszczać, że zaobserwowana w modelu pustym wariacja wyników na drugim poziomie związana jest w dużej mierze z procesem segregacji uczniów do klas. Daje się ją bowiem w dużej części wyjaśnić cechami indywidualnymi uczniów, bez uwzględnienia cech z poziomu klasy. Wyniki te nie mówią jednak o sile czy powszechności tego zjawiska.

### 3.6.2. Wyniki analiz dla testu świadomości językowej

W tabeli 3.3 przedstawiono wyniki analiz, w których zmienną zależną jest wynik testu świadomości językowej. Model (1) wskazuje na istnienie istotnego zróżnicowania wyników na każdym poziomie analizy. Zróżnicowanie to jest największe na poziomie uczniów, jednak wartości współczynników korelacji wewnątrzgrupowej wskazują, że podział uczniów na szkoły wyjaśnia ok. 6% zmienności wyników, a podział na klasy troszkę mniej – ok. 4%. Rozpatrując łącznie pogrupowanie uczniów na szkoły i klasy, można wyjaśnić ok. 10% wariacji wyników testu świadomości językowej.

**Tabela 3.3. Uwarunkowania osiągnięć w zakresie świadomości językowej: model kontrolny. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych**

Zmienna zależna: wynik testu świadomości językowej	Model (1)	Model (2)	Model (3)
<b>Oszacowanie efektów stałych</b>			
<i>Poziom ucznia</i>			
Test Matryc Ravena		<b>0,534</b> (0,016)	<b>0,447</b> (0,016)
Płeć dziecka <sup>a</sup>		<b>5,08</b> (0,47)	<b>5,34</b> (0,45)
Wiek dziecka w tygodniach		<b>0,07</b> (0,014)	<b>0,083</b> (0,014)
Opóźniony tok nauki		<b>-14,14</b> (1,74)	<b>-12,40</b> (1,68)
Przyspieszony tok nauki		<b>5,25</b> (2,18)	<b>4,86</b> (1,9)
HISEI			<b>0,064</b> (0,019)
Indeks zasobów materialnych rodziny dziecka			<b>0,125</b> (0,019)
Wykształcenie rodziców: średnie <sup>b</sup>			<b>3,33</b> (0,68)
Wykształcenie rodziców: policealne lub studia I stopnia <sup>b</sup>			<b>4,81</b> (0,93)
Wykształcenie rodziców: wyższe magisterskie, dr lub prof. <sup>b</sup>			<b>5,69</b> (0,96)
Stała	<b>98,89</b> (0,50)	5,66 (7,79)	-11,13 (7,78)
<b>Oszacowanie efektów losowych</b>			
Wariancja efektów szkół	<b>13,84</b>	<b>11,92</b>	<b>6,04</b>
Wariancja efektów oddziałów klasowych	<b>8,86</b>	<b>3,26</b>	1,93
Wariancja na poziomie ucznia	<b>202,64</b>	<b>135,02</b>	<b>121,73</b>
<i>Współczynniki korelacji wewnątrzgrupowej:</i>			
Poziom szkół	0,061	--	--
Poziom klas	0,039	--	--
Poziom klas i szkół łącznie	0,101	--	--
<b>Podsumowanie</b>			
pseudo R <sup>2</sup>	--	0,333	0,424
pseudo R <sup>2</sup> (p3)	--	0,139	0,563
pseudo R <sup>2</sup> (p2)	--	0,632	0,782
pseudo R <sup>2</sup> (p1)	--	0,334	0,399
Liczba uczniów	4738	4738	4738
Liczba klas	300	300	300
Liczba szkół	172	172	172

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

a – grupa odniesienia: chłopcy

b – grupa odniesienia: wykształcenie nieukończone podstawowe, podstawowe lub zasadnicze zawodowe

Źródło: opracowanie własne

## CZY SZKOŁA MA ZNACZENIE?

Uwzględnienie w analizie efektów stałych dla biopsychologicznych korelatów osiągnięć, zdefiniowanych na pierwszym poziomie, pokazało istnienie istotnych zależności pomiędzy nimi a wynikami testu świadomości językowej oraz zredukowało niewyjaśnioną wariancję o 33%. Największy odsetek zredukowanej wariancji zaobserwowano, podobnie jak dla testu umiejętności czytania, na poziomie klas.

Model (3) będzie w większości dalszych analiz przedstawianych w tym raporcie traktowany jako model kontrolujący podstawowe, niezależne od szkoły czynniki warunkujące osiągnięcia szkolne. W modelu tym, poza zmiennymi z modelu (2), zostały uwzględnione także wskaźniki statusu społeczno-ekonomicznego rodziny ucznia. Zmienne te zostały zdefiniowane jako efekty stałe na pierwszym poziomie analizy. Podobnie jak dla testu umiejętności czytania, nie znaleziono istotnych statystycznie interakcji między zmiennymi wchodzącymi w skład tego modelu.

Wszystkie zależności badane za pomocą modelu (3) są istotne statystycznie na poziomie istotności  $p < 0,05$ . Siła związku między poziomem inteligencji a wynikami testu osiągnięć jest praktycznie taka sama jak w przypadku testu umiejętności czytania. Wzrost poziomu inteligencji o jedno odchylenie standardowe pociąga za sobą wzrost wyników w teście świadomości językowej średnio o niecałe pół odchylenia standardowego. Dziewczynki osiągają przeciętnie o 5,35 punktu wyższe wyniki niż chłopcy. Różnica ta jest trochę wyższa niż w zakresie umiejętności czytania.

Związek wieku z osiągnięciami szkolnymi dla głównej kohorty wiekowej w obszarze świadomości językowej jest podobny, jak w przypadku testu umiejętności czytania: pozytywny, niezbyt silny, lecz istotny. Uczniowie o opóźnionym toku kształcenia osiągają natomiast wyniki średnio o 12,4 punktu (tj. o niecałe odchylenie standardowe) niższe niż koledzy z głównej kohorty wiekowej. Efekt dla uczniów o przyspieszonym toku nauki okazał się dla testu świadomości językowej dodatni i istotny statystycznie. Uczniowie ci uzyskali średnio wyniki o prawie 5 punktów wyższe niż ich koledzy z głównej kohorty wiekowej, mimo że są to uczniowie młodszy. Potwierdza to tylko fakt, że uczniowie wcześniej rozpoczynający naukę są średnio uczniami o wyższym poziomie uprzednich zdolności i w związku z tym, co do zasady, lepiej radzą sobie w zakresie osiągnięć szkolnych niż przeciętni uczniowie z głównej kohorty wiekowej.

Czynniki związane ze statusem społeczno-ekonomicznym rodziny ucznia są również pozytywnie i istotnie związane z jego osiągnięciami szkolnymi w zakresie świadomości językowej. Siła tych związków jest podobna jak w przypadku testu umiejętności czytania.

Wartość współczynnika pseudo  $R^2$  wskazuje, że uwzględnienie w modelu zmiennych z modelu (3) pozwala zredukować niewyjaśnioną wariancję wyników testu świadomości językowej o ok. 42% w stosunku do modelu pustego. Wariancja efektów losowych zmniejszyła

się na każdym poziomie analizy, jednak największy odsetek redukcji niewyjaśnionej wariancji zaobserwowano na poziomie klas (patrz pseudo  $R^2_{(p2)}$ ). W związku z tym, analogicznie, jak w przypadku testu umiejętności czytania, wariancja efektów losowych na poziomie klas dla modelu (3) okazała się nieistotnie statystycznie różna od zera. Oznacza to, że jeśli kontrolujemy opisane wyżej zmienne, nie możemy odrzucić hipotezy mówiącej o tym, że poszczególne oddziały wewnątrz szkół nie różnią się istotnie wynikami testu świadomości językowej. Natomiast podział na szkoły nadal wyjaśnia znaczącą statystycznie część zróżnicowania wyników.

### 3.6.3. Wyniki analiz dla testu umiejętności matematycznych

Tabela 3.4 prezentuje wyniki analiz, w których zmienną wyjaśnianą są umiejętności matematyczne. Model (1), będący modelem pustym bez dodatkowych zmiennych niezależnych, pokazuje, że podział uczniów na szkoły i klasy wyjaśnia podobną część zmienności wyników jak w przypadku testów językowych – tj. niecałe 10%, jeśli rozpatrywać pogrupowanie łącznie. Wartości współczynników korelacji wewnątrzgrupowej dla każdego z poziomów analizy mają zbliżone wartości, jednak zauważa się trochę większe znaczenie wariancji na poziomie klas.

Modele (2) i (3) są analogiczne do prezentowanych dla testów językowych, dostrzega się tu jednak pewne różnice w wynikach – dwie zmienne, które są znaczące dla przewidywania wyników testów umiejętności czytania i świadomości językowej, nie są istotne, jeśli wyjaśniamy wyniki testu matematycznego. Nie zaobserwowano istotnego związku umiejętności matematycznych z płcią, co oznacza, że wyniki dziewcząt i chłopców średnio nie różnią się od siebie. Porównując ten wynik do wyników innych badań należy mieć na uwadze to, że podczas konstrukcji testów osiągnięć do badania SUEK szczególną uwagę poświęcono analizie stroniczości zadań ze względu na płeć (wykonując analizę efektów DIF). Do testów osiągnięć nie włączono zadań, które charakteryzował silny efekt DIF za względu na płeć, który znajdował prawdopodobne potwierdzenie w treści zadań testowych.

Drugą zmienną, dla której – inaczej niż w przypadku testów językowych – nie wykryto związku z wynikami testu matematycznego, jest wskaźnik HISEI. Pozwala to przypuszczać, że umiejętności matematyczne są w mniejszym stopniu niż osiągnięcia językowe uwarunkowane statusem społecznym rodziny, a większe znaczenie odgrywa inteligencja ucznia (związek z wynikiem Testu Matrycy Ravena jest silniejszy niż dla testu czytania i świadomości językowej). Pozostałe wskaźniki statusu rodziny ucznia pozostają jednak w podobnym związku jak w pozostałych testach osiągnięć.

W związku z powyższym, z modelu kontrolnego dla analiz uwarunkowań efektywności nauczania w zakresie umiejętności matematycznych zdecydowano się wykluczyć płeć i wskaźnik HISEI. Model ten ma nieznacznie niższą wartość wskaźnika pseudo  $R^2$  niż model (3),

## CZY SZKOŁA MA ZNACZENIE?

jednak analiza statystyki *deviance* dla modelu (3) i (4) dla każdej z pięciu analiz wykorzystującej kolejne wartości PV pokazała, że pogorszenie dopasowania modelu po wykluczeniu zmiennych nieistotnych nie jest duże<sup>2</sup>. Model (4) przyjęto za najlepszy model kontrolny dla analiz uwarunkowań efektów kształcenia w zakresie umiejętności matematycznych.

**Tabela 3.4. Uwarunkowania osiągnięć w zakresie umiejętności matematycznych: model kontrolny. Wyniki trzypoziomowych analiz regresji, modele z losową stałą, odporne (robust) oszacowania błędów standardowych**

Zmienna zależna: wynik testu umiejętności matematycznych	Model (1)	Model (2)	Model (3)	Model (4)
<b>Oszacowanie efektów stałych</b>				
<b>Poziom ucznia</b>				
Test Matrycy Ravena		<b>0,616</b> (0,018)	<b>0,543</b> (0,017)	<b>0,544</b> (0,017)
Płeć dziecka <sup>a</sup>		-0,379 (0,524)	-0,20 (0,499)	
Wiek dziecka w tygodniach		<b>0,054</b> (0,015)	<b>0,065</b> (0,015)	<b>0,066</b> (0,015)
Opóźniony tok nauki		<b>-14,55</b> (2,42)	<b>-13,07</b> (2,44)	<b>-13,12</b> (2,42)
Przyspieszony tok nauki		<b>6,21</b> (3,1)	<b>5,92</b> (2,76)	<b>5,99</b> (2,78)
HISEI			0,029 (0,02)	
Indeks zasobów materialnych rodziny dziecka			<b>0,119</b> (0,018)	<b>0,126</b> (0,018)
Wykształcenie rodziców: średnie <sup>b</sup>			<b>3,18</b> (0,56)	<b>3,42</b> (0,55)
Wykształcenie rodziców: policealne lub studia I stopnia <sup>b</sup>			<b>3,25</b> (0,74)	<b>3,82</b> (0,74)
Wykształcenie rodziców: wyższe magisterskie, dr lub prof. <sup>b</sup>			<b>5,44</b> (0,99)	<b>6,52</b> (0,78)
Stała	<b>99,09</b> (0,47)	9,07 (8,15)	-5,41 (8,56)	-5,67 (8,58)
<b>Oszacowanie efektów losowych</b>				
Wariancja efektów szkół	<b>9,06</b>	<b>9,81</b>	<b>6,45</b>	<b>6,58</b>
Wariancja efektów oddziałów klasowych	<b>12,7</b>	<b>6,02</b>	<b>4,45</b>	<b>4,38</b>
Wariancja na poziomie ucznia	<b>203,74</b>	<b>123,36</b>	<b>114,36</b>	<b>114,52</b>
<b>Współczynniki korelacji wewnątrzgrupowej:</b>				
Poziom szkół	0,040	--	--	--
Poziom klas	0,056	--	--	--
Poziom klas i szkół łącznie	0,096	--	--	--

<sup>2</sup> Dla dwóch z pięciu analiz dla kolejnych wartości PV różnica wartości statystyki *deviance* była nieistotna statystycznie przy założonym poziomie istotności  $p < 0,05$ . Dla trzech okazała się istotna, jednak nie była duża (6,92; 8,48 i 14,87). Średnia z różnic dla każdej z pięciu analiz wyniosła 7,34, przy 2 stopniach swobody (w modelu (3) szacowano 14 parametrów, w modelu (4) – 12).

## MODELE ANALIZY RÓŻNICOWANIA WYNIKÓW NAUCZANIA

Zmienna zależna: wynik testu umiejętności matematycznych	Model (1)	Model (2)	Model (3)	Model (4)
<b>Podsumowanie</b>				
pseudo R <sup>2</sup>	--	0,383	0,445	0,443
pseudo R <sup>2</sup> (p3)	--	-0,083	0,288	0,274
pseudo R <sup>2</sup> (p2)	--	0,526	0,650	0,655
pseudo R <sup>2</sup> (p1)	--	0,395	0,439	0,438
Liczba uczniów	4633	4633	4633	4633
Liczba klas	300	300	300	300
Liczba szkół	172	172	172	172

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

a – grupa odniesienia: chłopcy

b – grupa odniesienia: wykształcenie nieukończone podstawowe, podstawowe lub zasadnicze zawodowe

*Źródło: opracowanie własne*

Jak już wspomniano, związek poziomu inteligencji z wynikami testu matematycznego okazał się trochę silniejszy niż w przypadku umiejętności czytania i świadomości językowej. Wzrost wyników testu Ravena o jeden punkt pociąga za sobą średni wzrost o 0,544 punktu w teście matematycznym, co oznacza średnią różnicę o ponad pół odchylenia standardowego na skali umiejętności matematycznych między uczniami różniącymi się o jedno odchylenie standardowe wynikiem testu inteligencji.

Wiek uczniów z głównej kohorty wiekowej ma podobne znaczenie jak w przypadku pozostałych testów. Jest istotnie, choć niezbyt silnie związany z umiejętnościami matematycznymi: uczniowie urodzeni we wcześniejszych miesiącach mają trochę wyższe wyniki niż ci urodzeni później. Uczniowie o opóźnionym toku kształcenia uzyskują wyniki na skali umiejętności matematycznych średnio o prawie jedno odchylenie standardowe niższe niż ich koledzy z głównej kohorty wiekowej. Efekt dla uczniów, którzy wcześniej rozpoczęli formalną naukę w szkole jest także znaczący: otrzymali oni średnio o ok. 6 punktów wyższe wyniki z testu matematycznego niż uczniowie z głównej kohorty wiekowej.

Zasobność gospodarstw domowych, z których pochodzą uczniowie, jest – podobnie jak dla pozostałych testów osiągnięć – również pozytywnie i istotnie związana z poziomem umiejętności matematycznych dzieci. Siła tego związku jest taka sama jak dla osiągnięć językowych. Wraz ze wzrostem wartości tego indeksu o jedno odchylenie standardowe (15 punktów na skali, w której jest wyrażone), wyniki testu matematycznego rosną przeciętnie przyrost o niecałe 2 punkty.

Wykształcenie rodziców także istotnie różnicuje osiągnięcia szkolne uczniów w zakresie umiejętności matematycznych. Zaobserwowany związek jest podobny jak dla pozostałych testów, a maksymalna różnica między dziećmi pochodzącymi z rodzin, w których najwyższym



## CZY SZKOŁA MA ZNACZENIE?

poziomem wykształcenia wśród rodziców jest niepełne podstawowe, podstawowe lub zasadnicze zawodowe a tymi z rodzin, w których co najmniej jeden z rodziców ma co najmniej wykształcenie wyższe magisterskie, wynosi 6,52 punktu.

Model kontrolny (4) pozwala zredukować niewyjaśnioną wariację o ok. 44%. Znow największy odsetek redukcji niewyjaśnionej wariacji zaobserwowano na poziomie klas, jednakże wariacja efektów losowych na każdym poziomie analizy pozostaje istotnie statystycznie większa od zera. Oznacza to, że mimo kontroli zmiennych uwzględnionych w modelu (4), podział na szkoły i klasy nadal wyjaśnia zmienność wyników testów umiejętności matematycznych.

### 3.6.4. Dyskusja wyników

Przedstawione powyżej analizy miały na celu wypracowanie modeli kontrolnych dla analiz uwarunkowań efektywności nauczania w zakresie trzech obszarów osiągnięć szkolnych: umiejętności czytania, świadomości językowej i umiejętności matematycznych. Pokazały one jakie zmienne powinny być wzięte pod uwagę podczas analizowania znaczenia innych czynników mogących mieć wpływ na osiągnięcia szkolne, by uniknąć skupiania się na korelacjach pozornych. Potwierdziły także wyniki licznych badań pokazujących, że wyniki testów osiągnięć są uwarunkowane w pewnej mierze cechami uczniów i ich rodzin, na które szkoła nie ma większego wpływu.

Jednym z ważniejszych i niebędących zaskoczeniem korelatów osiągnięć jest poziom inteligencji uczniów. Wynik testu Ravena jest istotnie pozytywnie związany z wynikami wszystkich trzech testów osiągnięć, z tym że związek z poziomem umiejętności matematycznych jest trochę silniejszy niż z umiejętnością czytania i świadomością językową. Istnienie silniejszego związku inteligencji z wynikami z przedmiotów matematycznych niż z przedmiotów humanistycznych potwierdzają również inne badania (Ferrer i McArdle, 2004; Heaven i Ciarrochi, 2012; Teo i in., 1996), także te nieliczne prowadzone na gruncie polskim (Murawska, 1990; Wieczorkowska i Siarkiewicz, 2007).

Płeć uczniów ma znaczenie tylko dla testów językowych, w których to dziewczynki wypadły trochę lepiej, co najpewniej związane jest z różnicami w zakresie rozwoju mowy, więc także kompetencji językowych (por. wyniki innych badań: Konarzewski, 2007). Dla umiejętności matematycznych nie zaobserwowano istotnych różnic pomiędzy dziewczętami i chłopcami. Wyniki innych polskich badań prowadzonych z udziałem dzieci rozpoczynających naukę szkolną także pokazują brak istotnych różnic międzypłciowych dla umiejętności matematycznych, w odróżnieniu od umiejętności językowych, w których delikatną przewagę mają dziewczynki (Karwowski i Dziedziewicz, 2012, s. 123–124). Pokazuje to, że powszechne przekonanie, iż dziewczynki są mniej uzdolnione od chłopców w zakresie umiejętności matematycznych nie znajduje potwierdzenia – przynajmniej na początku edukacji formalnej.

Wiek uczniów również jest istotnie związany z poziomem osiągnięć. Analiza z wytręceniem efektów dla opóźnionego lub przyspieszonego toku nauki pokazała, że starsi uczniowie mają większą szansę uzyskania wyższych wyników w testach. A uczniowie z głównej kohorty wiekowej mogą się różnić ze względu na wiek prawie o rok (tacy uczniowie średnio różnią się będą wynikami w testach o ok. jedną piątą odchylenia standardowego). Zależność ta jest istotna dla wszystkich trzech analizowanych obszarów osiągnięć. Efekt ten nie jest duży, choć silniejszy niż w przypadku zależności obserwowanej na sprawdzianie w szóstej klasie lub egzaminie gimnazjalnym<sup>3</sup> (Dolata i Pokropek, 2012). Dowodzi to tego, że znaczenie wieku biologicznego dla osiągnięć szkolnych jest większe dla młodszych uczniów i potrzeba świadomości tego faktu nie tylko w analizach i badaniach, ale także diagnozie indywidualnej i ewaluacji pracy szkół i nauczycieli, a może przede wszystkim w zarządzaniu oświatą na poziomie lokalnym i globalnym.

Uczniowie o opóźnionym lub przyspieszonym toku nauki także różnią się średnim poziomem osiągnięć od uczniów z głównej kohorty wiekowej. Zależność ta nie jest jednak związana z ich wiekiem (musielibyśmy się bowiem spodziewać wtedy odwrotnego kierunku zależności), ale ze specyficznymi powodami przynależności do którejś z tych grup. Opóźniony tok kształcenia jest związany przede wszystkim z trudnościami w nauce lub przyczynami ściśle z tym powiązаныmi, co powoduje, że uczniowie ci średnio uzyskują wyniki w testach osiągnięć o jedno odchylenie standardowe niższe niż przeciętnie uczniowie z głównej kohorty wiekowej. Natomiast uczniowie wcześniej rozpoczynający naukę są średnio uczniami o wyższym poziomie ogólnych zdolności i gotowości do nauki, których rodzice w szczególności interesują się ich edukacją i w związku z tym, co do zasady, lepiej radzą sobie w zakresie osiągnięć szkolnych niż przeciętnie uczniowie z głównej kohorty wiekowej. Efekt ten okazał się nieistotny tylko dla umiejętności czytania.

Status społeczno-ekonomiczny rodziny pochodzenia ucznia jest pozytywnie związany z wynikami testów osiągnięć. Ma on większe znaczenie dla umiejętności czytania i świadomości językowej, dla których to wszystkie wskaźniki statusu (HISEI, wykształcenie rodziców, indeks zasobów materialnych rodziny dziecka) okazały się miarami dopełniającymi, z których każda wyjaśnia część zmienności wyników testów. Dla poziomu umiejętności matematycznych uczniów wskaźnik HISEI nie był istotnie związany z wynikami testu, co pozwala przypuszczać, że umiejętności te są w mniejszym stopniu uzależnione od statusu rodziny, a większe znaczenie odgrywa inteligencja

<sup>3</sup> W modelu kontrolnym efekt wieku jest zaniżony przez łączne uwzględnienie wieku i inteligencji, nie można więc tych oszacowań traktować jako wiążących. Z pewnością jednak siła efektu po I etapie edukacyjnym jest większa niż po II etapie.

## CZY SZKOŁA MA ZNACZENIE?

uczni. Problematyka związku cech rodziny pochodzenia uczniów z wynikami osiągnięć została bardziej wnikliwie zanalizowana i omówiona w oddzielnym rozdziale.

Współczynniki korelacji wewnątrzgrupowej wyznaczone dla każdego testu na podstawie modeli pustych pokazały istnienie znaczącego zróżnicowania osiągnięć między szkołami oraz na poziomie klas. Pogrupowanie uczniów na szkoły i klasy, rozpatrywane łącznie, wyjaśnia 8–10% zmienności wyników. Udział wariancji na poziomie szkół i klas jest w miarę równomierny, z pewną przewagą zróżnicowania międzyszkolnego dla umiejętności czytania. Największe zróżnicowanie obserwuje się na poziomie uczniów, czyli wewnątrz klas. Problematyka zróżnicowania klas i szkół podstawowych ze względu na wyniki nauczania została szczegółowo przedstawiona w następnym rozdziale. Warto jednak już teraz zwrócić uwagę na to, że uwzględnienie w modelach zmiennych wyjaśniających z poziomu uczniów (jako efektów stałych) pozwoliło zredukować znaczną część wariancji na poziomie klas. W przypadku testu umiejętności czytania i testu świadomości językowej spowodowało to, że wariancja efektów losowych na poziomie klas okazała się nieistotnie statystycznie różna od zera. Oznacza to, że nie możemy stwierdzić, że podział uczniów na klasy wyjaśnia jakąś część zmienności wyników testów językowych. Dla testu umiejętności matematycznych wariancja ta jest istotnie różna od zera, jednak większe znaczenie dla wyjaśniania zmienności wyników ma podział uczniów na szkoły. Zróżnicowanie efektów na poziomie klas, zaobserwowane w modelu pustym, może być w dużej mierze związane z procesem segregacji uczniów do klas, można je bowiem w większości wyjaśnić cechami indywidualnymi uczniów, bez uwzględnienia cech z poziomu klasy.

### 3.7. Podsumowanie

Badania nad efektywnością kształcenia są przykładem badań, w których mamy do czynienia z hierarchiczną strukturą danych. Celem tego typu badań jest zbadanie zależności pomiędzy różnymi poziomami tej hierarchii, a w szczególności określenie skali międzyoddziałościowego i międzyszkolnego zróżnicowania wyników nauczania i wskazanie czynników szkolnych odpowiedzialnych za to zróżnicowanie. Hierarchiczna struktura danych wymaga jednakże odpowiednich metod analizy, które pozwalają uwzględnić ich wielopoziomowy charakter oraz umożliwia swobodne modelowanie zależności pomiędzy różnymi poziomami analizy. Metody wykorzystane do analizy danych na potrzeby niniejszego raportu, tzn. hierarchiczne modele liniowe, spełniają te wymogi i umożliwiają testowanie różnorodnych hipotez dotyczących zależności pomiędzy cechami środowiska szkolnego a wynikami uczniów.

Hipotezy o wpływie szkoły na efekty nauczania wymagają kontrolowania wielu czynników, które mogą wpływać na osiągnięte przez uczniów wyniki, ale nie zależą od szkoły. Kontrola tych

czynników umożliwia oszacowanie wielkości tej części zmienności wyników uczniów, która może być przypisana szkołom czy oddziałom. Pozwala ona także na większe zaufanie do odnalezionych zależności pomiędzy cechami środowiska szkolnego a wynikami uczniów, zapobiegając w dużym stopniu odkrywaniu korelacji pozornych, czyli tych związków między dwiema zmiennymi, które dają się wyjaśnić wpływem trzeciej zmiennej, nieuwzględnionej w modelu.

Analizy wyników pierwszej fazy badania SUEK przedstawione w tym rozdziale pokazały, jak duże znaczenie dla osiągnięć szkolnych po I etapie kształcenia mają charakterystyki indywidualne uczniów oraz środowisko rodzinne, z którego pochodzą. Cechy takie, jak inteligencja i powiązane z nią ogólne zdolności poznawcze, płeć, wiek czy status społeczno-ekonomiczny rodziny pochodzenia mogą być nazwane potencjałem ucznia, z którym przychodzi on do szkoły. Zadaniem szkół i nauczycieli jest natomiast jak najlepsze wykorzystanie tego potencjału, by uczeń mógł poczynić największy możliwy dla niego postęp. Publiczne szkoły rejonowe nie mają wpływu na to, jacy uczniowie do nich przyjdą. Jedne, choćby ze względu na swoje położenie, pracują z uczniami z trudnych środowisk społecznych, inne głównie z dziećmi inteligentów (skala tego zróżnicowania została pokazana w rozdziale poświęconym nierównościom społecznym). Trudno wymagać zatem, by szkoły różniące się potencjałem uczniów na wejściu, osiągały podobne wyniki nauczania. Co więcej, zmiany struktury społecznej i migracje przyczyniają się do tego, że także te same szkoły co roku mogą przyjmować uczniów o różnym potencjale. Tak więc ewaluacja pracy szkół oparta tylko na wynikach testów osiągnięć, nie pozwala poprawnie porównywać efektywności nauczania w poszczególnych szkołach, jak i w tej samej placówce w różnych latach.

### Literatura cytowana

- Babbie, E. (2002). *Podstawy badań społecznych*. Tłum. Betkiewicz W., Bucholc M., Gadomski P., Haman J., Jasiewicz-Betkiewicz A., Kłoskowska-Dudzińska A., Kowalski M., Mozga-Górecka M., Warszawa: Wydawnictwo Naukowe PWN.
- Dolata, R., Pokropek, A. (2012). Czy warto urodzić się w styczniu? Wiek biologiczny a wyniki egzaminacyjne. W: B. Niemierko, M. K. Szmigel (red.), *Regionalne i lokalne diagnozy edukacyjne* (s. 157–168). Wrocław: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- Ferrer, E., 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.
- Ganzeboom, H. B. G.; Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25, 201–239.

- Ganzeboom, H. B. G.; Treiman, D. J. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. W: Hoffmeyer-Zlotnik J. H. P, Wolf C. (red.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables* (s. 159–193). New York: Kluwer Academic Press.
- Heaven, P. L., Ciarrochi, J. (2012). When IQ is not everything: Intelligence, personality and academic performance at school. *Personality and Individual Differences*, 53, 518–522.
- Karwowski, M., Dziedziewicz, D. (2012). *Test umiejętności na starcie szkolnym. Podręcznik*. Warszawa: Instytut Badań Edukacyjnych.
- Konarzewski, K. (2007). *PIRLS 2006. Jak czytają dzieci w Polsce?* Warszawa: Centralna Komisja Egzaminacyjna.
- OECD (2010). *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II)*. <http://dx.doi.org/10.1787/9789264091504-en>. Data dostępu: 22.12.2012.
- Konarski, R. (2009). *Modele równań strukturalnych. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe PWN.
- Murawska, B. (1990). *Społeczne wyznaczniki sukcesu szkolnego uczniów klas początkowych*, Niepublikowana praca doktorska, Wydział Pedagogiczny UW.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., Rasbach, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 1 (60), 23–40.
- Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and data analysis methods*. Thousand Oaks, London, New Delhi: Sage Publications.
- Scientific Software International (2012). <http://www.ssicentral.com/hlm/example6-2.html>. Data dostępu: 19.12.2012.
- Snijders, T. A. B., Bosker, R. J. (2012). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage Publications.
- Teo, A., Carlson, E., Mathieu, P. J., Egeland, B., Sroufe, L. A. (1996). A prospective longitudinal study of psychosocial predictors of achievement. *Journal of School Psychology*, 34, 285–306.
- Wieczorkowska, G., Siarkiewicz, M. (2007). Inteligencja uczniów a wynik egzaminu. Egzamin, *Biuletyn badawczy*, 13, Warszawa: Centralna Komisja Egzaminacyjna.

## **Rozdział 4**

### **Międzyszkolne i międzyoddziałowe zróżnicowanie wyników nauczania**

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

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

Zróżnicowanie międzyszkolne na danym poziomie kształcenia zdefiniowane tak jak powyżej nie pozwala rozróżnić dwóch niezależnych zjawisk, które mogą mieć wpływ na jego wielkość. Pierwszym jest zróżnicowanie uczniów na wejściu ze względu na uprzednie osiągnięcia, ich cechy indywidualne i charakterystyki ich rodzin, drugim zaś różny poziom efektywności

## CZY SZKOŁA MA ZNACZENIE?

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

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

Drugim skrajnym scenariuszem jest ten, w którym procesy te wzajemnie się znoszą – w związku z tym, że oba procesy: segregacji i zróżnicowanego pod względem jakości kształcenia mogą nawzajem hamować swój wzrostowy wpływ na zróżnicowanie międzyszkolne, scenariusz o niwelowaniu zróżnicowania może przyjmować dwa warianty. W pierwszym, mamy do czynienia z wysokim poziomem segregacji uczniów „na wejściu”, ale system szkolny pełni funkcję kompensacyjną dla uczniów słabszych, jeśli chodzi o uprzednie osiągnięcia, w związku z czym szkoły, skupiające tych uczniów osiągają średnie wyniki nie gorsze od szkół, do których uczęszczają uczniowie lepsi, co do uprzednich osiągnięć lub o korzystniejszej konfiguracji cech indywidualnych i społecznych mających wpływ na osiągnięcia szkolne. W tę stronę mogą iść interpretacje niskiego zróżnicowania międzyszkolnego w sytuacji, gdy pomiar osiągnięć podsumowuje pewien okres kształcenia. W drugim wariantcie, system szkolny wymusza swoiste mieszanie się w szkołach uczniów lepszych i słabszych ze względu na uprzednie osiągnięcia, co prowadzi do tego, że średnie dla szkół są bardzo do siebie zbliżone, pomimo to, że szkoły charakteryzują się różną efektywnością kształcenia. Z tą sytuacją możemy mieć do czynienia w przypadku, gdy osiągnięcia szkolne uczniów są analizowane na progu jakiegoś okresu kształcenia.

W przypadku, gdy dysponujemy tylko wynikami uczniów, rozplątanie wpływu tych dwóch procesów nie jest możliwe. Sytuacja zmienia się jednak, gdy mamy dostęp do informacji

o uprzednich osiągnięciach uczniów lub tych indywidualnych cech uczniów, które mogą mieć wpływ na ich osiągnięcia, ale nie są zależne od szkoły. Kontrolując je, jesteśmy w stanie „wytrącić” efekt związany z nielosowym przypisaniem uczniów o różnych charakterystykach do klas i uzyskać informację o efektywności nauczania w danej szkole. Miara zróżnicowania międzyszkolnego w tej sytuacji pozwala określić, na ile szkoły różnią się co do średniej efektywności nauczania.

Jak widać miara zróżnicowania międzyszkolnego kryje w sobie bardzo dużo informacji o tym, co dzieje się w danym systemie kształcenia. W związku z tym jest dobrym wskaźnikiem charakteryzującym system oświaty na poziomie ogólnym. Pomimo uproszczonego obrazu rzeczywistości, w którym zakłada się, że uczniowie są bezpośrednio pogrupowani w szkołach, umożliwia ona interesujące międzynarodowe porównania, dzięki temu właśnie, że poziom szkoły jest jasno interpretowalny praktycznie we wszystkich krajach, w przeciwieństwie do relatywnie złożonych i niekonięcznie przystających do siebie sposobów organizacji nauczania w samych szkołach. Opis analiz różnych aspektów zróżnicowania międzyszkolnego na poziomie międzynarodowym zawarty jest w sprawozdaniach z badania PISA (OECD, 2009).

W polskim systemie szkolnym regularną analizę i monitoring zróżnicowania międzyszkolnego dla okresu jednolitego kształcenia umożliwia wprowadzony w 2002 roku system egzaminów zewnętrznych. Ogranicza się on jednak do dwóch pomiarów: pod koniec szkoły podstawowej (sprawdzian) oraz na zakończenie gimnazjum. Okazuje się, że analiza dynamiki zróżnicowania międzyszkolnego na podstawie danych egzaminacyjnych prowadzi do niezwykłych wniosków (Dolata, Jasińska, Modzelewski, 2012).

W niniejszym rozdziale zaprezentowane zostały wyniki analizy zróżnicowania dla szkół podstawowych po pierwszym etapie kształcenia (tj. po trzeciej klasie szkoły podstawowej) na podstawie badań SUEK. Wyniki te zostały skonfrontowane z analizami zróżnicowania pod koniec szkoły podstawowej oraz na zakończenie gimnazjum. Pozwala to na szersze spojrzenie na dynamikę różnicowania się szkół w Polsce.

Oszacowanie zróżnicowania międzyszkolnego zostało wyliczone z użyciem dwóch rodzajów hierarchicznych modeli liniowych: analizy wariancji z efektami losowymi oraz analizy kowariancji z efektami losowymi. Modele te mają szczególną zaletę w świetle przeprowadzanych analiz, są bowiem dosyć odporne na artefaktualne zwiększanie wariancji międzygrupowej, wynikającej z wahań w wielkości poszczególnych grup (szkół). Dzieje się tak dlatego, że średnie dla szkół są ściągane do średniej populacyjnej tym bardziej, im mniej uczniów stanowiło podstawę do jej wyliczenia. Wydaje się to rozsądnym założeniem – średnia otrzymana z wyników dla małej liczby uczniów jest mniej rzetelna niż ta, do której obliczenia wzięto wyniki kilkudziesięciu uczniów.



#### 4.1. Zróżnicowanie międzyszkolne wyników nauczania po pierwszym etapie kształcenia na podstawie wyników badania SUEK

Zacznijmy od przyjrzenia się poziomowi zróżnicowania międzyszkolnego ze względu na osiągnięcia szkolne po pierwszym etapie kształcenia, czyli na zakończenie nauczania zintegrowanego. Do analiz wykorzystano wyniki testów osiągnięć szkolnych z badania SUEK (opisane szczegółowo w rozdziale 2). Analizy wykonano za pomocą oprogramowania HLM6.06. Posłużono się metodą estymacji *full maximum likelihood* oraz wykorzystano warunkowe wagi próbkowania (patrz rozdział 1). Wskaźnik zróżnicowania międzyszkolnego uzyskano, dzieląc wariancję efektów losowych na poziomie szkół przez wariancję całkowitą, a wynik wyrażono w procentach. W tabeli 4.1 przedstawiono oszacowania efektów losowych uzyskane za pomocą dwupoziomowych modeli pustych z losową stałą. Wariancja efektów szkół dla każdego testu osiągnięć jest statystycznie istotnie różna od zera.

**Tabela 4.1. Zróżnicowanie międzyszkolne wyników nauczania po I etapie kształcenia – oszacowania na podstawie dwupoziomowych modeli z losową stałą (analiza wariacji z efektami losowymi)**

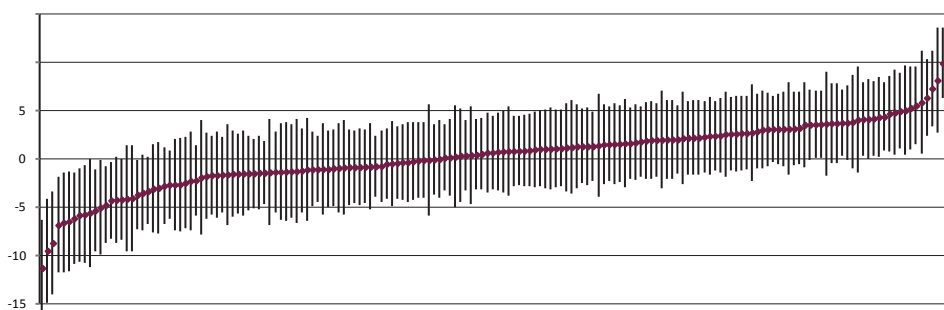
	Test umiejętności czytania	Test świadomości językowej	Test umiejętności matematycznych
<b>oszacowanie efektów losowych</b>			
wariancja efektów szkół	<b>15,13</b>	<b>20,16</b>	<b>17,71</b>
wariancja na poziomie uczniów	<b>209,32</b>	<b>206,63</b>	<b>208,72</b>
wskaźnik zróżnicowania międzyszkolnego	6,74%	8,89%	7,82%
liczba uczniów	4725	4725	4784
liczba szkół	172	172	172

Źródło: opracowanie własne

Analizy te pokazują, że podział uczniów na szkoły wyjaśnia istotną statystycznie część zróżnicowania wyników nauczania po pierwszym etapie kształcenia. Zróżnicowanie międzyszkolne jest na poziomie ok. 7–9%. Nie jest to dużo, jednak w takim stopniu wyniki, jakie uzyskują uczniowie, są związane z tym, do których szkół chodzą. By lepiej zrozumieć, co oznacza taka wartość wskaźnika zróżnicowania międzyszkolnego, warto przyrzeć się graficznej interpretacji uzyskanych wyników. Na rysunku 4.1 przedstawiono oszacowania średnich wyników szkół dla przykładowego testu – testu umiejętności czytania – wraz z 95% przedziałem ufności dla oszacowanej średniej. Wyniki szkół uporządkowano rosnąco ze względu na uzyskany średni wynik. Na osi pionowej przedstawiono skalę umiejętności czytania, w której średnia wyników w populacji uczniów wynosi 0,

a odchylenie standardowe wyników indywidualnych 15. Ograniczono się do prezentacji wykresu tylko dla tego testu, gdyż dla pozostałych wyglądają bardzo podobnie.

W hierarchicznych modelach liniowych efekty losowe na poziomie grup (w tym przypadku szkół) nie są, tak jak inne parametry modelu, szacowane podczas procesu estymacji. Estymowana jest ich wariancja ( $\tau^2$ ) i ewentualnie kowariancja z innymi efektami losowymi. Jeśli jednak interesują nas oszacowania efektów grupowych, możemy wyznaczyć bayesowskie predykcje *posteriori* (*empirical Bayes estimation*). Metoda ta pozwala wyznaczyć wartości tych efektów na podstawie dwóch informacji: danych o jednostkach przynależących do konkretnej grupy (wynikach uczniów z danej szkoły) i założenia modelu o normalnym rozkładzie efektów losowych o średniej 0 i wariancji równej  $\tau^2$  (Snijders i Bosker, 2012, s. 62–67). W ten sposób wyznaczono średnie dla szkół – jako oszacowania efektów losowych uzyskane opisaną powyżej metodą. Błędy standardowe zostały wyznaczone jako pierwiastek z wariancji rozkładu *posteriori*.



**Rysunek 4.1. Oszacowania średnich wyników szkół wraz z przedziałami ufności dla testu umiejętności czytania (średni wynik w populacji = 0, odchylenie standardowe = 15).**

Źródło: opracowanie własne

Wykres ten pozwala nam lepiej zobaczyć, że zdecydowana większość szkół nie różni się znacząco wynikami nauczania. Tylko na skrajach możemy dostrzec szkoły, których średnie wyniki są istotnie różne od średnich wyników w populacji, a tym samym różne od wyników szkół z drugiego krańca. Szkół takich jest ok. 20%. Jednak mimo że większość szkół osiąga średnio podobne wyniki, to jeśli spojrzymy na wielkość różnicy między szkołami o najniższych i najwyższych wynikach, to zobaczymy, że wynosi ona trochę ponad 20 punktów, czyli jeden i jedna trzecia odchylenia standardowego wyników indywidualnych! Wyniki te pokazują, iż mimo że patrząc całościowo na poziom zróżnicowania międzyszkolnego na tym etapie kształcenia, można mówić o w miarę jednorodnym systemie nauczania, to nie oznacza to, że w przypadku każdej pary szkół różnice te są niewielkie. Interpretacje zależą od przyjętej perspektywy.

## CZY SZKOŁA MA ZNACZENIE?

Zanim jednak postawiony zostanie wniosek, że to różna efektywność nauczania odpowiada za zaobserwowany poziom zróżnicowania międzyszkolnego, należy wytrącić wpływ tych czynników, które od szkoły nie zależą, czyli poziom ogólnych zdolności uczniów, ich cechy indywidualne i cechy środowiska rodzinnego, z którego pochodzą. Być może bowiem zróżnicowanie międzyszkolne wyników nauczania po pierwszym etapie edukacyjnym jest tylko rezultatem międzyszkolnej segregacji uczniów na wejściu. W tym celu policzono wskaźniki zróżnicowania międzyszkolnego na podstawie modeli, w których kontrolowano takie zmienne jak: inteligencja płynna uczniów, ich płeć, wiek, status społeczno-ekonomiczny wskazany wykształceniem rodziców, indeksem HISEI i indeksem zasobów materialnych rodziny dziecka. Były to dwupoziomowe modele z losową stałą i efektami stałymi dla zmiennych kontrolnych. Modele te są analogiczne do wypracowanych trzypoziomowych modeli kontrolnych opisanych w rozdziale 3, z tą jedyną różnicą, że uwzględniają tylko poziom szkoły i uczniów. W tabeli 4.2 przedstawiono interesującą nas w tym miejscu część wyników, czyli oszacowania efektów losowych. Wariancja efektów szkół dla każdego testu osiągnięć jest statystycznie istotnie różna od zera.

**Tabela 4.2. Zróżnicowanie międzyszkolne efektywności nauczania po I etapie kształcenia – oszacowania na podstawie dwupoziomowego modelu z losową stałą i efektami stałymi dla zmiennych kontrolnych (analiza kowariancji z efektami losowymi)**

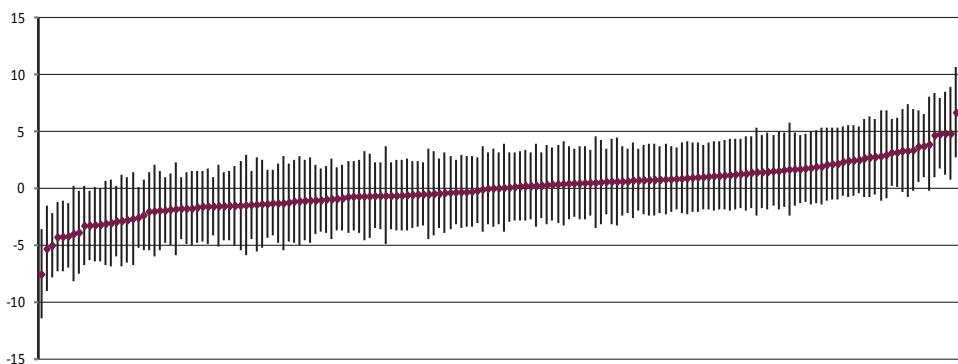
	Test umiejętności czytania	Test świadomości językowej	Test umiejętności matematycznych
<b>oszacowanie efektów losowych</b>			
wariancja efektów szkół	<b>7,89</b>	<b>7,48</b>	<b>8,95</b>
wariancja na poziomie uczniów	<b>142,66</b>	<b>122,44</b>	<b>116,83</b>
wskaźnik zróżnicowania międzyszkolnego	5,24%	5,76%	7,11%
liczba uczniów	4725	4725	4784
liczba szkół	172	172	172

Źródło: opracowanie własne

Wyniki te pokazują, że jeśli kontrolujemy cechy biologiczno-społeczne uczniów, to zróżnicowanie międzyszkolne osiągnięć szkolnych jest niższe. Tak więc, zgodnie z przewidywaniami, za część zróżnicowania międzyszkolnego wyników nauczania odpowiada zróżnicowanie międzyszkolne cech uczniów na wejściu. Uzyskane wyniki informują o tym, na ile szkoły różniłyby się między sobą ze względu na wyniki nauczania, jeśli wszystkie pracowałyby z uczniami o takich samych cechach biopsychicznych i społecznych. Można je więc interpretować w kategoriach zróżnicowania w zakresie wkładu szkoły w osiągnięcia uczniów, czy inaczej mówiąc

w zakresie efektywności nauczania. Tego typu wskaźniki nazywamy kontekstowymi wskaźnikami wyników nauczania. Pokazują one, że nawet przy kontroli mierzonych w tym badaniu cech uczniów, podział na szkoły wyjaśnia istotną statystycznie część zróżnicowania osiągnięć szkolnych uczniów.

Dodanie do modelu zmiennych kontrolnych, zredukowało jednak nie tylko wariancję efektów na poziomie szkół, ale w dużej mierze także wariancję na poziomie uczniów. Oznacza to, że sporą część zróżnicowania indywidualnego osiągnięć szkolnych możemy przypisać cechom biopsychologicznym uczniów i środowisku rodzinnemu, z którego pochodzą. Redukcja wariancji na poziomie uczniów, poza tą oczywistą interpretacją, prowadzi jednak do jeszcze jednej konsekwencji – zmniejszenia błędu standardowego oszacowania efektów grupowych. Co to oznacza? Spójrzmy na rysunek 4.2, na którym przedstawiono oszacowania kontekstowych wskaźników wyników nauczania dla testu umiejętności czytania wraz z 95% przedziałem ufności. Wyniki te w pewnym uproszczeniu mówią o tym, o ile średnio punktów został przesunięty rozkład wyników indywidualnych z zakresu umiejętności czytania w danej szkole w stosunku do rozkładu przewidywanego na podstawie cech uczniów uwzględnionych w modelu i zależności między tymi zmiennymi a zmienną zależną w populacji. Skala na, której przedstawiono wyniki, jest skalą indywidualnych wyników w teście czytania, o odchyleniu standardowym w populacji równym 15. Średnia w populacji tych efektów jest równa 0. Inaczej mówiąc, wyniki zobrazowane na wykresie, pokazują, ile średnio punktów w teście czytania dana szkoła „dodała” (lub „odjęła”) uczniom w stosunku do wyników przewidywanych na podstawie ich cech niezależnych od pracy szkoły. Efekty te zostały oszacowane w sposób analogiczny do omawianego dla modelu pustego z tym zastrzeżeniem, że zostały one policzone na podstawie modelu uwzględniającego zmienne wyjaśniające.



**Rysunek 4.2. Oszacowania edukacyjnej wartości dodanej szkół wraz z przedziałami ufności dla testu umiejętności czytania.**

*Źródło: opracowanie własne*

## CZY SZKOŁA MA ZNACZENIE?

Jak widzimy, na podstawie porównania wykresów znajdujących się na rysunkach 4.1 i 4.2, spadek wariancji międzyszkolnej uwidacznia się w bardziej płaskiej linii, która rysuje się nam z kropek oznaczających średnie dla wyznaczonych efektów. Widzimy jednak także, że przedziały ufności są węższe, dzięki czemu nadal ok. 10% szkół ma – wyznaczoną w ten sposób – efektywność nauczania różną od przeciętnej, a także dla trochę ponad 10% szkół możemy powiedzieć, że istotnie różnią się one efektywnością nauczania. I nadal różnica pomiędzy najlepszą i najslabszą szkołą jest niemała, bo w granicach jednego odchylenia standardowego. Dla testu matematycznego jest podobna, a dla testu świadomości językowej troszkę mniejsza, bo wynosi ok. dwóch trzecich odchylenia standardowego.

Powyższe analizy pokazują, że zróżnicowanie międzyszkolne wyników nauczania w zakresie umiejętności czytania, wiadomości językowej i umiejętności matematycznych po pierwszym etapie kształcenia jest względnie nieduże. Jest ono jeszcze mniejsze, jeśli policzymy je, kontrolując cechy uczniów i rodzin, z których pochodzą. Oznacza to, że w pewnym stopniu zróżnicowanie międzyszkolne wyników na poziomie pierwszego etapu edukacyjnego jest rezultatem zróżnicowania międzyszkolnego cech uczniów na wejściu. Jednak nawet po wytrąceniu tego zróżnicowania, szkoły różnią się w pewnym stopniu efektywnością nauczania, co prowadzi do dość sporych różnic, jeśli porównamy szkoły o skrajnych wynikach. Można więc powiedzieć, że patrząc całościowo na poziom zróżnicowania międzyszkolnego na tym etapie kształcenia mamy w miarę jednorodny system nauczania, co jednak nie oznacza, że nie odnajdujemy w nim szkół różniących się wynikami.

### **4.2. Zróżnicowanie międzyszkolne po pierwszym etapie kształcenia na tle drugiego i trzeciego etapu edukacji**

Właściwe zinterpretowanie faktu, że zróżnicowanie międzyszkolne wyników nauczania po I etapie edukacyjnym jest na poziomie 7–9%, wymaga ukazania tego zjawiska w szerszym kontekście. Jedną z ważnych perspektyw jest dynamiczne ujęcie problemu. O dynamice procesu możemy myśleć na dwa sposoby: możemy śledzić proces w kolejnych latach na tym samym etapie edukacji, a także możemy analizować zróżnicowanie międzyszkolne dla tej samej kohorty uczniów na kolejnych etapach kształcenia. Pierwsze ujęcie pozwala nam zbadać zmiany zachodzące w systemie edukacji, monitorować zaobserwowane procesy i na podstawie tych danych ocenić skutki podejmowanych działań na poziomie globalnym czy lokalnym. Dobrym przykładem takich badań, które przyczyniły się do odkrycia niezwykle znaczącego dla polityki oświatowej procesu, jest analiza trendów w zakresie różnicowania się gimnazjów w kolejnych latach przeprowadzona na podstawie danych z systemu egzaminów zewnętrznych (Dolata, 2009; Dolata i in., 2012). Drugie ujęcie pozwala w sposób bardziej całościowy

scharakteryzować system oświaty pod kątem jego jednorodności. Dzięki temu możemy porównać poziom zróżnicowania międzyszkolnego na kolejnych etapach kształcenia i przekonać się, czy założenie o jednolitości systemu edukacji na poziomie szkół podstawowych i gimnazjów znajduje odzwierciedlenie w rzeczywistości, a jeśli nie, to czy obserwujemy w tym zakresie jakieś zmiany i w którym momencie są one najbardziej znaczące.

Najważniejsze w kontekście tych analiz jest skupienie się na trzech punktach w czasie: pomiarze osiągnięć po I, II i III etapie edukacyjnym. Okres ten obejmuje obowiązkową naukę szkolną w szkołach z założenia jednolitych, do których uczniowie powinni być przyjmowani bez selekcji na wejściu. W celu znalezienia odpowiedzi na postawione wyżej pytania najlepiej byłoby wykorzystać dane podłużne obejmujące wspomniany okres (wykluczilibyśmy wtedy możliwość popełnienia błędów w oszacowaniach związanych z porównywaniem różnych kohort uczniów, które mogłyby mieć trochę inne cechy). Dodatkowo posiadanie informacji charakteryzujących uczniów i ich rodziny, pozwoliłoby na pogłębienie analiz przez kontrolowanie zmiennych mogących mieć wpływ na wielkość zróżnicowania. Nie dysponujemy w Polsce jeszcze takimi danymi. Uczniowie uczestniczący w badaniu SUEK dopiero w 2017 roku planowo zakończą III etap edukacji (naukę w gimnazjum). Nie oznacza to jednak, że do tego czasu pozostaje tylko siedzieć i czekać. Dysponujemy miarą osiągnięć szkolnych po II i III etapie nauczania – są to wyniki egzaminów zewnętrznych (sprawdzianu i egzaminu gimnazjalnego). Nie mogąc przeprowadzić analiz na tej samej kohorcie uczniów, zdecydowaliśmy się natomiast wybrać kohortę najbliższą tej, do której będziemy przyrównywać wyniki z badania SUEK, czyli uczniów, którzy przystępowali do egzaminu gimnazjalnego w 2012 roku, a do sprawdzianu w szóstej klasie szkoły podstawowej trzy lata wcześniej – czyli w 2009 roku.

### 4.2.1. Dane i metoda analizy

W analizach, oprócz danych z badania SUEK, wykorzystano dane ogólnokrajowe ze sprawdzianu w szóstej klasie szkoły podstawowej z 2009 roku oraz z egzaminu gimnazjalnego z 2012 roku. Wyniki egzaminacyjne znormalizowano metodą ekwicytylową, a następnie z analiz wykluczono dane pochodzące ze szkół, w których na poziomie ostatniej klasy uczyło się mniej niż 11 uczniów. Z danych dotyczących egzaminu gimnazjalnego wykluczono także szkoły dla dorosłych, szkoły specjalne oraz przyszpitalne. Kryterium tego nie można było wykorzystać dla danych ze sprawdzianu, ze względu na brak odpowiednich zmiennych w tych bazach. Jednak wspomniane szkoły są przede wszystkim szkołami małymi, dlatego zastosowanie kryterium wielkości, pozwoliło wykluczyć z analiz zdecydowaną ich większość. Wykluczenia te są podyktowane potrzebą ograniczenia definicji interesującej nas w analizach populacji do tej przyjętej w badaniu SUEK (szczegóły są opisane w rozdziale 1), w celu zapewnienia porównywalności wyników.

## CZY SZKOŁA MA ZNACZENIE?

W kontekście porównywalności wyników należy zwrócić jeszcze uwagę na trzy problemy. Pierwszy wynika z faktu, że dane podsumowujące I etap nauczania są próbką danych z populacji, a sposób doboru próby zakładał przebadanie maksymalnie dwóch klas w każdej szkole. Oznacza to, że w 72 na 172 szkoły próba założona nie objęła wszystkich uczniów w szkole. Dane egzaminacyjne poza tym, że obejmują całą populację szkół, zapewniają informację o wszystkich uczniach uczących się w danej szkole. Dane te różnią się więc przede wszystkim zrównoważeniem ze względu na liczebność uczniów w ramach szkół. Może to prowadzić do pewnych różnic w zakresie dekompozycji wariancji, jednak wydaje się, że stosowane w niniejszych analizach metody (dwupoziomowe modele puste z losową stałą) dają najlepsze oszacowania w takich sytuacjach.

Drugi problem dotyczy kwestii związanych z realizacją próby. Jak omówiono w rozdziale poświęconym realizacji badania, część szkół z próby głównej, które nie wzięły udziału w badaniu, zostało zastąpionych szkołami z próby zapasowej. Mimo iż analiza reprezentatywności próby pokazała, że rozkład indywidualnych wyników osiągnięć szkolnych uczniów dla próby zrealizowanej jest zbliżony do rozkładu tych wyników w populacji, to zaobserwowano pewne odchylenie od reprezentatywności próby ze względu na zróżnicowanie międzyszkolne (nie-wielka redukcja wariancji międzyszkolnej).

Trzeci problem związany jest z właściwościami psychometrycznymi testów. Oszacowania wariancji zależą bowiem od rzetelności pomiaru. Niestety, wykorzystane w analizach testy nie mają identycznej rzetelności. Porównanie rzetelności testów z badania SUEK z rzetelnością egzaminów zewnętrznych nie jest jednak bezproblemowe. Testy te zostały bowiem opracowane według różnych technologii. Dla testów z badania SUEK obliczenie dobrze znanego z klasycznej teorii testów współczynnika rzetelności (alfa Cronbacha) jest zasadniczo niemożliwe (patrz rozdział 2). Bez zastosowania analogicznej do przyjętej w badaniu SUEK metody skalowania wyników egzaminacyjnych, nie możemy natomiast wyznaczyć dla egzaminów zewnętrznych porównywalnych współczynników rzetelności, do tych obliczonych dla testów z badania SUEK. Ponadto należy zdawać sobie sprawę z tego, że przyjęcie różnych metod wyznaczania wyników uczniów, może także prowadzić do trochę odmiennych oszacowań wariancji. Znaczenie ma chociażby uwzględnienie niepewności pomiaru przez posługiwanie się wartościami PV (patrz rozdział 2)<sup>1</sup>. Pamiętanie o tym podczas interpretacji wyników jest waż-

<sup>1</sup> Wydaje się także, że uwzględnienie w wyznaczaniu wartości PV złożonego schematu doboru próby biorącego pod uwagę to, że uczniowie są „zagnieżdżeni” w klasach i szkołach, może mieć pewne znaczenie dla oszacowania wariancji międzygrupowej i na poziomie indywidualnym (może w pewnym stopniu wpłynąć na wzrost wariancji międzygrupowej i spadek wariancji na poziomie indywidualnym). Obecnie jednak, według naszej wiedzy, nie istnieje oprogramowanie, które umożliwiłoby uwzględnienie w sposób poprawny schematu doboru próby podczas szacowania wartości PV.

ne o tyle, że pokazuje, że nie możemy bezwzględnie porównywać oszacowanych wartości. Należy mieć na uwadze to, że są one pewnym przybliżeniem rzeczywistości, choć pozwalają naszkicować w sposób nadal wartościowy obraz interesującego nas zjawiska.

Dla każdego testu osiągnięć dekompozycję wariancji wykonano licząc, dwupoziomowe modele puste z losową stałą. Analizy wykonano metodą *full maximum likelihood* za pomocą oprogramowania HLM6.06. W analizach danych z badania SUEK wykorzystano warunkowe wagi próbkowania. Wskaźnik zróżnicowania międzyszkolnego uzyskano, dzieląc wariancję efektów losowych na poziomie szkół przez wariancję całkowitą, a wynik wyrażono w procentach.

### 4.2.2. Zróżnicowanie międzyszkolne na kolejnych etapach edukacji

W tabeli 4.3 przedstawiono wyniki analiz dla pomiarów podsumowujących kolejne etapy kształcenia. Dane te pokazują wyraźną tendencję wzrostową w zakresie poziomu zróżnicowania międzyszkolnego dla kolejnych etapów edukacji. Po I etapie kształcenia zróżnicowanie międzyszkolne kształtuje się na poziomie ok. 7–8%. Na zakończenie szkoły podstawowej wynosi ok. 11%, a po zakończeniu III etapu edukacyjnego podział na szkoły wyjaśnia ok. 14–18% zróżnicowania wyników. Pokazuje to, że system edukacji na poziomie szkoły podstawowej jest w miarę jednolity w porównaniu do III etapu nauczania. Wyniki uczniów szkół podstawowych w mniejszym stopniu niż w przypadku nauczania w gimnazjach są uwarunkowane tym, w jakiej szkole podejmują oni naukę.

**Tabela 4.3. Oszacowanie zróżnicowanie międzyszkolnego na kolejnych etapach edukacyjnych**

Pomiary osiągnięć na kolejnych etapach edukacyjnych		Wskaźnik zróżnicowania międzyszkolnego	Liczba szkół	Liczba uczniów
Badanie SUEK	test umiejętności czytania	6,74%	172	4725
	test świadomości językowej	8,89%	172	4725
	test umiejętności matematycznych	7,82%	172	4784
Sprawdzian w VI klasie szkoły podstawowej		10,93%	10041	382898
Egzamin gimnazjalny	język polski	18,13%	6171	384 891
	historia i wiedza o społeczeństwie	16,20%	6171	384 891
	matematyka	16,45%	6171	384 891
	przedmioty przyrodnicze	14,41%	6171	384 891

Źródło: opracowanie własne



## CZY SZKOŁA MA ZNACZENIE?

Przyglądając się wzrostowej tendencji dla kolejnych etapów nauczania, warto zwrócić uwagę na to, że różnica punktów procentowych w zakresie wskaźnika zróżnicowania międzyszkolnego jest między I i II etapem kształcenia jest nieduża. Większa jest natomiast między końcem szkoły podstawowej a gimnazjum, co daje podstawy do przypuszczeń, że nie mamy do czynienia z prostoliniowym wzrostem, który sugerowałby tylko zwiększanie się z wiekiem uczniów znaczenia szkoły dla wyników nauczania. Dysproporcja ta ma szansę być jeszcze większa, gdy kohorta uczniów, dla których dysponujemy pomiarem po I etapie nauczania, przystąpi do sprawdzianu i egzaminu gimnazjalnego. A to za sprawą faktu, że dla wyników sprawdzianu nie obserwuje się znaczącego trendu w zakresie interesującego nas zjawiska w kolejnych latach, co pozwala prognozować, że poziom zróżnicowania międzyszkolnego na tym etapie edukacji utrzyma się na podobnym poziomie (Dolata, 2011). W przypadku gimnazjów notuje się natomiast systematyczny wzrost zróżnicowania międzyszkolnego (w szczególności w dużych miastach). Nie bezzasadne jest więc przypuszczenie, że dla kohorty reprezentowanej w badaniu SUEK zróżnicowanie gimnazjów ze względu na wyniki będzie jeszcze wyższe.

O czym to może świadczyć? Za wzrost zróżnicowania międzyszkolnego między końcem szkoły podstawowej a gimnazjum odpowiadają zatem jakieś dodatkowe czynniki niezwiązane z naturalną tendencją zwiększania się znaczenia szkoły dla osiągnięć uczniów. Mogłoby to oznaczać, że gimnazja bardziej niż szkoły podstawowe różnią się efektywnością nauczania, lub że istnieją ukryte mechanizmy selekcji i autoselekcji na progu gimnazjum, które przyczyniają się do wzrostu zróżnicowania szkół ze względu na potencjał uczniów na wejściu. Zróżnicowanie na wejściu natomiast w oczywisty sposób prowadzi do zróżnicowania na wyjściu. Nie wykluczając istnienia pierwszego mechanizmu, drugi wydaje się niezwykle prawdopodobny. Tym bardziej, że badania pokazują (Dolata, 2009), że nie można wyjaśnić zjawiska rosnącego zróżnicowania się gimnazjów ze względu na wyniki egzaminacyjne segregacją przestrzenną, czy rozwojem sektora szkół niepublicznych. Procesy selekcji i autoselekcji na progu szkoły mogą zachodzić tam, gdzie rodzice mają do wyboru przynajmniej kilka szkół – czyli przede wszystkim w dużych i średniej wielkości miastach. I właśnie w tych lokalizacjach obserwuje się znacznie większe zróżnicowanie międzyszkolne wyników egzaminacyjnych (Dolata i in., 2012).

### 4.3. Zróżnicowanie międzyoddziałowe

W polskich realiach przez zdecydowaną większość czasu kształcenia uczniów, edukacja realizowana jest w ramach relatywnie stałych co do składu grupach – oddziałach klasowych (wyjątkiem od tej reguły często jest nauczanie języków obcych oraz etyki). Dodatkowo,

na pierwszym etapie kształcenia (tj. w klasach 1–3 szkoły podstawowej) w jednym oddziale klasowym uczy zazwyczaj tylko jeden nauczyciel. Powoduje to, że na problem zróżnicowania wyników nauczania warto spojrzeć w bardziej szczegółowy sposób – dokonując rozbicia tego zróżnicowania na trzy poziomy: szkół, oddziałów i uczniów.

Pojawia się zatem nowa miara – wskaźnik zróżnicowania międzyoddziałowego. Oznacza ona stopień, w jakim klasy różnią się od szkolnej średniej. Czysto teoretycznie moglibyśmy wyobrazić sobie system szkolny, w którym między szkołami nie ma żadnych różnic lub są one niewielkie, za to w samych szkołach mielibyśmy do czynienia z oddziałami znacząco lepszymi od średniej w populacji uczniów i klasami znacząco gorszymi. Bez miary zróżnicowania międzyoddziałowego moglibyśmy dojść do wniosku, że skoro szkoły nie różnią się co do średnich wyników, to, że system szkolny jest w miarę jednolity. Różnice mogą się jednak kryć do pewnego stopnia właśnie w zróżnicowaniu na poziomie oddziałów, które w ujęciu dwupoziomym (szkoła–uczniowie) było zawarte w ramach zróżnicowania uczniów.

Jeśli chodzi o interpretację zróżnicowania międzyoddziałowego, to jest ona zbliżona do interpretacji zróżnicowania międzyszkolnego. Podobnie jak w tamtym przypadku, należy pamiętać, że na wielkość obserwowanego zróżnicowania międzyoddziałowego mogą mieć wpływ dwa zjawiska: różny skład oddziałów pod względem uprzednich osiągnięć lub cech indywidualnych, które wpływają na osiągnięcia, oraz zróżnicowana efektywność nauczania w tych klasach. To ostatnie zjawisko jest o tyle interesujące w kontekście pierwszego etapu kształcenia, że w trakcie tego okresu nauki, zazwyczaj tylko jeden nauczyciel pracuje z uczniami. Oznacza to, że efektywność nauczania rozumiana jako różnica średniej wyników uczniów tego oddziału od średniej dla szkoły, przy kontroli cech indywidualnych uczniów, może być przypisywana sposobowi nauczania praktykowanemu przez tych nauczycieli. Otwiera się pole do wymiany doświadczeń i dobrych praktyk między nauczycielami edukacji wczesnoszkolnej, które może się przełożyć na poprawę jakości nauczania na pierwszym etapie edukacyjnym.

Z metodologicznego punktu widzenia warto dodać, że przy zastosowaniu trzypoziomowej analizy regresji z losowymi efektami, oszacowania średnich dla szkół są bardziej ściągane do średniej populacyjnej, niż przy modelu dwupoziomym. Dzieje się tak dlatego, że oszacowania średniej na poziomie szkoły są mniej rzetelne w trzypoziomym modelu, ponieważ model ten uwzględnia fakt, że w ramach szkoły uczniowie są podzieleni na oddziały, a w ramach oddziału mogą być do siebie bardziej podobni, niż miałyby to miejsce, gdyby stanowili losową próbkę uczniów w szkole (Raudenbush i Bryk, 2002). Co za tym idzie oszacowanie wariancji międzyszkolnej jest stosunkowo niższe niż dla modelu dwupoziomowego. W tabeli 4.4 przedstawiono wyniki analiz modeli trzypoziomowych. Dla każdego testu policzono dwa modele: model pusty z losową stałą uwzględniający tylko pogrupowanie uczniów w szkołach i oddziałach oraz model kontrolny, będący modelem z losową stałą i efektami stałymi z poziomu

## CZY SZKOŁA MA ZNACZENIE?

uczniów dla zmiennych kontrolujących efekt segregacji na wejściu. Modele kontrolne wytrącają zatem efekt wpływu na zróżnicowanie wyników cech ucznia, które są niezależne od szkoły, a – z drugiej strony – powiązane są z wynikami nauczania. Modele te, oraz wszystkie uwzględnione w nich zmienne, zostały szczegółowo opisane w rozdziale 3, dlatego nie będą tu ponownie omawiane. W tabeli 4.4 zaprezentowano tę część wyników, która z punktu widzenia omawianego tu problemu jest znacząca, czyli oszacowania efektów losowych na każdym z analizowanych poziomów oraz wskaźniki zróżnicowania na poziomie szkół i oddziałów. Wskaźniki te obliczono w sposób analogiczny do wcześniej omawianego – wariancję efektów na danym poziomie (szkół lub klas) podzielono przez wariancję całkowitą, a wynik wyrażono w procentach.

**Tabela 4.4. Zróżnicowanie międzyszkolne i międzyoddziałowe osiągnięć szkolnych po pierwszym etapie kształcenia na podstawie wyników badania SUEK. Oszacowania dla modelu pustego i modelu kontrolnego**

Model:	Test umiejętności czytania		Test świadomości językowej		Test umiejętności matematycznych	
	pusty	kontrolny	pusty	kontrolny	pusty	kontrolny
wariancja efektów szkół	<b>9,08</b>	<b>6,50</b>	<b>13,84</b>	<b>6,04</b>	<b>9,06</b>	<b>6,58</b>
wariancja efektów oddziałów klasowych	<b>8,97</b>	1,96	<b>8,86</b>	1,93	<b>12,7</b>	<b>4,38</b>
wariancja na poziomie ucznia	<b>205,48</b>	<b>141,71</b>	<b>202,64</b>	<b>121,73</b>	<b>203,74</b>	<b>114,52</b>
wskaźnik zróżnicowania międzyszkolnego	4,06%	4,33%	6,14%	4,66%	4,02%	5,24%
wskaźnik zróżnicowania na poziomie klas	4,01%	1,31%	3,93%	1,49%	5,63%	3,49%
liczba uczniów	4738	4738	4738	4738	4633	4633
liczba klas	300	300	300	300	300	300
liczba szkół	172	172	172	172	172	172

pogrubbym drukiem opisywane są wartości istotne na poziomie istotności  $p < 0,05$

Źródło: opracowanie własne

Wyniki te pokazują, że jeśli interesuje nas zróżnicowanie wyników nauczania bez kontroli innych zmiennych, to zarówno podział na szkoły, jak i podział na oddziały wewnątrz szkół istotnie wyjaśnia część wariancji wyników. Inaczej mówiąc, analiza ta pokazuje nam, że nie

tylko zróżnicowanie szkół ze względu na wyniki jest znaczące (co mogliśmy zaobserwować wcześniej), ale także zróżnicowanie oddziałów wewnątrz szkół ma znaczenie. Udział wariacji na poziomie szkół i oddziałów jest w miarę równomierny, z pewną przewagą zróżnicowania międzyszkolnego dla umiejętności czytania. Największe natomiast zróżnicowanie obserwuje się oczywiście na poziomie uczniów, czyli wewnątrz oddziałów.

Niezwykle interesujące jest jednak porównanie wyników tych analiz z oszacowaniami z modeli kontrolnych, w których wytrącamy wpływ na osiągnięcia szkolne tych cech, które od szkół nie zależą. Okazuje się, że jeśli kontrolujemy w modelu cechy biopsychologiczne i społeczne uczniów, to wariancja wyników na poziomie oddziałów ulega znaczącej redukcji. Dla testu umiejętności czytania i świadomości językowej staje się nieistotnie statystycznie różna od zera. Dla testu umiejętności matematycznych jest istotnie różna od zera, jednak większe znaczenie dla wyjaśniania zmienności wyników ma podział uczniów na szkoły. Wyniki te pokazują, że za zróżnicowanie osiągnięć szkolnych na poziomie klas trzecich odpowiada w dużej mierze nielosowy przydział uczniów do oddziałów. Gdyby bowiem podział uczniów na oddziały nie dokonywał się na podstawie cech uczniów związanych z ich poziomem ogólnych zdolności poznawczych, wiekiem i statusem rodziny pochodzenia, to cechy te rozłożyłyby się losowo między oddziały i dodanie do modelu zmiennych je kontrolujących, nie wpłynęłoby w tak dużym stopniu na zmiany w zakresie oszacowania wariacji efektów na poziomie oddziałów.

Wyliczone dla modeli kontrolnych wskaźniki zróżnicowania międzyszkolnego i międzyoddziałowego informują o tym, na ile szkoły i oddziały wewnątrz szkół różniłyby się między sobą ze względu na wyniki nauczania, jeśli wszystkie pracowałyby z uczniami o takich samych cechach. Okazuje się, że różnice między poszczególnymi oddziałami w ramach tych samych szkół byłyby bardzo niewielkie. Oznacza to, że w ramach tych samych szkół uczniowie z różnych oddziałów są nauczani z podobną efektywnością, w szczególności w zakresie języka polskiego. Pewne różnice międzyoddziałowe dostrzega się dla nauczania matematyki. Jest to ciekawe w szczególności z tego powodu, że na tym etapie kształcenia ci sami nauczyciele kształcą w zakresie edukacji polonistycznej i matematycznej. Ponieważ różnice w efektywności nauczania w różnych klasach w tych samych szkołach w dużej mierze można przypisać wysiłkom nauczycieli, wyniki te pokazują, że ci sami nauczyciele z porównywalną efektywnością uczą języka polskiego, natomiast różnią się w zakresie efektywności nauczania matematyki.

### 4.4. Podsumowanie

Miara zróżnicowania międzyszkolnego na danym etapie kształcenia jest przydatnym wskaźnikiem charakteryzującym system szkolny. Pozwala na określenie, na ile zróżnicowanie wyników nauczania uczniów może zostać wyjaśnione podziałem uczniów na szkoły. Miara ta może służyć również za wskaźnik jednolitości systemu kształcenia. Umożliwia także interesujące porównania międzynarodowe, w których można odnieść stopień zróżnicowania systemu szkolnego do osiągniętych wyników.

Na poziomie trzeciej klasy szkoły podstawowej obserwujemy relatywnie niewielkie zróżnicowanie międzyszkolne wyników. Po uwzględnieniu zmiennych odpowiadających za indywidualne cechy uczniów i ich rodzin, czyli kontrolując różnice w składach szkół, oszacowanie odsetka zróżnicowania wyników, który można przypisać szkołom, jest jeszcze mniejsze. Można więc powiedzieć, że za część zróżnicowania średnich wyników szkół odpowiada nielosowy podział uczniów na szkoły.

Stały monitoring zróżnicowania międzyszkolnego na poziomie jednolitego kształcenia umożliwia krajowy system egzaminacyjny, który dysponuje populacyjnymi danymi odnośnie wyników uczniów na sprawdzianie w klasie szóstej oraz egzaminie gimnazjalnym, czyli po drugim i trzecim etapie kształcenia. Wiadomo, że zróżnicowanie międzyszkolne gimnazjów w dużych miastach wykazuje trend wzrostowy, natomiast zróżnicowanie średnich wyników szkół podstawowych ze sprawdzianu utrzymuje się na w miarę stałym poziomie. Oznacza to, że na progu gimnazjum – w dużych miastach – mamy do czynienia z silnym procesem selekcji do szkół. Dla pełniejszego obrazu procesów zachodzących w systemie oświaty, przydałby się dobry pomiar diagnostyczny, także po etapie edukacji wczesnoszkolnej. Pozwoliłoby to zbadać hipotezę o rosnącym zróżnicowaniu międzyszkolnym wyników pomiędzy pierwszym a drugim etapem edukacyjnym, a także dałoby narzędzia do ewaluacji efektywności nauczania szkół podstawowych, co pozwoliłoby na skuteczniejszą politykę edukacyjną oraz na upowszechnienie dobrych praktyk zidentyfikowanych w szkołach o wyższej niż średnia efektywności nauczania.

Pełniejszy obraz działania procesów, które wpływają na zróżnicowanie wyników uczniów, możemy uzyskać, uwzględniając dodatkowo podział uczniów na klasy. Otrzymujemy wtedy informacje o tym, na ile podział uczniów na oddziały w obrębie szkoły, wyjaśnia całkowite zróżnicowanie wyników lub, mówiąc inaczej, na ile oddziały w danej szkole różnią się między sobą średnimi wynikami. Kontrolując przy tym wpływ cech indywidualnych ucznia oraz jego rodziny, na które szkoła nie ma wpływu, otrzymujemy miarę zróżnicowania wyników uczniów, którą możemy przypisać różnej efektywności nauczania w poszczególnych szkołach i oddziałach.

Na poziomie pierwszego etapu edukacyjnego uczniowie zazwyczaj kształceni są przez jednego nauczyciela, co pozwala na pewniejsze powiązanie zróżnicowania efektów nauczania na poziomie oddziały z pracą nauczyciela. Po uwzględnieniu wpływu czynników indywidualnych i społecznych, oddziały nie różnią się jednak między sobą efektywnością nauczania w dziedzinie nauki czytania i świadomości językowej. Niewielkie zróżnicowanie efektywności nauczania dla oddziałów obserwujemy w wypadku matematyki. Oznacza to, że być może istnieją różne praktyki związane z nauczaniem tego przedmiotu na etapie edukacji wczesnoszkolnej. Z pewnością warto przyrzeć się bliżej tym najlepszym, aby móc je upowszechnić w systemie.

Z badań *Szkolnych uwarunkowań efektywności kształcenia* wyłania się obraz w dużym stopniu jednolitego systemu szkolnego na pierwszym etapie edukacyjnym. Istnieją szkoły o efektywności nauczania odbiegającej od średniej krajowej, ale jest ich stosunkowo niewiele (ok. 10%).

### Literatura cytowana

- Dolata, R. (2009). Cicha rewolucja w polskiej oświacie – proces różnicowania się gimnazjów w dużych miastach. W: B. Niemierny, M. K. Szmigiel (red.), *Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej* (s. 111–123). Kielce: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- 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, 41–47.
- Dolata, R. (2011). *Analiza różnicowania się systemu oświaty w Polsce na poziomie szkół podstawowych i gimnazjów, edycja 2011*. Centralna Komisja Egzaminacyjna, [http://ewd.edu.pl/downloads/publikacje/analiza\\_roznicowania\\_2011.pdf](http://ewd.edu.pl/downloads/publikacje/analiza_roznicowania_2011.pdf), data dostępu: 19.12.2012.
- OECD (2010). *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes* (Volume II). <http://dx.doi.org/10.1787/9789264091504-en>. Data dostępu: 22.12.2012.
- Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and data analysis methods*. Thousand Oaks, London, New Delhi: Sage Publications.
- Snijders, T. A. B., Bosker, R. J. (2012). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage Publications.



## Rozdział 5

### Inteligencja a wyniki nauczania

Inteligencja zajmuje w schemacie badania SUEK kluczowe miejsce. Uczenie się szkolne jest nierozdzielnie związane z procesami poznawczymi. Opanowanie nowych wiadomości czy umiejętności wymaga m.in. percepcji bodźców, zasobów uwagi i pamięci, czy też zdolności rozumowania. Jedną z indywidualnych właściwości każdego człowieka, która wyznacza te procesy, jest właśnie inteligencja. Uwzględnianie jej w badaniu efektów nauczania szkolnego wydaje się więc naturalne i nie wymaga szczególnego uzasadniania. Bez kontroli tego czynnika nie sposób byłoby skutecznie kontrolować pozaszkolnych czynników odpowiedzialnych za osiągnięcia szkolne.

W niniejszym rozdziale zajmujemy się znaczeniem inteligencji dla wyników nauczania na koniec I etapu edukacyjnego w trzech obszarach: kompetencji matematycznych, czytania oraz świadomości językowej. Przedstawione zostaną wybrane ujęcia teoretyczne inteligencji oraz wynikające z nich powiązania tego konstruktów z wynikami nauczania. Wskazane zostaną trudności w badaniu tej zależności i w wyciąganiu wniosków dotyczących jej siły. Poruszony zostanie również problem możliwego zróżnicowania siły zależności między inteligencją a wynikami nauczania w poszczególnych szkołach i klasach.

#### 5.1. Koncepcje inteligencji

Inteligencja stanowi przedmiot zainteresowania psychologii od jej początków jako dyscypliny naukowej, jednak pomimo ponad stulecia badań nie wypracowano jej jednolitej definicji. 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 powstałej w tym nurcie zdolności te mają strukturę hierarchiczną lub są od siebie niezależne. Autorem pierwszej czynnikowej teorii inteligencji był Charles Spearman, który sprowadzał pomiar inteligencji do określenia proporcji dwóch zdolności potrzebnych do rozwiązania konkretnego zadania poznawczego: czynnika ogólnego  $g$  i czynnika specyficznego  $s$ . Zadania, które były dobrze „nasycone” czynnikiem  $g$  traktował jako te bardziej diagnostyczne w ocenie poziomu inteligencji ogólnej (Nęcka, 2003, s. 31). Zarówno Spearman jak i jego następcy, tworząc własne hierarchiczne



## CZY SZKOŁA MA ZNACZENIE?

modele inteligencji, próbowali definiować czynnik g, przy czym istotne jest, że pierwotnie był on traktowany jako konstrukt teoretyczny służący wyjaśnianiu korelacji między wynikami różnych testów i pozwalający określić ich przydatność w diagnozie poziomu inteligencji ogólnej.

Jednym z popularniejszych ujęć hierarchicznych jest teoria Cattella, która postuluje, iż inteligencja ogólna obejmuje dwa komponenty – inteligencję płynną oraz skryształowaną (por. 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. Grupa czynników środowiskowych dotyczy stanu zdrowia, odżywienia i innych biologicznych aspektów funkcjonowania organizmu, ale wyklucza efekty związane z treningiem i wychowaniem. Stanowi ona podbudowę do rozwoju inteligencji skryształowanej, obejmującą zgromadzone w toku życia wiadomości, umiejętności i doświadczenie istotne w danym kontekście kulturowym, a także zdolność do korzystania z nich (por. Nęcka, 2003; Wills, Dumont i Kaufman, 2011).

Przykładem modelu zakładającego równorzędność zdolności poznawczych jest model Guilforda. Punktem wyjścia dla jego koncepcji była nie analiza czynnikowa wyników testów, tylko model teoretyczny, do którego Guilford dobierał lub wymyślał poszczególne zadania testowe. Wyróżnia on trzy wymiary opisujące zdolności poznawcze: rodzaj przetwarzanego materiału (figuralny, semantyczny, behawioralny, symboliczny), rodzaj operacji, która jest uruchamiana (poznawanie, operacje pamięciowe, wytwarzanie konwergencyjne, dywergencyjne, ocena efektów myślenia) oraz otrzymywany wytwór czynności intelektualnych (jednostki, klasy, relacje, systemy, przekształcenia i implikacje). Pozwoliło to na wyróżnienie 120 różnych zdolności umysłowych, a po rewizji modelu (dokonanej przez autora) i rozbiciu materiału figuralnego na słuchowy i wzrokowy – aż 180 (por. Nęcka, 2003; Wills i in., 2011).

We współczesnych ujęciach inteligencji odchodzi się od jej statycznego (strukturalnego) rozumienia na rzecz ujęć dynamicznych. Nęcka (2009, s. 26) proponuje, by przyjąć, iż „inteligencja jest procesem mobilizacji zasobów poznawczych umysłu ludzkiego”. Zasoby te obejmują formalne cechy systemu poznawczego (tempo przetwarzania informacji, mobilność procesów umysłowych), jego centralnych modułów nadzorujących i kontrolujących (np. 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, 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 podobny sposób kompensowane mogą być inne zasoby.

Z dynamicznym ujęciem inteligencji korespondują zmiany w rodzaju zadań stawianych przed uczniami na kolejnych etapach edukacji. W początkowym etapie edukacji istotne jest trenowanie funkcji wykonawczych dzieci, przede wszystkim kontroli procesów uwagi i pamięci roboczej, które to funkcje umożliwiają przyswajanie nowej wiedzy w sposób systematyczny, rozwijając tym samym umiejętności samoregulacyjne ucznia (Welsh i Pennington, 1988; Bull i Scerif, 2001). Mimo różnorodności modeli pamięci roboczej większość z nich podkreśla jej istotną rolę w kontrolowanym przetwarzaniu informacji, sprawowanym przez związany z procesami uwagi mechanizm planowania i nadzoru (Engle, Sędek, von Hecker i McIntosh, 2004). Poza tym selektywna uwaga jest warunkiem efektywnego uczenia się, gdy lekcje prowadzone są w nowych, często trudnych poznawczo warunkach (Sękowski, 2001), np. w licznych klasach.

Trening ukierunkowany na mobilizację zasobów poznawczych, sprzyja rozwiązywaniu nowych problemów i zadań, przed którymi stawiany jest uczeń. Badania podłużne prowadzone przez Wyrwicz (2005) akcentują znaczenie zdolności poznawczych przedszkolaków, przede wszystkim pamięci, w wyjaśnianiu zmienności wyników w nauce szkolnej.

## 5.2. Inteligencja a osiągnięcia szkolne

Definiowanie inteligencji jako zdolności do przystosowania się do środowiska kieruje uwagę na jej rolę jako predyktora szeroko pojętego sukcesu życiowego jednostki, np. osiągniętego prestiżu zawodowego, wysokości dochodów, zdrowia, wykształcenia itp. (np. Adey, Csapo, Demetriou, Hautamaki i Shayer, 2007; Cheng i Furnham, 2012; Der, Batty, Deary, 2009; Sternberg, Giorenko i Bundy, 2001). W nurcie tym mieszczą się także osiągnięcia szkolne. Inteligencja identyfikowana jest jako jeden z najważniejszych i najbardziej stabilnych predyktorów sukcesu szkolnego dziecka, który pozwala wyjaśniać od kilkunastu do kilkudziesięciu procent zmienności wyników w nauce (np. Bossaert, Doumen, Buyse i Verschueren, 2011; Deary, Strand, Smith i Fernandes 2007; Greven, Harlaar, Kovas, Chamorro-Premuzic i Plomin, 2009; Hofer, Kuhnle, Kilian i Fries, 2012; Niesser i in., 1996; Teo, Carlson, Mathieu, Egeland i Sroufe, 1996; Rescorla i Rosenthal, 2004; Sternberg i in., 2001; Watkins, Lei i Canivez, 2007). Relacja ta widoczna jest pomimo różnorodności stosowanych zarówno ujęć i miar inteligencji, jak i aspektów i miar osiągnięć szkolnych, choć zwykle bywa silniejsza dla przedmiotów ścisłych (np. matematyka, fizyka) niż humanistycznych (np. historia, język ojczysty; np. Ferrer i McArdle, 2004; Heaven i Ciarrochi, 2012; Teo i in., 1996). Jej siła zmienia się również na różnych etapach edukacyjnych. W okresie edukacji szkolnej jest silniejsza, a słabnie na etapie studiów wyższych, podczas których na znaczeniu zyskują czynniki natury motywacyjnej, zainteresowania i cechy osobowości

## CZY SZKOŁA MA ZNACZENIE?

(Furnham, Chamorro-Premuzic i McDougall, 2003; Lin i Humphreys, 1977). Słabsza korelacja na późniejszych etapach edukacji może wynikać z mniejszych różnic wewnątrzgrupowych (efekt selekcji) oraz z tego, że pod wpływem wiedzy i doświadczeń, nabywanych na kolejnych etapach edukacji, wzrasta znaczenie inteligencji skryzalizowanej (Strelau i Zawadzki, 2008).

Podobna słabnąca moc predykcyjna ujawnia się w przypadku pamięci roboczej. Duża w szkole podstawowej, słabnie na kolejnych etapach nauki, w coraz mniejszym stopniu wyjaśniając zróżnicowanie osiągnięć szkolnych, np. wraz z wiekiem przestaje być istotnie związana z osiągnięciami matematycznymi (Reuhkala, 2001). Należy zauważyć, że zarówno inteligencja płynna, jak i pamięć robocza, są niezależnymi predyktorami osiągnięć szkolnych (Gathercole, Alloway, Willis i Adams, 2006; za: Cowan i Alloway, 2009).

Czynnikiem, który utrudnia analizę zależności między inteligencją a osiągnięciami szkolnymi, jest zróżnicowanie konceptualizacji i pomiaru inteligencji oraz osiągnięć szkolnych w badaniach empirycznych. W zakresie osiągnięć szkolnych w literaturze spotyka się wskaźniki takie jak np. oceny semestralne i końcoworoczne oceny szkolne oraz ich średnią, pozyskiwane z dokumentacji szkoły lub raportowane przez osoby badane (np. Hofer i in. 2012; Morrison, Rimm-Kauffman i Pianta, 2003), oceny poziomu umiejętności uczniów dokonywane przez nauczycieli (np. Greven i in., 2009), wyniki egzaminów zewnętrznych (np. Deary i in., 2007) czy standaryzowane testy osiągnięć (np. Bossaert i in., 2011; Teo i in. 1996; Hofer i in. 2012).

W zakresie pomiaru inteligencji ogólnej stosowane są standardowe testy inteligencji ogólnej (np. Graziano, Reavis, Keane i Calkins, 2007; Rescola i in., 2004; Watkins i in., 2007), podtesty różnych skal (np. Greven i in., 2009; Teo i in., 1996), z których wyników w toku analizy czynnikowej ekstrahuje się czynnik *g* (np. Deary i in. 2007) oraz łączone lub używane pojedynczo testy niewerbalne i werbalne (np. Bossaert i in. 2011; Hofer i in., 2012; Morrison i in., 2003). Te ostatnie pojawiają się jednak rzadziej, czego przyczyną może być między innymi ich fragmentaryczność w ujmowaniu zdolności poznawczych człowieka.

Spośród wymienionych powyżej metod pomiaru najczęściej przytaczanymi w literaturze światowej wydają się być miary inteligencji ogólnej, obejmujące różnorodne aspekty funkcjonowania poznawczego (werbalne, niewerbalne, przestrzenne) i dające tym samym najpełniejszy obraz możliwości poznawczych jednostki. Z drugiej jednak strony wykorzystywanie miar inteligencji ogólnej w badaniach edukacyjnych jest równoznaczne z uwzględnieniem tych jej aspektów, które obejmują zgromadzoną przez jednostkę wiedzę i doświadczenie, a więc także wiedzę szkolną. W sytuacji badania zależności inteligencja – osiągnięcia staje się to problematyczne, z uwagi na nakładanie się zakresów znaczeniowych mierzonych konstruktów.

Powyższe różnice, na poziomie operacjonalizacji zarówno osiągnięć jak i inteligencji, w połączeniu z odmiennymi schematami badawczymi (przekrojowe *versus* podłużne/z odroczo-nym pomiarem osiągnięć) oraz różnicami w wieku osób badanych sprawiają, iż trudno jest

precyzyjnie określić siłę omawianej zależności. Szacunkowo przyjąć można, iż w przypadku badań podłużnych wykorzystujących miary inteligencji ogólnej oraz miary inteligencji zawierające zarówno komponenty werbalne, jak i niewerbalne, raportowana w literaturze światowej siła relacji kształtuje się na poziomie 0,6–0,8 (np. Greven i in., 2009; Deary i in., 2007; Rescorla i in., 2004; Teo i in., 1996) i zależec może w pewnym stopniu od kontrolowanych zmiennych (np. przy kontroli wcześniejszych osiągnięć spada do poziomu 0,15–0,35, Watkis i in., 2007). Słabsze zależności z osiągnięciami notuje się w przypadku inteligencji płynnej – w zależności od wieku osób badanych kształtują się one na poziomie od około 0,2 do około 0,4 (Bossart i in., 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 przedstawia Dolata (2008), wskazując, że na etapie szkoły podstawowej (od około II 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.

Wspomniany już problem odrębności (lub jej braku) inteligencji i osiągnięć szkolnych został przez badaczy zauważony, prowadząc do dyskusji nad relacjami między nimi. 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 (por. Watkins i in., 2007). Trzecie ujęcie omawianej relacji postuluje, że inteligencja stanowi źródło osiągnięć szkolnych (por. Watkins i in., 2007).

### **5.3. Zróznicowanie siły zależności pomiędzy osiągnięciami a inteligencją w szkołach**

O możliwym zróznicowaniu siły zależności między inteligencją a poziomem osiągnięć (zależnym od szkoły) wnioskować można na podstawie predykcji wynikających z triarchicznej teorii inteligencji Sternberga (2003). Wyróżnia ona trzy typy inteligencji: analityczną, praktyczną i twórczą. Tradycyjne nauczanie szkolne wydaje się faworyzować osoby o wysokich zdolnościach analitycznych (por. Sternberg i Spear-Swerling, 2002), tj. takie, które dobrze radzą sobie z abstrakcyjnymi problemami wymagającymi analizy, oceny, porównania, czy wyszukiwania powiązań. Właśnie takie zadania dominują w klasie szkolnej oraz na sprawdzianach, a ich rozwiązywanie sprzyja rozwojowi inteligencji analitycznej. Pozostałe dwa typy uzdolnień wykorzystywane są w niewielkim stopniu. Co więcej, zdolności analityczne to ta właściwość poznawcza, którą zazwyczaj mierzą też testy zdolności poznawczych (Sternberg, 2003),

## CZY SZKOŁA MA ZNACZENIE?

czyli m.in. testy inteligencji. Jednakże z badań wynika, iż nauczanie w sposób wykorzystujący i rozwijający inne rodzaje inteligencji (twórczą i praktyczną), sprzyja osiągnięciu przez uczniów wyższych wyników w nauce na różnych etapach edukacji w porównaniu do standardowych metod nauczania (Sternberg, 2003) oraz wykorzystuje potencjał tych dzieci, które mają dobrze rozwinięte zdolności praktyczne i/lub są kreatywne, lecz słabiej radzą sobie z problemami analitycznymi.

Mimo znaczących różnic w ujęciu inteligencji wg Catella i Sternberga (m.in. w zakresie struktury i możliwości rozwoju inteligencji), istnieją przesłanki, by sądzić, iż sposób nauczania może stanowić czynnik moderujący siłę zależności między inteligencją dziecka a jego osiągnięciami szkolnymi. Ściślej, nauczanie, które uwzględnia także wykorzystywanie wiedzy w praktyce oraz proponowanie nowych idei i wypracowywanie własnych rozwiązań (czyli inteligencję praktyczną i twórczą), może zmniejszać siłę zależności między inteligencją uczniów a ich osiągnięciami. Jednak testy osiągnięć szkolnych bazują raczej na wiedzy, dlatego hipoteza ta wymaga założenia dotyczącego mechanizmu tej relacji. Jeżeli metody te, dzięki dopasowaniu do mocnych stron uczniów (zwłaszcza twórczych i uzdolnionych praktycznie), czynią naukę bardziej interesującą dla dzieci oraz pomagają im lepiej zrozumieć i przyswoić materiał, to przekłada się to na wyniki testów osiągnięć. Z drugiej jednak strony pamiętać należy, że inteligencja twórcza zawiera elementy inteligencji płynnej (radzenia sobie z zupełnie nowymi problemami; Sternberg, 2003), więc strategie ją wykorzystujące, mogą też nasilać siłę omawianej zależności.

Dodatkowym czynnikiem, który może wpływać na sposób nauczania i może moderować siłę zależności pomiędzy inteligencją a osiągnięciami uczniów, są subiektywne teorie zdolności nauczyciela. Naiwne teorie zdolności to przekonania na temat stopnia, w jakim różne właściwości człowieka (inteligencja, zdolności matematyczne, werbalne itp.) są stałe/modyfikowalne. „Teoretycy stałości” są przekonani o ich niezmienności oraz braku możliwości ich rozwijania, natomiast „teoretycy zmienności” wierzą, że zdolności można rozwijać. Co ważne, ludzie mogą żywić różne przekonania na temat różnych zdolności, np. mogą wierzyć, że zdolności matematyczne są niezmienne, ale językowe podlegają rozwojowi (Dweck, Mangels i Good 2004; Dweck i Molden, 2005). Teoretycy przyrostu przekonani są o znaczeniu wysiłku w zdobywaniu umiejętności i gotowi są go podejmować; teoretycy stałości – o znaczeniu zdolności (Blackwell, Tresniewski i Dweck, 2007). Co za tym idzie, teoretycy przyrostu częściej interpretują trudności i porażki w sposób adaptacyjny (zbyt krótki czas przygotowań, niewystarczających wysiłek), natomiast teoretycy stałości częściej wykazują atrybucje nieadaptacyjne, tj. do braku zdolności (Blackwell i in., 2007, Lachowicz-Tabaczek, 2004, Robins, Pals, 2002). Wiąże się to z kontrolowalnością zdarzeń – wysiłek i praca są pod kontrolą jednostki, a zdolności – nie.

Przypuszczać można, że nauczyciel postrzega swoich uczniów, dokonuje interpretacji ich sukcesów i porażek oraz dobiera formy pracy z nimi zależnie od własnego przekonania na temat stałości/zmienności zdolności intelektualnych. Nauczyciele – teoretycy stałości, którzy zaobserwują u swoich uczniów trudności w nauce, pomimo dużych nakładów pracy mogą wnioskować o niskich zdolnościach ucznia. Z kolei nauczyciele – teoretycy zmienności przekonani będą o znaczeniu wkładanego przez dziecko wysiłku dla rozwoju jego zdolności, stąd porażki będą atrybuować np. do niewystarczającego wysiłku. Teoretykom stałości będzie więc zaklasyfikować uczniów do jednej ze stereotypowych grup (np. uczniowie słabi, średni, zdolni) i interpretować ich sukcesy i porażki w kategoriach posiadania lub braku danego rodzaju zdolności. W konsekwencji pierwsze otrzymane przez ucznia oceny, a co za tym idzie jego sukces lub porażka, mogą doprowadzić do wypracowania przez nauczyciela trudnej do zmiany opinii o uczniu. Opinia ta działać może jak samospełniająca się przepowiednia – w przypadku negatywnej oceny zdolności ucznia nauczyciel może rezygnować z podejmowania próby dopasowania form pracy do konkretnego ucznia, uznając, iż jest on po prostu mało zdolny i nic na to nie można poradzić. Zupełnie inny styl pracy z uczniami może pojawić się u nauczycieli-teoretyków zmienności, których subiektywna teoria zdolności oraz orientacja sprawnościowa wymuszają inne spojrzenie na ucznia i jego zdolności. Ci nauczyciele nie będą traktować zdolności swoich uczniów jako stałej, niemożliwej do zmiany cechy. Można się spodziewać, że wyjdą z założenia, że porażka nie skazuje uczniów na problemy z danym przedmiotem. Szukając przyczyn ich porażki, mogą próbować wprowadzać zmiany w podejściu ucznia do nauki (np. konieczność większej liczby ćwiczeń) lub we własnych strategiach uczenia. W efekcie działania nauczyciela zostaną ukierunkowane na podtrzymanie lub zwiększenie poziomu motywacji uczniów do zrozumienia nauczanego przedmiotu i wzbogacania własnej wiedzy. W takim przypadku uczniowie mogą skoncentrować się na rozwijaniu i regularnym ćwiczeniu swoich kompetencji, w mniejszym zaś stopniu na tym, jak oceni ich nauczyciel, co istotnie redukuje stres ucznia związany z sytuacją uczenia się szkolnego. Przypuszczać też można, że nauczyciel przekazywać będzie uczniom swoje przekonania dotyczące stałości/zmienności zdolności i w ten sposób będzie wpływał na interpretację sukcesów i porażek przez uczniów oraz przekonania o znaczeniu wysiłku w osiągnięciu sukcesu szkolnego. Wiadomo jednocześnie, że naiwne teorie inteligencji są jednym z predyktorów ocen szkolnych (por. Offors, Andersson 2007; Blackwell i in., 2007), mogą w znacznym stopniu wpływać na osiągnięcia uczniów.

Podsumowując, nauczyciele w zależności od wyznawanych teorii zdolności mogą w odmienny sposób interpretować wyniki i pracę uczniów na lekcji, co może przełożyć się na zróżnicowanie w obrębie siły zależności pomiędzy zdolnościami intelektualnymi uczniów a ich osiągnięciami. Przekonanie o zmienności i możliwości rozwijania zdolności przez nauczyciela

## CZY SZKOŁA MA ZNACZENIE?

wiązać się będzie z mniejszą siłą zależności inteligencja – osiągnięcia, natomiast przekonanie o stałości – ze zwiększeniem tej siły. Patrząc z perspektywy teorii Sternberga oraz teorii Dweck, przypuszczać można, że zróżnicowanie siły omawianej zależności powinno wystąpić przede wszystkim na poziomie klas i atrybuowane powinno być do procesów w nich zachodzących, na które wpływa nauczyciel.

Z badań nad efektywnością uczenia się wiemy, że determinuje ją wiele czynników (Galloway, 1988), wśród których zazwyczaj wyróżnia się trzy klasy zmiennych: właściwości ucznia, sytuację uczenia się i wspomniane wyżej działania i właściwości nauczyciela (Włodarski i Matczak, 1996). Sytuacja uczenia się definiowana jest jako warunki zewnętrzne i sposoby postępowania, w wyniku których uczeń zdobywa wiadomości i umiejętności (Ledzińska i Czerniawska, 2011), czyli m.in. metody nauczania, system kar i nagród oraz sposoby udzielania informacji zwrotnej uczniom, stosowane materiały, a także inne warunki zewnętrzne. W szkołach podstawowych nauczyciele we wszystkich oddziałach na danym poziomie dysponują zazwyczaj tymi samymi podręcznikami, uczestniczą w tych samych szkoleniach, korzystają z tych samych materiałów. Na tej podstawie możemy przypuszczać, iż zróżnicowanie siły zależności pomiędzy inteligencją a osiągnięciami uczniów może pojawić się również na poziomie szkół.

### 5.4. Problem badawczy

Jeden z celów badania szkolnych uwarunkowań efektywności kształcenia stanowi weryfikacja zależności między poziomem inteligencji dziecka a poziomem jego osiągnięć szkolnych na pierwszym etapie edukacji, przy uwzględnieniu potencjalnego moderacyjnego wpływu szkoły i oddziału klasowego. W badaniu przyjęto trzecie z omawianych wcześniej podejść do badania relacji inteligencja – osiągnięcia, zgodnie z którym inteligencja stanowi źródło osiągnięć szkolnych. Za podstawę teoretyczną przyjęto model inteligencji Cattela (Nęcka, 2003), koncentrując się na wyodrębnionej w nim inteligencji płynnej. Przyjęto, że inteligencja płynna stanowi podstawę do nabywania wiedzy i umiejętności szkolnych, lecz nie jest z nimi tożsama.

Na podstawie przyjętego podejścia do analizowania relacji inteligencja – osiągnięcia oraz wyników wcześniejszych badań realizowanych w tym paradygmacie, skoncentrowanych na wyjaśnianiu czynników moderujących wpływ inteligencji na osiągnięcia uczniów, zakładano iż wraz z poziomem inteligencji dziecka wzrastają jego osiągnięcia szkolne, a siła tej zależności zależy może od środowiska klasowego i szkolnego.

## 5.5. Pomiar inteligencji

Do pomiaru inteligencji płynnej wykorzystano Test Matryc Ravena, wersja Standard, forma Klasyczna (TMS-K), który choć wyrasta z innej koncepcji teoretycznej, wykorzystywany jest w tym celu ze względu na podobieństwo spearmanowskiego edukacyjnego komponentu inteligencji (mierzonego przez test) do inteligencji płynnej w ujęciu Cattella (Jaworowska, Szustrowa, 2000). Test składa się 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.

Za wskaźnik poziomu inteligencji płynnej uczniów przyjęto oszacowania poziomu cechy ukrytej (EAP) uzyskane w toku analizy IRT w jednowymiarowym modelu dwu-parametrycznym z wykorzystaniem metody estymacji MML (por. np. de Ayala, 2009). Uwzględnionych zostało 5429 obserwacji. W celu uzyskania precyzyjniejszych oszacowań parametrów zadań najłatwiejszych i najtrudniejszych (a tym samym lepszych oszacowań poziomu inteligencji płynnej uczniów) przeprowadzona została analiza wielogrupowa IRT<sup>1</sup> z wykorzystaniem rekordów wykonania TMS-K zebranych wśród dzieci młodszych oraz starszych w badaniu SUEK, czyli takich, wśród których średni poziom cechy jest odpowiednio niższy oraz wyższy w porównaniu do uczniów z próby. Byli to uczniowie pierwszej klasy szkoły podstawowej i pierwszej klasy gimnazjum przebadani w ramach projektu badawczego EWD<sup>2</sup>. Obliczenia przeprowadzone zostały za pomocą pakietu MIRT (Glas, 2010). Uzyskane oszacowania EAP przeniesiono na skalę o średniej 100 i odchyleniu standardowym 15 z uwzględnieniem wag próbkowania. Informacje o sposobie pomiaru oraz przygotowaniu wskaźników pozostałych zmiennych uwzględnionych w analizach przedstawione zostały we wcześniejszych rozdziałach.

<sup>1</sup> Przeprowadzenie analizy wielogrupowej IRT podyktowane było faktem, iż test zawiera część zadań bardzo łatwych i bardzo trudnych dla uczniów z próby SUEK, tj. takich, na które bardzo niewielu uczniów udziela nieprawidłowej lub prawidłowej odpowiedzi. Takie rozkłady odpowiedzi nie pozwalają na dobre oszacowanie parametrów zadań, co z kolei przekłada się na słabe szacowanie poziomu umiejętności uczniów (por. Verhelst, 2009). Dzięki analizie wielogrupowej IRT możliwe częściowe rozwiązanie tego problemu dzięki szacowaniu parametrów zadań na podstawie odpowiedzi z wszystkich trzech prób. Różniły się one poziomem cechy, a więc też rozkładami odpowiedzi na zadania najłatwiejsze i najtrudniejsze.

<sup>2</sup> Projekt badawczy "Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD)" współfinansowany z funduszy Unii Europejskiej; badanie wśród uczniów klas I szkół podstawowych oraz gimnazjów zrealizowano na początku 2010 roku; uwzględniono dane odpowiednio 5596 i 5565 uczniów.



### 5.6. Wyniki

#### 5.6.1. Ogólny plan analiz

W celu weryfikacji hipotez przeprowadziliśmy serię wielopoziomowych analiz regresji, w których zmienne zależne stanowiły kolejno osiągnięcia uczniów z zakresu matematyki, czytania oraz świadomości językowej. Poprzedziła je analiza statystyk opisowych. Modelowanie przeprowadzono w czterech krokach – w pierwszym kroku dla każdej zmiennej zależnej zbudowano model pusty z uwzględnieniem poziomu klas i szkół. W drugim kroku zbudowano wielopoziomowe modele kontrolne z losowymi stałymi uwzględniające podstawowe zmienne kontrolowane w badaniu (oba rodzaje modeli zostały opisane szczegółowo w rozdziale 3). Modele te posłużyły jako punkty odniesienia w zakresie weryfikacji polepszenia dopasowania modeli po dodaniu zmiennych i/lub parametrów. W trzecim kroku, celem weryfikacji hipotez dotyczących zróżnicowania siły zależności między poziomem inteligencji a poziomem osiągnięć w szkołach i klasach, dodano losowe nachylenia na poziomie klas i szkół dla zmiennej inteligencji. W dalszych częściach rozdziału będziemy się odwoływać do opisanych wyżej modeli za pomocą liczb porządkowych.

Obliczenia przeprowadzone zostały na zmiennych wycentrowanych wokół średniej w całej próbie (centrowaniu podlegały zmienne ilościowe, tj. liczba lat nauki rodziców w latach, wartości indeksu dóbr oraz oszacowania poziomu inteligencji przed usunięciem z analiz obserwacji z brakami danych). Uwzględniono dane jedynie tych uczniów, dla których dostępny był komplet zmiennych. Wszystkie analizy przeprowadzono z wykorzystaniem wag warunkowych generowanych oddzielnie dla analiz różniących się zmienną zależną (poziom osiągnięć z matematyki, czytania i świadomości językowej; wagi opisane zostały w rozdziale 3). Ze względu na wykorzystanie w analizach PV jako oszacowań poziomu osiągnięć uczniów (opisanych szczegółowo w rozdziale 2), obliczenia do każdego rodzaju modelu dla każdego typu testu osiągnięć powtórzone zostały pięciokrotnie, z wykorzystaniem każdego z pięciu PV jako zmiennej zależnej. Przedstawione w tabelach parametry stanowią wypadkowe wyników z pięciu powtórzeń tego samego modelu. Analizy przeprowadzono z wykorzystaniem oprogramowania HLM 6.06. Podkreślić należy, że analizy wykonano z wykorzystaniem PV obliczonych według nieco innej procedury niż opisana w rozdziale 2. Różnica polegała na przeprowadzeniu warunkowania ze względu na wszystkie zmienne niezależne uwzględniane w analizach (a także pewne zmienne, które planuje się wykorzystać w analizach w przyszłości). Informacje na temat celów i logiki warunkowania PV czytelnik znajdzie w pracy Von Daviera, Gonzaleza i Mislevy'ego (2009). Warunkowanie przeprowadzono zgodnie z procedurami stosowanymi w badaniach PIRLS i TIMSS (IEA, 2013).

### 5.6.2. Statystyki opisowe

W analizach z wykorzystaniem testu matematycznego jako zmiennej zależnej uwzględniono dane 4805 uczniów z 300 oddziałów i 172 szkół, co na poziomie indywidualnym stanowi 72,73% pierwotnej liczebności próby oraz 85,42% uczniów zrekrutowanych (tj. uczniów, których rodzice wyrazili zgodę na uczestnictwo dziecka w badaniu). Zbliżone odsetki obserwacji uwzględnione zostały w analizach z testem czytania oraz świadomości językowej jako zmiennymi zależnymi – w obu przypadkach stanowiły one 71,73% pierwotnej liczebności próby uczniów oraz 85,76% próby uczniów zrekrutowanych (liczba uczniów = 4739, liczba oddziałów = 300, liczba szkół = 172).

Statystyki opisowe dla wyników testu matematycznego, czytania i świadomości językowej znajdują się w tabeli 5.1. Ze względu na wykorzystanie w analizie wartości PV, dla każdego testu przedstawiono informacje z wykorzystywanych PV.

**Tabela 5.1. Testy osiągnięć szkolnych z matematyki, czytania i świadomości językowej – statystyki opisowe dla oszacowań umiejętności uczniów. Wyniki wystandaryzowane i przeniesione na skalę o średniej 100 i odchyleniu standardowym 15 z uwzględnieniem wag próbkowania**

	PV	liczebność	średnia	odch. std.	min.	maks.	skośność	kurtoza
test matematyczny	PV1	4805	100,64	14,88	47,03	146,67	-0,042	-0,013
	PV2		100,77	15,04	44,59	151,83	-0,038	0,037
	PV3		100,70	14,86	50,47	150,11	-0,075	0,000
	PV4		100,69	14,91	41,72	155,70	-0,047	0,091
	PV5		100,68	15,08	49,04	153,69	-0,024	0,056
test czytania	PV1	4739	100,62	14,86	54,18	149,30	0,012	-0,247
	PV2		100,73	14,95	50,11	148,32	0,050	-0,214
	PV3		100,64	14,89	52,64	150,57	0,024	-0,187
	PV4		100,81	14,93	50,25	154,78	0,041	-0,267
	PV5		100,85	14,83	51,51	150,29	-0,001	-0,239
test świadomości językowej	PV1	4739	100,80	14,82	54,18	149,30	-0,148	-0,059
	PV2		100,81	14,80	50,11	148,32	-0,144	-0,032
	PV3		100,94	14,84	52,64	150,57	-0,156	0,055
	PV4		101,00	14,87	50,25	154,78	-0,151	0,011
	PV5		100,79	15,01	51,51	150,29	-0,147	0,076

Źródło: opracowanie własne

## CZY SZKOŁA MA ZNACZENIE?

Średni wynik dla uczniów, których dane zostały wykorzystane w analizach, wyniósł ok. 101 punktów w każdym z testów, przy odchyleniu standardowym bliskim wartości 15. Rozkłady okazały się symetryczne (wartości bezwzględne skośności były nie wyższe niż 0,16), w przypadku testu pisania odbiegały od typowego dla rozkładu normalnego spiętrzenia (kurtoza od -0,267 do -0,187).

W przypadku wyników Testu Matryc Ravena w każdej z przeprowadzonych analiz średni wynik bliski był wartości 101 przy odchyleniu standardowym około 14,6. Rozkłady okazały się przesunięte w stronę wyników wysokich (skośność około -0,36) i nieco bardziej spiętrzone niż jest to typowe dla rozkładu normalnego (kurtoza około 0,26) w analizach dla czytania i świadomości językowej. Dane znajdują się w tabeli 5.2.

**Tabela 5.2. Test Matryc Ravena – statystyki opisowe dla analiz z wykorzystaniem wyników testów osiągnięć szkolnych z matematyki, czytania i świadomości językowej. Wyniki wystandaryzowane i przeniesione na skalę o średniej 100 i odchyleniu standardowym 15 z uwzględnieniem wag próbkowania**

Test Matryc Ravena	zmienna zależna w analizach: osiągnięcia szkolne w zakresie		
	matematyki	czytania	świadomości językowej
liczebność	4805	4739	4739
średnia	100,68	100,71	100,71
mediana	100,65	101,55	101,55
odch. st.	15,08	14,55	14,55
minimum	49,04	44,04	44,04
maximum	153,69	145,41	145,41
skośność	-0,024	-0,327	-0,327
kurtoza	0,056	0,263	0,263

Źródło: opracowanie własne

### 5.6.3. Inteligencja płynna a osiągnięcia w zakresie matematyki

W tabeli 5.3 przedstawiono szczegółowe wyniki analiz dla kolejnych modeli (pustego, kontrolnego, weryfikowanego i zmodyfikowanego), w których zmienną zależną stanowiły osiągnięcia szkolne uczniów z matematyki. Model (1) został opisany w rozdziale 3, wobec czego nie będzie omawiany. Model (2) stanowi zmodyfikowany wariant modelu podstawowego (również opisanego w rozdziale 3). Modyfikacja polegała na wyrażeniu wieku rodziców w liczbie lat potrzebnej do osiągnięcia danego poziomu wykształcenia. Jeżeli rodzice różnili się pod

tym względem, wybierano wartość wyższą. Jeżeli odnotowano brak danych dla jednego z rodziców, uwzględniano dane od drugiego rodzica. W kolejnym kroku dodano parametry wariancji nachyleń dla inteligencji. Oszacowania wariancji nachyleń dla oddziałów klasowych w zakresie Testu Matryc Ravena w modelu (3) wskazują, iż siła zależności między poziomem osiągnięć dziecka a jego poziomem inteligencji płynnej różni się między szkołami ( $p < 0,001$ , jednak wewnątrz szkół nie różni się między klasami ( $p > 0,500$ ).

**Tabela 5.3. Wyznaczniki osiągnięć matematycznych: znaczenie inteligencji płynnej i zróżnicowanie międzyoddziałowe oraz międzyszkolne w zakresie siły zależności osiągnięcia – inteligencja płynna przy kontroli czynników socjoekonomicznych**

**Wyniki trzypoziomowych analiz regresji, odporne (*robust*) oszacowania błędów standardowych**

zmienna zależna: osiągnięcia szkolne z matematyki	model (1)	model (2)	model (3)	model (4)
<b>oszacowanie efektów stałych</b>				
<b>poziom ucznia</b>				
stała	<b>99,09</b> (0,535)	<b>100,61</b> (0,399)	<b>100,56</b> (0,399)	<b>100,56</b> (0,399)
wiek dziecka w tygodniach		<b>0,065</b> (0,013)	<b>0,065</b> (0,013)	<b>0,065</b> (0,013)
opóźniony tok nauki		<b>-14,19</b> (1,825)	<b>-14,04</b> (1,821)	<b>-14,09</b> (1,815)
przyspieszony tok nauki		<b>6,16</b> (2,970)	<b>6,71</b> (2,831)	<b>6,67</b> (2,834)
wykształcenie rodziców w latach nauki		<b>0,842</b> (0,106)	<b>0,832</b> (0,105)	<b>0,833</b> (0,105)
indeks zasobów materialnych rodziny dziecka		<b>0,117</b> (0,018)	<b>0,117</b> (0,019)	<b>0,118</b> (0,018)
inteligencja (TMS-K)		<b>0,537</b> (0,019)	<b>0,541</b> (0,018)	<b>0,540</b> (0,018)
<b>oszacowanie efektów losowych</b>				
wariancja efektów dla stałych dla szkół	<b>12,90</b>	<b>8,27</b>	<b>8,37</b>	<b>8,21</b>
wariancja nachyleń dla szkół dla inteligencji			<b>0,0070</b>	<b>0,0082</b>
wariancja efektów dla stałych dla oddziałów klasowych	<b>18,03</b>	<b>7,20</b>	<b>6,89</b>	<b>7,04</b>
wariancja nachyleń dla oddziałów klasowych dla inteligencji			0,0016	
wariancja na poziomie ucznia	<b>194,49</b>	<b>108,83</b>	<b>107,03</b>	<b>107,07</b>

liczba szkół = 172, liczba oddziałów = 300, liczba uczniów = 4805

Pogrubionym drukiem zaznaczono wartości istotne statystycznie na poziomie  $p < 0,05$ . W nawiasach podano błędy standardowe.

Źródło: opracowanie własne

W celu uzyskania oszacowania wariancji siły tej zależności w różnych szkołach, zbudowano model (4) zmodyfikowany, wykluczając parametr losowych nachyleń dla inteligencji dla oddziałów klasowych. Zgodnie z oczekiwaniami, wariancja nachyleń dla szkół pozostała istotna statystycznie i wyniosła 0,0082 ( $p < 0,001$ ), przy wartości 0,540 dla współczynnika regresji

## CZY SZKOŁA MA ZNACZENIE?

dla inteligencji. Wraz ze wzrostem poziomu inteligencji ucznia o 1 punkt, jego wyniki w teście matematycznym rosną przeciętnie o 0,54 punktu, jednakże w zależności od szkoły, do której uczęszcza, siła tej zależności różni się. W 95% szkół wraz ze wzrostem wyniku w Teście Matrycy Ravena o 1 punkt poziom osiągnięć szkolnych dziecka rośnie od 0,36 do 0,72 punktu (gdy wartość pozostałych zmiennych pozostaje na niezmiennym poziomie).

W celu weryfikacji modeli sprawdzono istotność statystyczną poprawy dopasowania w stosunku do modelu prostszego. Szczegółowe wyniki testowania istotności poprawy dopasowania modeli przedstawiono w tabeli 5.4. Ze względu na fakt, iż każdy model powtarzany był pięciokrotnie (dla każdej z pięciu PV jako zmiennej zależnej), porównywano dopasowania odpowiednich modeli dla odpowiednich PV. Model (4) okazał się istotnie lepiej dopasowany do danych niż model (2) (kontrolny) w czterech z pięciu powtórzeń. Brak istotnej statystycznej poprawy dopasowania modelu (4) w stosunku do (2) dla jednej z PV sprawia, że można za wystarczający do opisu badanej zależności przyjąć oszczędniejszy model (2). Jednak trzeba podkreślić, że międzyszkolna zmienność siły korelacji inteligencja – osiągnięcia szkolne z matematyki jest ciekawym fenomenem wartym dalszych analiz.

**Tabela 5.4. Weryfikacja poprawy dopasowania modeli. Wyniki testowania istotności poprawy dopasowania modelu bardziej złożonego w porównaniu z oszczędniejszym**

	model (2) k = 12
	(PV1) $\chi^2 = 16,63$
	(PV2) $\chi^2 = 14,51$
model (4)	(PV3) $\chi^2 = 14,75$
k = 14	(PV4) $\chi^2 = 4,21$
	(PV5) $\chi^2 = 20,46$
	df = 2

Pogrubionym drukiem zaznaczono wartości istotne statystycznie na poziomie  $p < 0,05$ .

k – liczba parametrów modelu

df – liczba stopni swobody

(PV1) – (PV5) – porównanie dla powtórzeń modeli, w którym zmienną zależną jest PV o danym numerze porządkowym

Źródło: opracowanie własne

### 5.6.4. Inteligencja płynna a osiągnięcia szkolne w zakresie czytania

Analizy, w których zmienną zależną stanowiły oszacowania poziomu umiejętności uczniów w zakresie czytania, przeprowadzone zostały według schematu zastosowanego dla umiejętności matematycznych. Ich wyniki przedstawione zostały w tabeli 5.5.

**Tabela 5.5. Wyznaczniki osiągnięć z zakresu czytania: znaczenie inteligencji płynnej**

**i różnicowanie międzyoddziałowe i międzyszkolne w zakresie siły zależności osiągnięcia – inteligencja płynna przy kontroli czynników socjoekonomicznych. Wyniki trzypoziomowych analiz regresji, odporne (*robust*) oszacowania błędów standardowych**

zmienna zależna: osiągnięcia szkolne w zakresie czytania	model (1)	model (2)	model (3)	model (4)	model (5)
<b>oszacowanie efektów stałych</b>					
<b>poziom ucznia</b>					
stała	<b>99,23</b> (0,504)	<b>99,03</b> (0,513)	<b>99,00</b> (0,516)	<b>98,96</b> (0,524)	<b>98,96</b> (0,528)
pleć dziecka <sup>a</sup>		<b>2,97</b> (0,669)	<b>2,97</b> (0,670)	<b>2,98</b> (0,679)	<b>2,97</b> (0,679)
wiek dziecka w tygodniach		<b>0,052</b> (0,015)	<b>0,052</b> (0,015)	<b>0,052</b> (0,015)	<b>0,052</b> (0,015)
opóźniony tok nauki		<b>-9,51</b> (2,147)	<b>-9,50</b> (2,146)	<b>-9,48</b> (2,149)	<b>-9,48</b> (2,160)
przyspieszony tok nauki		2,95 (2,352)	3,03 (2,358)	3,16 (2,445)	3,18 (2,268)
wykształcenie rodziców w latach nauki		<b>0,862</b> (0,140)	<b>0,992</b> (0,113)	<b>0,989</b> (0,112)	<b>0,990</b> (0,112)
HISEI		0,027 (0,018)			
indeks zasobów materialnych rodziny dziecka		<b>0,125</b> (0,016)	<b>0,132</b> (0,019)	<b>0,132</b> (0,019)	<b>0,132</b> (0,019)
inteligencja (TMS-K)		<b>0,400</b> (0,019)	<b>0,401</b> (0,019)	<b>0,404</b> (0,019)	<b>0,403</b> (0,019)
<b>oszacowanie efektów losowych</b>					
wariancja efektów dla stałych dla szkół	<b>11,24</b>	<b>8,35</b>	<b>8,61</b>	<b>8,44</b>	<b>8,44</b>
wariancja nachyleń dla szkół dla inteligencji				<b>0,0018</b>	<b>0,0029</b>
wariancja efektów dla stałych dla oddziałów klasowych	<b>15,30</b>	<b>6,26</b>	<b>6,17</b>	<b>6,33</b>	<b>6,22</b>
wariancja nachyleń dla oddziałów klasowych dla inteligencji				0,0026	
wariancja na poziomie ucznia	<b>193,74</b>	<b>133,19</b>	<b>133,28</b>	<b>132,35</b>	<b>132,64</b>

liczba szkół = 172, liczba oddziałów = 300, liczba uczniów = 4739

Pogrubionym drukiem zaznaczono wartości istotne statystycznie na poziomie istotności  $p < 0,05$ , w nawiasach podano błędy standardowe.

a – grupa odniesienia: chłopcy

Źródło: opracowanie własne

Model (2) stanowi zmodyfikowany wariant modelu podstawowego. Ponadto wskaźnik HISEI okazał się w nim nieistotnym statystycznie predyktorem osiągnięć w zakresie czytania, nie był włączany do kolejnych modeli. Oszacowania wariancji nachyleń dla oddziałów klasowych w zakresie inteligencji w modelu (4), testującym postawione hipotezy, wskazują, iż siła zależności między poziomem osiągnięć dziecka z czytania i jego poziomem

## CZY SZKOŁA MA ZNACZENIE?

inteligencji płynnej nie różni się między szkołami ( $p > 0,071$ ) oraz wewnątrz szkół nie różni się między oddziałami ( $p > 0,456$ ). Ze względu na bliską istotności statystycznej wariancji nachyleń dla szkół, zdecydowano się przetestować model z wykluczoną wariancją nachyleń dla oddziałów, lecz pozostawioną wariancją nachyleń dla szkół. Oszacowanie wariancji siły zależności inteligencja – osiągnięcia na poziomie szkół w modelu (5) wyniosło 0,0029 ( $p < 0,0018$ ) przy wartości 0,40 dla współczynnika regresji dla inteligencji. Wraz ze wzrostem wyników w Teście Matryc Ravena o 1 punkt, wyniki ucznia w teście czytania rosną przeciętnie o 0,41 punktu, jednakże w zależności od szkoły, do której uczęszcza, siła tej zależności różni się. W 95% szkół wraz ze wzrostem wyniku w Teście Matryc Ravena o 1 punkt, poziom osiągnięć szkolnych dziecka rośnie od 0,35 do 0,46 punktu (gdy wartość pozostałych zmiennych pozostanie na niezmiennym poziomie).

Weryfikacja poprawy dopasowania kolejnych modeli przeprowadzona została zgodnie ze schematem zastosowanym w przypadku modeli dla osiągnięć z matematyki. Szczegółowe dane dotyczące testowania istotności poprawy dopasowania modeli znajdują się w tabeli 5.6. Okazało się, że model (5) nie jest istotnie lepiej dopasowany do danych niż model kontrolny w każdym z pięciu powtórzeń. W sensie statystycznym pomimo istotności parametru nachylenia dla inteligencji na poziomie szkół stanowi to podstawę do jego odrzucenia. Należy jednak pamiętać, że po raz drugi zaobserwowano istotne statystycznie, międzyszkolne zróżnicowanie siły zależności inteligencja – osiągnięcia szkolne.

**Tabela 5.6. Weryfikacja poprawy dopasowania modeli. Wyniki testowania istotności poprawy dopasowania modelu bardziej złożonego w porównaniu z oszczędniejszym**

	model (2) k = 12
	(PV1) $\chi^2 = 1,91$
	(PV2) $\chi^2 = 4,00$
model (5)	(PV3) $\chi^2 = 1,92$
k = 14	(PV4) $\chi^2 = 0,092$
	(PV5) $\chi^2 = 5,03$
	df = 2

Pogrubionym drukiem opisywane są wartości istotne na poziomie  $p < 0,05$ .

k – liczba parametrów modelu

df – liczba stopni swobody

(PV1) – (PV5) – porównanie dla powtórzeń modeli, w którym zmienną zależną jest PV o danym numerze porządkowym

Źródło: opracowanie własne

### 5.6.5. Inteligencja płynna a osiągnięcia szkolne w zakresie świadomości językowej

W tabeli 5.7 przedstawiono szczegółowe wyniki analiz, w których zmienną zależną stanowią osiągnięcia szkolne uczniów z zakresu świadomości językowej.

**Tabela 5.7. Wyznaczniki osiągnięć z zakresu świadomości językowej: znaczenie inteligencji płynnej i zróżnicowanie międzyoddziałowe i międzyszkolne w zakresie siły zależności osiągnięcia – inteligencja płynna przy kontroli czynników socjoekonomicznych. Wyniki trzypoziomowych analiz regresji, odporne (*robust*) oszacowania błędów standardowych**

zmienna zależna: osiągnięcia szkolne w zakresie świad. jęz.	model (1)	model (2)	model (3)	model (4)
<b>oszacowanie efektów stałych</b>				
<b>poziom ucznia</b>				
stała	98,95 (0,570)	<b>97,88</b> (0,463)	<b>97,84</b> (0,459)	<b>97,84</b> (0,720)
pleć dziecka <sup>a</sup>		<b>5,17</b> (0,373)	<b>5,17</b> (0,370)	<b>5,17</b> (0,442)
wiek dziecka w tygodniach		<b>0,068</b> (0,014)	<b>0,068</b> (0,014)	<b>0,068</b> (0,013)
opóźniony tok nauki		<b>-12,88</b> (1,943)	<b>-12,77</b> (1,914)	<b>-12,82</b> (1,690)
przyspieszony tok nauki		<b>3,71</b> (1,661)	<b>4,08</b> (1,599)	<b>4,03</b> (1,906)
wykształcenie rodziców w latach nauki		<b>0,906</b> (0,125)	<b>0,907</b> (0,124)	<b>0,909</b> (0,681)
HISEI		<b>0,042</b> (0,016)	<b>0,041</b> (0,016)	<b>0,041</b> (0,016)
indeks zasobów materialnych rodziny dziecka		<b>0,122</b> (0,017)	<b>0,123</b> (0,017)	<b>0,123</b> (0,020)
inteligencja (TMS-K)		<b>0,441</b> (0,017)	<b>0,444</b> (0,017)	<b>0,444</b> (0,016)
<b>oszacowanie efektów losowych</b>				
wariancja efektów dla stałych dla szkół	<b>18,02</b>	<b>8,41</b>	<b>8,18</b>	<b>8,19</b>
wariancja nachyleń dla szkół dla inteligencji			<b>0,0026</b>	<b>0,0051</b>
wariancja efektów dla stałych dla oddziałów klasowych	<b>15,14</b>	<b>6,58</b>	<b>6,65</b>	<b>6,65</b>
wariancja nachyleń dla oddziałów klasowych dla inteligencji			0,0044	
wariancja na poziomie ucznia	<b>191,72</b>	<b>114,79</b>	<b>113,38</b>	<b>113,68</b>

liczba szkół = 172, liczba oddziałów = 300, liczba uczniów = 4739

Pogrubionym drukiem zaznaczono wartości istotne na poziomie istotności  $p < 0,05$ . W nawiasach podano błędy standardowe.

a – grupa odniesienia: chłopcy.

Źródło: opracowanie własne

Oszacowania wariancji nachyleń dla oddziałów klasowych w zakresie inteligencji w modelu (3), testującym postawione hipotezy, wskazują, iż siła zależności między poziomem



## CZY SZKOŁA MA ZNACZENIE?

osiągnięć dziecka w zakresie świadomości językowej i jego poziomem inteligencji płynnej różni się między szkołami ( $p < 0,001$ ), jednak wewnątrz szkół nie różni się między klasami, ( $p > 0,5$ ). Oszacowanie wariancji siły badanej zależności w szkołach wyniosło w modelu (4) (z wykluczonym parametrem wariancji nachyleń dla inteligencji dla oddziałów klasowych) 0,0051, przy wartości 0,44 dla współczynnika regresji dla inteligencji. Wraz ze wzrostem wyniku ucznia w Teście Matryc Ravena o 1 punkt, wyniki w teście świadomości językowej rosną przeciętnie o 0,44 punktu, jednakże w zależności od szkoły, do której dziecko uczęszcza, siła tej zależności różni się. W 95% szkół wraz ze wzrostem poziomu inteligencji o 1 punkt, poziom osiągnięć szkolnych dziecka rośnie od 0,37 do 0,52 punktu (gdy wartość pozostałych zmiennych pozostanie na niezmiennym poziomie).

Weryfikacja poprawy dopasowania kolejnych modeli, przeprowadzona zgodnie z wcześniej stosowanym schematem, nie pozwala jednak przyjąć modelu (4) jako ostatecznego. Dane na temat testowania poprawy dopasowania modeli znajdują się w tabeli 5.8. Model (4) nie jest istotnie lepiej dopasowany do danych niż model kontrolny w każdym z powtórzeń. Pomimo istotności parametru nachylenia dla inteligencji na poziomie szkół, reguły wnioskowania statystycznego wskazują, że należy go odrzucić na rzecz oszczędniejszego modelu (2). Jednak fenomen międzyszkolnego zróżnicowania siły korelacji inteligencja – osiągnięcia znalazł potwierdzenie dla trzeciego obszaru osiągnięć szkolnych. Z pewnością warto w przyszłych badaniach kontynuować eksplorację tego zjawiska.

**Tabela 5.8. Weryfikacja poprawy dopasowania modeli. Wyniki testowania istotności poprawy dopasowania modeli bardziej złożonych w porównaniu z oszczędniejszymi**

	model (2) k = 12
	(PV1) $\chi^2 = 1,27$
	(PV2) $\chi^2 = 9,60$
model (4)	(PV3) $\chi^2 = 9,10$
k = 14	(PV4) $\chi^2 = 12,16$
	(PV5) $\chi^2 = 17,34$
	df = 2

Pogrubionym drukiem zaznaczono wartości istotne na poziomie  $p < 0,05$ .

k – liczba parametrów modelu

df – liczba stopni swobody

(PV1) – (PV5) – porównanie dla powtórzeń modeli, w którym zmienną zależną jest PV o danym numerze porządkowym

Źródło: opracowanie własne

## 5.7. Podsumowanie

Celem analiz było znalezienie odpowiedzi na pytanie o siłę zależności pomiędzy inteligencją płynną a osiągnięciami szkolnymi uczniów w zakresie matematyki, czytania i świadomości językowej. Badano także zróżnicowanie siły tej zależności między szkołami i klasami.

Wyniki potwierdziły istnienie zależności pomiędzy poziomem inteligencji płynnej ucznia a jego osiągnięciami szkolnymi w trzech badanych obszarach: matematyce, czytaniu i świadomości językowej. Gdy wynik w Teście Matrycy Ravena rośnie o 1 punkt, wynik ucznia w testach rośnie o odpowiednio: 0,54, 0,40 i 0,44 punktu. Silniejsza siła zależności w przypadku matematyki jest zgodna z wynikami innych badań (np. Ferrer i McArdle, 2004; Heaven i Ciarrochi, 2012; Teo i in., 1996).

Ponadto analizy wykazały, że nie ma istotnych różnic między oddziałami klasowymi w zakresie siły zależności osiągnięcia – inteligencja. Pojawiają się one na poziomie szkół w każdym z trzech analizowanych obszarów wiedzy, jednak modele zawierające te efekty nie pozwalają lepiej statystycznie wyjaśniać osiągnięć uczniów w porównaniu do modeli ich nieuwzględniających. Mimo braku przewagi predykcyjnej modeli z losowymi nachyleniami dla inteligencji hipoteza międzyszkolnego zróżnicowania związku inteligencji i osiągnięć szkolnych warta jest dalszych analiz.

Brak zróżnicowania siły zależności inteligencja – osiągnięcia zarówno na poziomie oddziałów jest wynikiem zastanawiającym i wymaga interpretacji. Wydaje się, że procesy bezpośrednio powiązane ze zdobywaniem wiedzy, a więc zachodzące w klasie i w głównej mierze zależne od nauczyciela, choć częściowo powinny odpowiadać za siłę omawianej zależności. Uzyskane wyniki wskazują jednak, iż statystycznie znaczące zróżnicowanie na poziomie klas pod tym względem nie występuje. Może być to z jednej strony wynikiem spójności systemu szkolnego na poziomie nauczania początkowego, mogącym manifestować się w obszarach takich jak posługiwanie się tymi samymi podręcznikami we wszystkich oddziałach na konkretnym poziomie nauczania, korzystanie z tych samych materiałów przez nauczycieli, rozwój kompetencji wszystkich nauczycieli, ujednolicone metody uczenia i sposoby pracy z uczniami, dobra współpraca między nauczycielami nauczania początkowego, a także prowadzenie zajęć wyrównawczych i/lub pozalekcyjnych dla uczniów z różnych oddziałów.

Z drugiej strony brak zróżnicowania międzyoddziałowego może być częściowo pochodną schematu doboru próby. Został on opisany w rozdziale 1, jednak przypomnijmy, iż w ok. 1/4 szkół badano jeden oddział klasowy (gdyż była to jedyna klasa w szkole), natomiast w pozostałych szkołach, co najmniej dwuoddziałowych, badano 2 klasy (losując je, gdy w szkole funkcjonowały więcej niż 2 oddziały). Stąd dane mogą nie w pełni odzwierciedlać faktyczne zróżnicowanie międzyoddziałowe. Zatem przyczyną braku istotnego statystycznie zróżnicowania nachyleń może być niewystarczające moc testów na istotność efektu losowego dla zróżnicowania siły zależności inteligencja – osiągnięcia na poziomie klas. Problem ten wymaga więc dalszych badań.

## CZY SZKOŁA MA ZNACZENIE?

Otrzymane wyniki pozwalają wyciągnąć wnioski odnośnie siły zależności między inteligencją płynną a osiągnięciami szkolnymi, przy czym warto przypomnieć o konieczności zachowania ostrożności w wyciąganiu kategoriycznych wniosków o charakterze przyczynowym. Potrzeba ostrożności wynika z korelacyjnego charakteru badania, który nie pozwala na wykazanie istnienia tego rodzaju zależności. Poza tym efekt wynikać może nie z wpływu samej inteligencji, lecz np. wsparcia, zachęt czy możliwości oferowanych osobom spostrzegającym jako bardziej inteligentne przez ludzi i instytucje, które inteligencję cenią (Sternberg i in., 2001). Opisywane zależności traktowano w kategoriach przyczynowo-skutkowych ze względu na przyjęcie w pracy założenia teoretycznego o wpływie inteligencji na osiągnięcia szkole.

### Literatura cytowana

- de Ayala, R. J. (2009). *Theory and practice of Item Response Theory*. New York, NY: Guilford Press.
- Adey, P., Csapo, B., Demetrieou, A., Hautamaki i J., Shayer, M. (2007). Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review*, 2, 75–97.
- Blackwell, L. S., Trzesniewski, K. H. i Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246–263.
- Bossaert, G., Doument, 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, 47–57.
- Bull, R. i Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching and working memory. *Developmental Neuropsychology*, 19, 273–293.
- Cheng, H. i Furnham, A. (2012). Childhood cognitive ability, education, and personality traits predict attainment in adult occupational prestige over 17 years. *Journal of Vocational Behavior*, 81, 218–226.
- Cowan, N. i Alloway, T. (2009). Development of working memory in childhood. [W:] M. L. Courage, N. Cowan (Eds.), *The development of memory in infancy and childhood*. Hove, East Sussex, UK: Psychology Press, 303–342.
- Deary, I. J., Strand, S., Smith, P. i Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21.
- Der, G., Batty, G. D. i Deary, I. J. (2009). The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth. *Intelligence*, 39(6), 537–580.

- Dolata, R. (2008). *Szkoła, segregację, nierówności*. Warszawa: UW.
- Dweck, C. S. i Grant, H. (2008). Self-Theories, goals and meaning [in:] J. Y. Shah, W. L. Gardner (Eds.) *Handbook of motivation science*. New York: The Guilford Press.
- Dweck, C. S., Mangels, J. A. i Good, C. (2004). Motivational effects on attention, cognition, and performance [in:] D. Y. Dai, R. J. Sternberg (Eds.). *Motivation, emotion, and cognition. Integrative perspectives on intellectual functioning and development*. Mahwah, NJ: Lawrence Erlbaum, 41–55.
- Dweck, C. S. i Molden, D. C. (2005). Self-Theories. Their impact on competence motivation and acquisition [in:] A. J. Elliot, C. S. Dweck (Eds.) *Handbook of competence and motivation*. New York: Guilford Press, 122–140.
- Engle, R. W., Sędek, G., von Hecker, U. i McIntosh, D. N. (2004). *Cognitive limitations in aging and psychopathology: Attention, working memory, and executive functions*. Cambridge, UK: Cambridge University Press.
- 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.
- 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, 49–66.
- Galloway, C. (1988). *Psychologia uczenia się i nauczania*. Warszawa: PWN.
- Glas, C. A. (2010). *Preliminary manual of the software program Multidimensional Response Theory (MIRT)*. Pobrano z: [http://www.utwente.nl/gw/omd/afdeling/temp\\_test/mirt-manual.pdf](http://www.utwente.nl/gw/omd/afdeling/temp_test/mirt-manual.pdf).
- Graziano, P. A., Reavis, R. D., Keane, S. P. i Calkins, S. D. (2007). The role of emotion regulation in children's' early academic success. *Journal of School Psychology*, 45, 3–19.
- Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T. i Plomin, R. (2009). School achievement is predicted by self-perceived abilities – but for genetic rather environmental reasons. *Psychological Science*, 20(6), 753–762.
- Heaven, P. L. i Ciarrochi, J. (2012). When IQ is not everything: Intelligence, personality and academic performance at school. *Personality and Individual Differences*, 53, 518–522.
- 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, 368–375.

## CZY SZKOŁA MA ZNACZENIE?

- Jaworowska, A. i Szustrowa, T. (2000). *Test Matryc Ravena w wersji Standard. Formy: Klasyczna, Równoległa, Plus. Polskie standaryzacje*. Wydanie drugie. Warszawa: Pracownia Testów PTP.
- Klingberg, T., Fernell, E., Olsen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C. G., Forssberg, H. i Westerberg, H. (2005). Computerized training of working memory in children with ADHD – a randomized, controlled trial. *Journal of American Academy of Child and Adolescent Psychiatry*, 44(2), 177–186.
- Lachowicz-Tabaczek, K. (2004). *Potoczne koncepcje świata i natury ludzkiej*. Gdańsk: GWP.
- Lin, P. C. i Humphreys, L. G. (1977). Predictions of academic performance in graduate and Professional school. *Applied Psychological Measurement*, 1, 249–257.
- Morrison, E. F., Rimm-Kauffman, S. i Pianta, R. (2003). A longitudinal study of mother-child interactions at school entry and social and academic outcomes in middle school. *Journal of School Psychology*, 41, 185–200.
- Nęcka, E. (2003). *Inteligencja: geneza, struktura, funkcje*. Gdańsk: GWP.
- Nęcka, E. (2009). Inteligencja jest procesem [w:] J. Koziński (red.). *Nowe idee w psychologii*. Gdańsk: GWP.
- Niesser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. i Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 26(2), 77–101.
- Ollfors, M. i Andersson, S. I. (2007). Ability of stress, sense of control, and self-theories to predict Swedish high school students' final grades. *Educational Research and Evaluation*, 13(2), 143–169.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Rescorla, L. i Rosenthal, A. S. (2004). Growth in standardized ability and achievement test scores from 3rd to 10th grade. *Journal of Educational Psychology*, 96(1), 85–96.
- Reuhkala, M. (2001). Mathematical Skills in Ninth-graders: Relationship with visuospatial abilities and working memory. *Educational Psychology*, 21, 387–399.
- Robins, R. W. i Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity*, Vol 1, 313–336.
- Schatz, J., Kramer, J. H., Albin, A. i Matthey, K. K. (2000). Processing speed, working memory, and IQ. *Neuropsychology*, 14, 189–200.
- Sękowski, A. (2001). *Osiągnięcia uczniów zdolnych*. Lublin: Wydawnictwo Towarzystwa Naukowego KUL.

- Sternberg, R. (2003). *Wisdom, intelligence and creativity synthesized*. Cambridge: Cambridge University Press.
- Sternberg, R., Giorenko, E. L. i Bundy, D. A. (2001). The predictive power of IQ. *Merrill-Palmer Quarterly, Vol. 47(1)*, 1–41.
- Sternberg, R. i Spear-Swerling, L. (2003). *Jak nauczyć dzieci myślenia*. Gdańsk: GWP.
- Strelau, J. i Zawadzki, B. (2008). Psychologia różnic indywidualnych. W: J. Strelau i D. Doliński (red.). *Psychologia. Podręcznik akademicki, tom 1 (765–846)*, Gdańsk: GWP.
- Teo, A., Carlson, E., Mathieu, P. J., Egeland, B. i Sroufe, A. L. (1996). *Journal of School Psychology, Vol. 34(3)*, 285–306.
- Verhelst, N. (2009). Probabilistyczna teoria wyniku zadania. [W:] H. Szaleniec (red.). Teoria wyniku zadania IRT. *Zastosowania w polskim systemie egzaminów zewnętrznych*. Warszawa: CKE.
- Von Davier, M., Gonzalez, E. i Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series, 2*, 9–36.
- Watkins, M. W., Lei, P.-W. i Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence, 35*, 59–68.
- Welsh, M. C. i Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology, 4*, 199–230.
- Wills, J. O., Dumont i R. Kaufman, A. S. (2011). Factor-analytic model of intelligence. [in:] R. J. Sternberg i S. B. Kaufman (Eds.) *The Cambridge Handbook of Intelligence*. New York, NY: Cambridge University Press.
- Włodarski, Z. i Matczak, A. (1996). *Wprowadzenie do psychologii*. Warszawa: WSiP.
- Wyrwicz, N. (2005). Zdolności poznawcze jako predyktor osiągnięć szkolnych uczniów. *Psychologia. Edukacja i Społeczeństwo, 2*, 157–169.



## Rozdział 6

### Lęk a wyniki nauczania

Znaczenie indywidualnych charakterystyk uczniów rozpatrywane w kontekście ich osiągnięć szkolnych, kieruje naszą uwagę na te z nich, które samodzielnie lub wchodząc w interakcje z charakterystykami szkoły i klasy, przyczyniają się do polepszenia lub pogorszenia osiągniętych wyników. W wcześniejszym rozdziale braliśmy pod uwagę znaczenie inteligencji, która wyznacza zakres możliwości poznawczych dziecka. Obraz ten nie byłby jednak pełen bez uwzględnienia aspektu emocjonalnego. Znaczenie przeżywanych emocji, ich rola regulacyjna i motywacyjna podkreślana jest przez wiele współczesnych koncepcji, np. inteligencji emocjonalnej Daniela Golemana (1997) czy rozwoju kompetencji emocjonalnych (Salovey, Sluyter, 1999).

System emocjonalny jest głównym czynnikiem motywacyjnym, Izard i Ackerman (2005) twierdzą, że każda z emocji pełni rolę w organizowaniu zachowania ludzi poprzez wpływ na procesy spostrzegania, rozumowania oraz bezpośrednio na podejmowanie czynności. Ważne dla poruszanego przez nas problemu jest spostrzeżenie, że relacje między emocjami a zachowaniem mają swoje początki na wczesnych etapach życia i pozostają stabilne przez długi czas. Zgodnie z rozwojowym ujęciem emocji zmienia się sposób ich ekspresji, jednak związek pomiędzy wzbudzoną emocją a reakcjami behawioralnymi pozostaje w miarę trwały.

Najszerzej rzecz ujmując, poruszany przez nas problem dotyczy związków pomiędzy charakterystykami emocjonalnymi a funkcjonowaniem szkolnym. Można postawić pytanie, w jaki sposób przeżywane emocje wiążą się z podejmowaniem określonych działań lub przeciwnie – prowadzą do zaprzestania tych działań, np. w postaci wycofania się. Jak zauważają Izard i Ackerman (2005, s. 331)

„Stany uczuciowo-motywacyjne zawierają pewną informację i stanowią źródło sygnałów do podejmowania decyzji oraz działania. Można to rozumieć jako tendencje do spostrzegania i myślenia zgodnie z informacją zawartą w emocji. Chociaż odczucie emocji różni się jakościowo od procesów myślowych i decyzyjnych, to generując sygnały z reguły mobilizuje ono – gwałtownie i automatycznie – system poznawczy. Integracja, czy też skoordynowana interakcja emocji z odpowiednim poznaniem, prowadzi do zachowania adaptacyjnego”.



## CZY SZKOŁA MA ZNACZENIE?

Część emocji, takich jak na przykład złość czy strach, uruchamia tendencje do działania sprzyjające dość konkretnym celom takim jak atak lub obrona (Izard, Ackerman, 2005, s. 331). Inne emocje mogą powodować ogólną aktywację organizmu i jego gotowość do odpowiadania na napływające informacje, ale nie są powiązane z żadnym konkretnym typem reakcji behawioralnej.

Problem emocji i ich znaczenie w procesach regulacji zachowania można rozpatrywać również przez pryzmat motywacji, która wiąże się z przeżywaniem określonych stanów emocjonalnych. Według E. Deciego (1992) zainteresowanie i radość to połączenie emocji, które jest typowe dla wewnętrznej motywacji. Niezbyt mocny lęk może występować wraz z poczuciem bycia czymś zainteresowanym lub zaciekawionym. W takiej sytuacji możemy obserwować, jak te dwie emocje układają się w charakterystyczny wzorec dążenia–unikania. Jednostka zbliża się do jakiegoś obiektu/osoby motywowana zainteresowaniem/ciekawością, a z drugiej strony unika tej nowej sytuacji, co motywowane jest lękiem. Natomiast nasilony lęk może przyczynić się do zmniejszania się ciekawości i w konsekwencji rezygnacji z podejmowania działania.

Z punktu widzenia analizy uwarunkowań wyników nauczania najbardziej interesującą emocją jest lęk. Zadajemy pytanie o związki pomiędzy poziomem lęku traktowanym przez nas jako indywidualna charakterystyka uczniów a poziomem osiągnięć szkolnych w zakresie umiejętności matematycznych, czytania i świadomości językowej.

### 6.1. Koncepcje i rodzaje lęku

Lęk możemy zdefiniować jako przewidywanie przyszłych zagrożeń lub nieszczęść, któremu towarzyszą uczucia dysforyczne lub somatyczne objawy napięcia. Strach różni się od lęku przede wszystkim tym, że jest wywoływany przez rozpoznawalny bodziec. Zatem lęk jest przedbodźcowy (antycypowany w stosunku do bodźca zagrażającego), a strach jest pobodźcowy (wywołany określonym zdarzeniem); (Öhman 2005 s. 720). Według Epsteina (1972 za Öhman, 2005 s. 720).

„...strach stanowi motyw unikania. Gdyby nie istniały żadne ograniczenia (wewnętrzne lub zewnętrzne), strach prowadziłby do działania jakim jest ucieczka. Lęk można zdefiniować jako nierozwiązany strach, czy też stan nieukierunkowanego pobudzenia następującego po spostrzeżeniu zagrożenia”.

Jak zauważa J. Kagan (2008 s. 26) „lęk we współczesnym świecie jest większym ciężarem niż złość, jest on mniej przystosowawczy w świecie, w którym podejmowanie ryzyka

i poznawanie ludzi jest często niezbędne do osiągnięcia sukcesu zawodowego”. Zatem nadmiernie nasilone przeżywanie lęku może blokować zachowania niezbędne do radzenia sobie z wyzwaniami szkolnymi.

Lęk może być spostrzegany zarówno jako stan emocjonalny, który jest wywoływany w konkretnym kontekście i ma ograniczony czas trwania, jak i jako cecha osobowości, typowa dla jednostki niezależnie od momentu i sytuacji (Öhman 2005 s. 721).

Bardzo ogólnie objawy lęku można podzielić na te związane z somatyczną hiperaktywnością (pocenie się, rumienienie, płytki oddech, kołatanie serca, dyskomfort w jamie brzusznej) oraz te związane z lękiem poznawczym bądź psychicznym (natrętne i niechciane myśli, martwienie się, ruminacje, niepokój i czasami uczucie nadmiernego napięcia emocjonalnego).

Począwszy od pracy Cattela (1966, za Tucholska i Steuden, 1990), który wyodrębnił lęk jako trwałą dyspozycję osobowościową i lęk jako aktualnie wzbudzony stan, posługiwano się rozróżnieniem na lęk-cechę i lęk-stan. To rozróżnienie zostało zaadaptowane i wykorzystane przez Charlesa Spilberga (por. Papay, Costello, Hedl, Speilberger, 1975) do stworzenia koncepcji lęku i odpowiadających tej koncepcji narzędzi pomiarowych. Kluczowe rozróżnienie na dwa rodzaje lęku związane jest ze spostrzeżeniem, że różnice w nasileniu obserwowanych lub relacjonowanych przez badanych objawów możemy odnosić zarówno do cech sytuacji, która może być mniej lub bardziej lękotwórcza, jak i do cech osoby. Poniżej przedstawiono zestawienie charakterystyk obu ujęć lęku.

**Tabela 6.1. Lęk jako cecha i jako stan – definicje**

lęk jako cecha	lęk jako stan
<ul style="list-style-type: none"> <li>■ jest to dyspozycja osoby do reagowania w określony sposób</li> <li>■ jeśli sytuacja zostanie zinterpretowana jako trudna/zagrażająca, mogą pojawić się objawy lęku</li> <li>■ ten rodzaj lęku czyni jednostkę podatną na spostrzeżenie szerokiego zakresu sytuacji jako zagrażających; taka ocena sytuacji może prowadzić do reagowania stanami lęku nieproporcjonalnymi do faktycznego zagrożenia</li> <li>■ jest nabywany w procesie wychowania</li> <li>■ istotne są doświadczenia z okresu dzieciństwa szczególnie te z osobami znaczącymi</li> <li>■ ważne dla kształtowania się tego typu lęku jest doświadczenie zagrożenia utratą aprobaty i miłości rodzicielskiej</li> </ul>	<ul style="list-style-type: none"> <li>■ jest określonym sposobem reagowania na sytuację trudną</li> <li>■ wiąże się z przeżywaniem obawy i napięcia</li> <li>■ objawom psychicznym towarzyszy aktywacja autonomicznego układu nerwowego</li> <li>■ charakterystyczne dla tego typu lęku jest to, że zmienia się pod wpływem czynników zewnętrznych (z sytuacji na sytuację)</li> </ul>

Źródło: opracowanie na podstawie Tucholska i Steuden (1990)

## CZY SZKOŁA MA ZNACZENIE?

W przyjętej koncepcji nasilenie lęku-cechy wyznacza zakres i intensywność reakcji na sytuację, które zostaną zinterpretowane jako mogące nieść zagrożenie dla Ja. Osoby z wysokim poziomem lęku-cechy będą miały zatem również większą tendencję do reagowania lękiem w danej sytuacji, a co za tym idzie, będą ujawniać również wyższy poziom lęku jako stanu. Dokonywanie pomiaru lęku kwestionariuszem STAIC (polska adaptacja: Inwentarz stanu i cechy lęku dla dzieci; Jaworowska, 2005) daje możliwość diagnozy obu typów lęku. Należy zastanowić się, czy i w jakim stopniu wyniki uzyskiwane przez badane dzieci w zakresie lęku-stanu są pochodną lękowej interpretacji sytuacji, co byłoby związane z wysokim poziomem lęku-cechy. Na ile natomiast stan lęku został wzbudzony przez „obiektywne” właściwości sytuacji związanej z pomiarem. Problem ten zostanie poruszony w części poświęconej analizie wyników badań.

Ważnych informacji na temat uwarunkowań obu typów lęku dostarczają badania z zakresu genetyki zachowania. Badania przeprowadzone z udziałem 1058 bliźniąt w wieku pomiędzy 8–16 lat wykazały, że lęk jako stan pozostawał pod znaczącym wpływem czynników środowiskowych zarówno w grupie kobiet, jak i w grupie mężczyzn. Wpływ środowiska specyficznego w grupie mężczyzn tłumaczył 84% zmienności wyników, a w grupie kobiet 64%. Przy czym w grupie kobiet środowisko wspólne wywierało umiarkowany wpływ na poziom lęku i czynnik ten tłumaczył 36% zmienności wyników. Kiedy natomiast spojrzymy na lęk ujmowany jako cecha, pozostawał on pod umiarkowanym wpływem czynników genetycznych (31% wyjaśnianej zmienności) i specyficznego środowiska (54% wyjaśnianej zmienności). Środowisko wspólne wyjaśniało około 15% zmienności. W odniesieniu do lęku jako cechy nie zaobserwowano znaczących różnic pomiędzy kobietami i mężczyznami (Lau, Eley i Stevenson 2006).

W odniesieniu do sytuacji szkolnej, jak zauważają J. Papay i Ch. Spilberger (1986), dzieci z wysokim poziomem lęku-cechy częściej doświadczają stanu lęku, który negatywnie wpływa na ich osiągnięcia. Badacze przypuszczają również, że doświadczenia szkolne mogą wzbudzać lęk sytuacyjny (lęk-stan), który, jeśli staje się chroniczny, może przyczyniać się do podwyższenia poziomu lęku-cechy. Jednak te stwierdzenia muszą być traktowane bardzo ostrożnie ze względu na jednorazowy charakter dokonywanych pomiarów, który nie pozwala kontrolować poziomu lęku przed rozpoczęciem szkoły i jego zmian w trakcie nauki.

## 6.2. Lęk a osiągnięcie szkolne

Jak wskazuje Cz. Czabała i in. (2005) zaburzenia lękowe stanowią jeden z najbardziej rozpowszechnionych typów problemów psychicznych doświadczanych przez uczniów. Ich nasilenie w populacji wzrasta w przypadku uczniów starszych (gimnazjum i szkoła średnia). W przypadku młodzieży pomiędzy 12 a 19 rokiem życia badanej przez Wolańczyka (2002) 17% zgłaszało różnorodne problemy emocjonalne i behawioralne, w tym między innymi związane z przeżywaniem nadmiernego lęku. W niniejszym rozdziale nie poruszamy problemu zaburzeń lękowych, jednak warto zwrócić uwagę, że bardzo duże nasilenie lęku-cechy może wiązać się z występowaniem zaburzeń lękowych. Szczególnie dotyczy to tzw. zespołu uogólnionego lęku. P. Kendall pisze (2004 s. 101–102):

„Cechą charakterystyczną tego zaburzenia jest nierealistyczny lęk i niepokój, które nie są powiązane z żadną konkretną sytuacją czy zewnętrznym stresem. Dzieci z zespołem uogólnionego lęku martwią się, jak będą cenione w przyszłości, niepokoją je społeczne kontakty, aktywność rodzinna, sport, kwestie zdrowia, i martwią się po prostu, co się może zdarzyć jutro, a nawet za godzinę. Ten chronicznie utrzymujący się lęk jest określany jako „lękowe pogotowie”, czyli stała gotowość do reagowania na sygnały w kategorii zagrożeń”.

Dzieci cierpiące na ten rodzaj zaburzenia mogą doświadczać różnorodnych kłopotów w codziennym funkcjonowaniu w szkole, zarówno w obszarze realizacji zadań związanych z nauką, jak i w obszarze kontaktów z rówieśnikami i dorosłymi.

Prześledzenie wyników badań dotyczących lęku i poziomu osiągnięć szkolnych dostarcza argumentów na temat słuszności włączenia do programu naszych badań tej zmiennej. Obserwowane związki pomiędzy natężeniem lęku mierzonym za pomocą kwestionariuszy wypełnianych przez dzieci i ich rodziców oraz nauczycieli wskazuje na umiarkowanie negatywne zależności z różnymi wskaźnikami osiągnięć, np. standardowymi testami czy ocenami szkolnymi. Wyniki tych badań zostały przedstawione w tabeli 6.2. Należy zwrócić uwagę, że w przypadku części badań istotne związki lęku z osiągnięciami dotyczyły tylko dzieci, u których zaobserwowano wysoki poziom lęku lub u których już wcześniej zdiagnozowano zaburzenia lękowe.

**Tabela 6.2. Przegląd badań na związkami lęku z osiągnięciami szkolnymi**

Autorzy badania	Badana grupa	Zastosowane narzędzia	Najważniejsze wyniki
Milgram, R. M., Milgram, N. A. (1977). The effect of test content and context on the anxiety-intelligence relationship. <i>The Journal of Genetic Psychology</i> , 130, 121–127.	177 uczniów (99 chłopców i 78 dziewcząt) z klas 4–6	<ul style="list-style-type: none"> <li>■ tysunkowy test rozumienia humoru</li> <li>■ Test Matrycy Ravena</li> <li>■ werbalny test inteligencji Milta</li> <li>■ test osiągnięć szkolnych (czytanie, geografia, arytmetyka)</li> <li>■ General Anxiety Test Scale</li> </ul>	<ul style="list-style-type: none"> <li>■ brak korelacji uogólnionego lęku z osiągnięciami (zarówno u chłopców, jak i u dziewcząt)</li> <li>■ wśród dziewcząt ujemne korelacje lęku z testem rozumienia humoru, Testem Matrycy Ravena i werbalnym testem inteligencji (odpowiednio: -0,35; -0,52; -0,35)</li> </ul>
McCann, S. J., Meen, K. S. (1984). Anxiety, ability, and academic achievement. <i>Journal of Social Psychology</i> , 124, 257–258.	222 uczniów klas 11 i 12	<ul style="list-style-type: none"> <li>■ Test Matrycy Ravena</li> <li>■ STAI</li> <li>■ średnia ocen z języka ojczystego</li> </ul>	<p>W całej próbie:</p> <ul style="list-style-type: none"> <li>■ lęk i zdolności skorelowane na poziomie -0,16</li> <li>■ zdolności i osiągnięcia skorelowane (0,48)</li> <li>■ osiągnięcia i lęk nieskorelowane</li> </ul> <p>W próbie podzielonej w punkcie mediany ze względu na zdolności:</p> <ul style="list-style-type: none"> <li>■ w dolnej grupie korelacja lęku z osiągnięciami: -0,16</li> <li>■ w górnej grupie korelacja lęku z osiągnięciami: 0,2</li> </ul>
Ialongo, N., Edelson, G., Werthamer-Larsson, L., Crockett, L., Kellam, S. (1995). The significance of self-reported anxiety symptoms in first grade children: Prediction to anxious symptoms and adaptive functioning in fifth grade. <i>Journal of Child Psychology and Psychiatry</i> , 36, 427–437.	<ul style="list-style-type: none"> <li>■ Zbadano 1197 dzieci w pierwszej klasie</li> <li>■ w wieku 7 lat</li> <li>■ Dokonano dwóch pomiarów lęku w odstępie 4 miesięcy</li> </ul>	<ul style="list-style-type: none"> <li>■ Revised Children's Manifest Anxiety Scale kwestionariusz przeznaczony do samodzielnego wypełniania dla dzieci w wieku od 6 do 19 lat</li> <li>■ The California Achievement Test (CAT; Forms E &amp; F).</li> </ul>	<ul style="list-style-type: none"> <li>■ dzieci, które znalazły się w najwyższym kwartylu pod względem poziomu lęku (pomiar jesienią), miały najwyższe prawdopodobieństwo znalezienia się w najniższym kwartylu pod względem umiejętności czytania (pomiar na wiosnę)</li> <li>■ związki te były istotne zarówno w grupie chłopców, jak i dziewcząt</li> <li>■ nie obserwowano natomiast istotnych statystycznie związków między poziomem lęku mierzonego w trakcie pierwszego pomiaru (jesień) a wynikami w zakresie umiejętności matematycznych mierzonych wiosną</li> </ul>
Newbegin, I., Owens, A. (1996). Self esteem and anxiety in secondary school achievement. <i>Journal of Social Behavior and Personality</i> , 11(3), 521–530.	chłopcy w wieku 12–17 lat (7–12 klasa)	<ul style="list-style-type: none"> <li>■ kwestionariusz lęku-cechy i leku-stanu</li> <li>■ lęku egzaminacyjnego (jako cechy)</li> <li>■ samoocena</li> <li>■ oceny na koniec semestru z języka ojczystego i matematyki</li> </ul>	<ul style="list-style-type: none"> <li>■ słaba negatywna korelacja poziomu lęku z oceną na koniec semestru z matematyki (-0,16)</li> <li>■ brak korelacji z oceną z języka ojczystego; zaobserwowana korelacja pomiędzy oceną z matematyki a lękiem przestała być istotna, gdy kontrolowano lęk testowy (rozumiany jako cecha)</li> <li>■ oba rodzaje lęku korelowały na poziomie 0,6</li> </ul>
Jaworowska, A. (2005). <i>Inwentarz Stanu i Cechy Lęku dla Dzieci (STAI-C) C. D. Spielberga, C. D. Edwardsa, R. E. Lushene'a, J. Monturioniego i D. Platzeck. Warszawa: PTP.</i>	<ul style="list-style-type: none"> <li>■ 391 dziewcząt i 354 chłopców z klas piątych i szóstych</li> <li>■ 508 dziewcząt i 476 chłopców z klasy pierwszej i drugiej gimnazjum</li> </ul>	<ul style="list-style-type: none"> <li>■ STAI</li> <li>■ oceny szkolne</li> </ul>	<ul style="list-style-type: none"> <li>■ ujemna korelacja lęku jako cechy z ocenami szkolnymi uczniów klas 5 i 6 SP; zależność ta obserwowana była przede wszystkim wśród chłopców</li> <li>■ wysoki poziom lęku skorelowany ujemnie z ocenami z historii (-0,16), matematyki (-0,13), biologii (-0,12) oraz ze średnią ocen (-0,12)</li> <li>■ w przypadku dziewcząt –korelacja była widoczna tylko w odniesieniu do ocen z matematyki (-0,11)</li> </ul>

<p>Grover, R., Ginsburg, G. &amp; Jalongo, N. (2007). Psychosocial outcomes of anxious first graders: a seven year follow-up. <i>Depression &amp; Anxiety</i>, 24, 410–420.</p>	<ul style="list-style-type: none"> <li>■ pomiar pierwszy: 149 uczniów klasy pierwszej w wieku od 5 do 8 lat (72 dziewczynki)</li> <li>■ Pomiar drugi: 129 uczniów klasy ósmej</li> </ul>	<ul style="list-style-type: none"> <li>■ narzędzia do oceny poziomu lęku wypełniali zarówno uczniowie jak i rodzice oraz nauczyciele</li> <li>■ Baltimore How I Feel—Young Child Version, Child Report</li> <li>■ Baltimore How I Feel—Young Child Version, Parents Report</li> <li>■ Shy Behaviour subscale pochodząca z narzędzia dla nauczycieli (TOSCA)</li> <li>■ Osiągnięcia szkolne: w klasie pierwszej – The fourth edition of the Comprehensive Test of Basic Skills; w klasie ósmej – Kaufman Test of Educational Achievement</li> </ul>	<ul style="list-style-type: none"> <li>■ najważniejszy wynik dotyczył związków pomiędzy poziomem lęku a osiągnięciami szkolnymi. Na podstawie przeprowadzonych procedur dla każdej ze skal osobno (samocena, ocena rodziców i nauczycieli) dokonano podziału dzieci na przejawiające symptomy lęku i dzieci nie lękowe</li> <li>■ istotne związki pomiędzy osiągnięciami zarówno w klasie pierwszej, jak i w klasie ósmej, zaobserwowano tylko w przypadku ocen dokonywanych przez nauczycieli</li> <li>■ dzieci uznane za lękowe miały istotnie większą szansę na znalezienie się w grupie mającej najniższy poziom wyników (matematyka i czytanie)</li> <li>■ analogiczne związki zaobserwowano w klasie ósmej</li> </ul>
<p>Mazzone, L., Ducci, F., Scoto, M. C., Passaniti, E., D'Arrigo, V. G., Vitello, B. (2007). The role of anxiety in a community sample of children and adolescents. <i>BioMed Central Public Health</i>, 7, 347–411.</p>	<ul style="list-style-type: none"> <li>■ zbadano dzieci w wieku: 8–10 lat (131), 11–13 (267), 14–16 (80)</li> </ul>	<ul style="list-style-type: none"> <li>■ Multidimensional Anxiety Scale for Children (MASC);</li> <li>■ wynik powyżej 65 świadczyły o występowaniu lęku</li> <li>■ oceny szkolne (skategoryzowane na 10 stopniowej skali, poniżej 6 uznawano je jako niewystarczające)</li> </ul>	<ul style="list-style-type: none"> <li>■ w całej zbadanej grupie 7,3% dzieci miało wyniki świadczące o występowaniu lęku</li> <li>■ w grupie tych dzieci 37,1% miało oceny szkolne na poziomie niewystarczającym, a 28,6% na poziomie dobrym lub bardzo dobrym</li> <li>■ w grupie nie lękowej oceny na poziomie dobrym lub bardzo dobrym miało 56,1% dzieci</li> </ul>
<p>Hughes, A. A., Lourea-Waddell, B., Kendall, P. C. (2008). Somatic complaints in children with anxiety disorders and their unique prediction of poorer academic performance. <i>Child Psychiatry and Human Development</i>, 39, 211–220.</p>	<ul style="list-style-type: none"> <li>■ 108 uczniów w wieku 8–14 lat w tym 69 spełniających kryteria diagnostyczne zaburzeń lękowych</li> </ul>	<ul style="list-style-type: none"> <li>■ Multidimensional Anxiety Scale for Children (MASC)</li> <li>■ Rodzice wypełniali The Child Behaviour Checklist (CBCL)</li> <li>■ Nauczyciele wypełniali The Teacher Report Form (diagnozujący m.in. osiągnięcia szkolne)</li> </ul>	<ul style="list-style-type: none"> <li>■ uczniowie z zaburzeniami lękowymi: przejawiali więcej objawów somatycznych</li> <li>■ uczniowie z zaburzeniami lękowymi mieli niższe osiągnięcia szkolne</li> <li>■ predyktorem osiągnięć nie był poziom lęku, lecz objawy somatyczne</li> <li>■ w przypadku kwestionariusza wypełnianego przez dziecko miara objawów somatycznych stanowiła element skali ogólnej lęku. Obie skale są skorelowane na poziomie 0,8</li> </ul>

Źródło: opracowanie na podstawie literatury wymienionej w tabeli

## CZY SZKOŁA MA ZNACZENIE?

Cześć badaczy wskazuje na odmienne wyniki, które nie potwierdzają związków pomiędzy lękiem a osiągnięciami. Vitaro i in. (2005) stwierdzili, że poziom lęku mierzony po raz pierwszy, kiedy badani byli w przedszkolu, nie był lepszym predyktorem nieukończenia przez nich szkoły średniej niż pozostałe czynniki ryzyka takie jak płeć, trudności w funkcjonowaniu społecznym i ekonomicznym rodziny. Również w przypadku studentów lęk nie stawał się ważniejszym predyktorem problemów w nauce w stosunku do innych czynników ryzyka (Strahan, 2003). Na obraz związków pomiędzy lękiem a osiągnięciami nakładają się dwa ważne problemy, które ograniczają możliwości jednoznacznego opisu tych zależności. Po pierwsze zaobserwowano, że zależności te mają charakter krzywoliniowy, np. nastolatki, którzy mają umiarkowany poziom lęku, odnoszą większe sukcesy szkolne niż ich rówieśnicy charakteryzujący się lękiem wysokim lub niskim (Duchesne i in., 2008). Po drugie, dostrzega się problem związany z faktem, że w większości badań stosuje się jeden pomiar lęku jako predyktor późniejszych osiągnięć, co nie pozwala na uwzględnienie zmian w intraindywidualnym poziomie lęku ani zmian związanych z procesami rozwojowymi np. nasileniem się poziomu lęku występującym u adolescentów (op. cit.).

Prześledzenie związków pomiędzy lękiem a poziomem osiągnięć szkolnych skłania do postawienia pytania o mechanizmy, które mogą być odpowiedzialne za taki obraz zależności. I. Krejtz zauważa (2012 s. 101):

„Badania nad funkcjonowaniem poznawczym osób lękowych wskazują, iż poczucie niepokoju może prowadzić do zawężania pola uwagi, ograniczając je do bodźców stanowiących źródło lęku, co utrudnia koncentrację na zadaniu. Zasoby poznawcze takich osób mogą być w dużym stopniu pochłonięte monitorowaniem własnych stanów, obserwowaniem reakcji otoczenia oraz koniecznością przełamywania negatywnych myśli na swój temat(...).”

Zgodnie z teorią wydajności przetwarzania (Owens i in., 2008) wysoki poziom lęku angażuje zasoby poznawcze, a powiązane z niepokojem myśli, które nie mają związku z wykonywanym zadaniem, zmniejszają wydolność pamięci roboczej. Zmniejszona wydolność zwiększa prawdopodobieństwo spadku wydajności przetwarzania i jednoczesnego wzrostu wysiłku lub czasu koniecznego do osiągnięcia typowego poziomu wykonania zadań poznawczych (zwłaszcza tych wymagających dużych zasobów pamięci roboczej). W badaniu przeprowadzonym przez Owensa (op. cit.) opisano istotne związki pomiędzy lękiem-cechą (mierzonym za pomocą STAIC) a osiągnięciami z matematyki. Związek ten był mediowany przez werbalną pamięć roboczą. Wysoki poziom lęku wiąże się z pogorszeniem funkcjonowania pamięci roboczej, co z kolei wiązało się z istotnie niższymi wynikami w matematycznej części testu

osiągnięć szkolnych. Nie zaobserwowano natomiast takich związków dla wyników w teście z języka ojczystego.

Charakterystyczne dla dzieci z wysokim poziomem lęku przetwarzanie informacji nadmiernie angażujące zasoby poznawcze, które są niezbędne do nauki, może powodować w dłuższej perspektywie obniżanie się wyników. Jednak zależności te mogą mieć również charakter dwustronny. Pogorszenie się wyników i trudności w nauce wzmagać będą lęk zarówno w konkretnej sytuacji, np. na danym przedmiocie, ale również jeśli stan taki utrzymuje się dłużej, może się on zgeneralizować na inne sytuacje szkolne i pozaszkolne (por. Mazzone i. in., 2007). Interesujące zależności opisano w odniesieniu do związków pomiędzy poziomem inteligencji a lękiem i poziomem osiągnięć. W badaniach H. Billa i in. (1991) wykazano, że w przypadku dziewcząt niska inteligencja na początku nauki szkolnej może być czynnikiem ryzyka rozwinięcia się zaburzeń lękowych (i silniejszego lęku w ogóle), gdyż powoduje gorsze radzenie sobie z zadaniami szkolnymi. Taka sytuacja na początku nauki jest bardziej typowa dla dziewczynek, od których generalnie oczekuje się lepszych wyników niż w przypadku chłopców. Zatem kłopoty z nauką powodują nasilenie się lęku, który z kolei wtórnie może przyczyniać się do obniżania się wyników szkolnych.

Nasilenie lęku ulega zmianom na przestrzeni późnego dzieciństwa i adolescencji. Młodszy uczeń przed wejściem w okres dorastania doświadcza niższego lęku i rzadziej diagnozuje się u nich zaburzenia lękowe niż w przypadku adolescentów. W odniesieniu do tej drugiej grupy nasilenie lęku wydaje się wiązać z przemianami okresu dojrzewania, np. wzmożoną tendencją do reagowania silnymi emocjami oraz labilnością emocjonalną charakterystyczną dla nastolatków (Mazzone, 2007).

Dla rozpatrywanych przez nas związków pomiędzy lękiem a osiągnięciami istotne jest również uwzględnienie innych zmiennych, które mogą znacząco wpływać na obserwowane zależności. W modelu opisywanym przez S. Duchense i in. (2008) dotyczącym przyczyn przedwczesnego wypadania z systemu edukacji, autorzy wskazują na dwie grupy zmiennych, które mogą pośredniczyć w związkach pomiędzy lękiem a nieukończeniem edukacji. Są to charakterystyki związane z uczniem, takie jak np. agresywność czy nadaktywność, oraz charakterystyki związane z rodziną dziecka, takie jak np. niski status społeczno-ekonomiczny, niskie wykształcenie matki, samotne rodzicielstwo. W literaturze pojawiają się także informacje o możliwej interakcji między lękiem jako cechą a inteligencją – w przypadku uczniów o niskim poziomie inteligencji lęk wiązać może się ze spadkiem poziomu osiągnięć, przy wysokim poziomie inteligencji lęk może prowadzić do wyższych osiągnięć (McCaan i Meen, 1984). Zatem w prowadzonych analizach zasadne jest uwzględnienie/kontrolowanie zarówno takich charakterystyk ucznia jak poziom inteligencji, jak i zmiennych opisujących szczegółowo jego sytuację rodzinną.



## CZY SZKOŁA MA ZNACZENIE?

Kolejna kwestia, na którą chcielibyśmy zwrócić uwagę, dotyczy odkrytej w wielu opisanych powyżej badaniach zależności pomiędzy silnym lękiem lub zaburzeniami lękowymi a niskim poziomem osiągnięć. Pragniemy podkreślić, że w klinicznych studiach dotyczących zaburzeń lękowych jest to silna i powtarzająca się zależność (por. Van Ameringen, 2003). Jednak w badaniach populacyjnych prowadzonych na grupach nieklinicznych wskazuje się na to, że umiarkowany poziom lęku może spełniać rolę mobilizatora i uruchamiać zasoby poznawcze i behawioralne, które są niezbędne do radzenia sobie z nauką w szkole a poprzez to przyczynić się do zwiększenia poziomu osiągnięć. A tylko bardzo wysoki lub bardzo niski poziom lęku zakłóca naukę poprzez dezorganizację działania (wysoki lęk) lub brak mobilizacji do działania (bardzo niski lęk, por. Duchense i in., 2008).

Lęk jest istotny w analizach osiągnięć szkolnych nie tylko ze względu na możliwość zakłócania procesu uczenia się, ale również z tego względu, że objawy lękowe utrudniają realizację ważnych zadań rozwojowych, takich jak nawiązywanie i podtrzymywanie relacji z rówieśnikami i nauczycielami (Grover i in., 2007). Nasilone przeżywanie lęku niezależnie od tego, jakie są jego przyczyny, może mieć długotrwałe konsekwencje w postaci mniejszych osiągnięć życiowych, np. mężczyźni mają mniej stabilny przebieg zatrudnienia i odnoszą mniej sukcesów zawodowych, kobiety natomiast mają mniejsze szanse na zakończenie edukacji na poziomie średnim (Caspi i in., 1988).

Należy podkreślić, że wzajemne interakcje i procesy sprzężeń zwrotnych pomiędzy lękiem a poziomem osiągnięć i codzienną praktyką szkolną powinny być brane pod uwagę nie tylko w kontekście poszukiwania uwarunkowań efektywności kształcenia, ale również w celu wskazywania na możliwości zapobiegania powstawaniu zaburzeń lękowych i niepowodzeń szkolnych.

### 6.3. Metoda

Na podstawie przeprowadzonej analizy literatury można przypuszczać, że istnieją związki pomiędzy poziomem doświadczanego przez dziecko lęku a poziomem jego osiągnięć szkolnych. Rozbieżności w przedstawionych wynikach analizowanych badań sprawiają, że postanowiono postawić pytanie dotyczące tego, czy i na ile osiągnięcia szkolne związane są z poziomem lęku traktowanego jako cecha. Postawiono również dodatkowe pytanie o związki pomiędzy symptomami napięcia emocjonalnego (mogącymi być symptomami lęku) relacjonowanymi przez rodziców (opiekunów) dziecka a jego poziomem osiągnięć szkolnych, a także o możliwe ich interakcje z poziomem inteligencji dziecka. W ramach eksploracji obszaru zróżnicowania międzyklasowego i międzyszkolnego, postawiono także pytania także o to, czy siła zaobserwowanych zależności między lękiem a osiągnięciami różni się między szkołami i oddziałami klasowymi.

Do pomiaru poziomu lęku jako cechy wykorzystano kwestionariusz STAI-C Spielberga i współpracowników w adaptacji Jaworowskiej (2005). Uczniowie podczas sesji testowej wypełniali obie skale narzędzia, w analizach wykorzystana została jednak tylko skala lęku jako cechy. Składa się ona z 20 pytań dotyczących częstotliwości doświadczania niepokojów i zmartwień (np. „Chce mi się płakać”; „Przejmuję się szkołą”), na które osoba badana udziela odpowiedzi na trzypunktowej skali (rzadko/czasem/często). Wskaźnik poziomu lęku jako cechy utworzono poprzez oszacowanie wartości czynnikowych dla poszczególnych osób badanych na podstawie zweryfikowanego modelu konfirmacyjnej analizy czynnikowej. Następnie uzyskane wyniki zostały wystandaryzowane i przeniesione na skalę o średniej 100 i odchyleniu standardowym 15 (z wykorzystaniem wag próbkowania opisanych w rozdziale metodologicznym raportu). Szczegóły dotyczące właściwości psychometrycznych narzędzia znajdują się w dalszej części rozdziału.

Wskaźników napięcia emocjonalnego doświadczanego przez dziecko dostarczyły dane z ankiety wypełnianej przez rodzica (opiekuna prawnego) dziecka. Pytania dotyczyły częstotliwości występowania u dziecka różnego rodzaju dolegliwości i objawów somatycznych oraz przeżywanych emocji (np. Czy obecnie Pani/Pana dziecko... Czy obecnie obserwuje Pani/Pan następujące objawy u dziecka...?). Osoby badane odpowiadały na pytania na czteropunktowej skali. Dwie spośród trzech utworzonych skal zostały wykorzystane w analizach – skale psychologicznych oraz somatycznych objawów napięcia emocjonalnego. Ich wskaźniki powstały poprzez oszacowanie wartości czynnikowych dla poszczególnych osób badanych na podstawie eksploracyjnego modelu strukturalnego (ESEM). Wskaźniki następnie wystandaryzowano oraz przedstawiono na skali o średniej 100 i odchyleniu standardowym 15 (przekształcenia zostały przeprowadzone z wykorzystaniem wag warunkowych opisanych w rozdziale metodologicznym). Szczegóły dotyczące procedury tworzenia wskaźników znajdują się w dalszej części rozdziału.

Zdecydowaliśmy się na włączenie do analiz zarówno wskaźników lęku pochodzących z samoopisu, którego dokonywały dzieci (lek-cecha), jak i z informacji uzyskanych od rodziców. Te dwa źródła informacji uzupełniają się, choć każde z nich wnosi swoją informację na temat emocjonalnego funkcjonowania dziecka. Jak pokazują badania (Duchesne i in., 2008), samoopisy dokonywane przez dzieci oraz opisy rodziców i nauczycieli mogą się różnić.

Podkreślenia wymaga fakt, iż pomiary te dokonane zostały jednorazowo, gdy uczniowie uczęszczali do klasy trzeciej szkoły podstawowej. W związku z tym przedstawione analizy zakładają ich stałość pomiędzy trzecią i czwartą klasą, choć, jak wcześniej wspomniano, mogą one ulegać zmianom w toku rozwoju dziecka. Do problemu tego powrócimy w dyskusji wyników.

### 6.4. Wyniki

#### 6.4.1. Ogólny plan analiz

W celu weryfikacji hipotez przeprowadziliśmy serię wielopoziomowych analiz regresji, w których zmienne zależne stanowiły kolejno osiągnięcia uczniów z zakresu matematyki, czytania oraz świadomości językowej. Poprzedziła je analiza statystyk opisowych. Modelowanie przebiegło etapowo – w pierwszym kroku dla każdej zmiennej zależnej zbudowany został model pusty z uwzględnieniem poziomu klas i szkół. W drugim kroku powstały wielopoziomowe modele uwzględniające podstawowe zmienne kontrolowane w badaniu (oba rodzaje modeli zostały opisane szczegółowo w rozdziale metodologicznym). Modele te posłużyły jako punkty odniesienia w zakresie weryfikacji polepszenia dopasowania modeli po dodaniu zmiennych i/lub parametrów w kolejnych etapach. W trzecim kroku, celem weryfikacji hipotez dotyczących znaczenia lęku jako cechy oraz obserwowanych przez rodziców objawów napięcia emocjonalnego (psychologicznych i somatycznych) dla osiągnięć szkolnych i zróżnicowania siły tych zależności w szkołach i klasach, wszystkie te zmienne zostały jednocześnie wprowadzone do modelu wraz z interakcjami z inteligencją i losowymi nachyleniami dla szkół i klas dla wszystkich efektów głównych. W następnym etapie te z losowych nachyleń, które okazały się nieistotne statystycznie, były kolejno wykluczane aż do momentu, w którym pozostały jedynie nachylenia istotne statystycznie. Następnie usuwano z modelu predyktory nieistotne. Procedura ta jest zgodna z proponowaną przez Snijdersa i Boskera (2012). W dalszych częściach rozdziału odwoływać będziemy się do opisanych wyżej modeli za pomocą liczb porządkowych, przy czym model pusty oznaczony będzie numerem 1, model kontrolny – numerem 2. W razie specyfikacji kolejnych modeli, będą one oznaczane odpowiednimi numerami porządkowymi.

Obliczenia przeprowadzone zostały na zmiennych wycentrowanych na średniej w całej próbie (centrowaniu podlegały zmienne ilościowe, tj. wartości indeksu dóbr, wykształcenia rodziców w latach, oszacowania poziomu inteligencji, lęku jako cechy oraz napięcia emocjonalnego i jego somatycznych objawów<sup>1</sup>). Uwzględniono dane jedynie tych uczniów, dla których dostępny był komplet zmiennych<sup>2</sup>. Wszystkie analizy przeprowadzone zostały z wykorzystaniem wag warunkowych generowanych oddzielnie dla analiz różniących się zmienną zależną (poziom osiągnięć

<sup>1</sup> Centrowanie przeprowadzone zostało przed usunięciem obserwacji z brakami danych.

<sup>2</sup> Ze względu na przypuszczenie, iż uczniowie o bardzo niskich wynikach w Teście Matryc Ravena mogli w rzeczywistości nie ukończyć jego wypełniania lub odpowiadać przypadkowo, a zatem ich dane są niewiarygodne, z analiz wykluczone zostały obserwacje znajdujące się o więcej niż 3 odchylenia standardowe poniżej średniej dla Testu Matryc Ravena. Spowodowało to zmniejszenie liczebności w analizach dla testu matematycznego o 23 obserwacje, w analizach dla testów czytania i świadomości językowej – o 18 obserwacji.

szkolnych z matematyki, czytania i świadomości językowej; wagi opisane zostały w rozdziale 1). Ze względu na wykorzystanie w analizach PV jako oszacowań poziomu osiągnięć uczniów, obliczenia do każdego modelu dla każdego typu testu osiągnięć powtórzone zostały pięciokrotnie, z wykorzystaniem każdego z pięciu PV jako zmiennej zależnej. Przedstawione w tabelach parametry stanowią wypadkowe wyników z pięciu powtórzeń tego samego modelu. Analizy przeprowadzono z wykorzystaniem oprogramowania HLM 6.06. Podkreślić należy, że analizy wykonano z wykorzystaniem PV obliczonych według nieco innej procedury niż opisana w rozdziale 4. Różnica polegała na przeprowadzeniu warunkowania ze względu na wszystkie zmienne niezależne uwzględnianie w analizach (a także pewne zmienne, które planuje się wykorzystać w analizach w przyszłości). Informacje na temat celów i logiki warunkowania PV czytelnik znajdzie w pracy Von Daviera, Gonzaleza i Mislevy’ego (2009). Warunkowanie przeprowadzono zgodnie z procedurami stosowanymi w badaniach PIRLS i TIMMS (IEA, 2013).

**6.4.2. Analiza psychometryczna narzędzi i tworzenie wskaźników**

W ramach badania właściwości psychometrycznych zastosowanych narzędzi przeprowadziliśmy szereg analiz mających na celu weryfikację ich struktury czynnikowej (w przypadku podskali C-2 z kwestionariusza STAI-C) oraz odkrycie nieobserwowalnych wymiarów w odniesieniu do wskaźników napięcia emocjonalnego podawanych przez rodziców (opiekunów) dziecka.

W ramach analizy psychometrycznej podskali C-2 kwestionariusza STAI-C przeprowadziliśmy konfirmacyjną analizę czynnikową dla zmiennych porządkowych<sup>3</sup> z wykorzystaniem oprogramowania MPlus 6.1. Próba liczyła 5429 obserwacji<sup>4</sup>. Zbudowany model okazał się dobrze dopasowany do danych (RMSEA=0,026; TLI=0,966). Szczegółowe informacje na temat dopasowania modelu znajdują się w tabeli 6.3.

**Tabela 6.3. Dopasowanie weryfikowanego modelu strukturalnego podskali C-2 kwestionariusza STAIC**

miara dopasowania	weryfikowany model
$\chi^2$	784,787; df = 170, p < 0,0001
RMSEA	0,026 (PU: 0,024–0,028)
CFI	0,970
TLI	0,966

PU – przedział ufności

Źródło: opracowanie własne

<sup>3</sup> Metoda estymacji: WLSMV, uwzględniono zgrupowanie uczniów w szkołach.

<sup>4</sup> Liczba pełnych rekordów wykonania wyniosła 5101; braki danych na poszczególnych pozycjach kwestionariusza usuwano parami; odsetek braków danych dla każdej pary zmiennych wyniósł nie więcej niż 1,4%.

## CZY SZKOŁA MA ZNACZENIE?

Pod względem wartości standaryzowanych ładunków czynnikowych model okazał się akceptowalny. Dla poszczególnych pozycji wahały się one od 0,43 (C2\_8) do 0,7 (C2\_19). Wartości ładunków powyżej 0,6 zanotowano dla 6 pozycji, poniżej 0,5 – dla 4 pozycji. Szczegółowe informacje zawarto w poniższej tabeli.

**Tabela 6.4. Wartości ładunków czynnikowych dla pozycji podskali C-2 kwestionariusza STAIC**

pozycja	treść pozycji	ładunek czynnikowy	
		niestandaryzowany	standaryzowany
C2_1	chce mi się płakać	1	0,488
C2_2	za bardzo się martwię	1,198	0,585
C2_3	czuję się nieszczęśliwy	1,067	0,521
C2_4	mam kłopoty	1,190	0,581
C2_5	jakiś nieważne myśli chodzą mi po głowie i dręczą	1,271	0,620
C2_6	denerwuję się, gdy ktoś przygląda się, jak odrabiam lekcje	1,093	0,534
C2_7	czuję przyspieszone bicie serca	1,043	0,509
C2_8	przejmuję się szkołą	0,880	0,429
C2_9	trudno mi się na coś zdecydować	1,122	0,548
C2_10	czuję się skrępowany w obecności dorosłych	1,278	0,624
C2_11	ściska mnie w gardle	1,135	0,554
C2_12	boję się bardziej, niż to okazuję	1,369	0,668
C2_13	trudno mi odpowiadać przed całą klasą	1,168	0,570
C2_14	pocą mi się ręce	0,923	0,451
C2_15	boję się, że mogę zrobić coś złego	1,303	0,636
C2_16	jestem zdenerwowany, gdy muszę pierwszy zacząć rozmowę	1,222	0,596
C2_17	martwię się czymś, co może się zdarzyć	1,396	0,681
C2_18	mam kłopoty z zaśnięciem	0,978	0,477
C2_19	wpadam w zakłopotanie	1,423	0,699
C2_20	martwię się, co inni o mnie myślą	1,302	0,636

Wszystkie ładunki są istotne statystycznie.

Źródło: opracowanie własne

W celu utworzenia wskaźników napięcia emocjonalnego doświadczanego przez dziecko oraz jego obserwowanych objawów somatycznych, przeprowadziliśmy serię eksploracyjnych analiz czynnikowych dla zmiennych porządkowych<sup>5</sup>. W ich toku wyselekcjonowana została pula zmiennych, które następnie weszły w skład czynników oraz określona została optymalna struktura czynnikowa (na podstawie wykresów osypiska, wartości własnych czynników oraz odsetka wyjaśnianej przez nie wariancji).

**Tabela 6.5. Wartości miar dopasowania eksploracyjnego modelu strukturalnego kwestionariusza objawów napięcia emocjonalnego i dolegliwości somatycznych dziecka raportowanych przez rodziców**

miara dopasowania	model ESEM
$\chi^2$	2228,452; df = 187; p < 0,001
RMSEA	0,044 (PU: 0,043–0,046)
CFI	0,960
TLI	0,946

PU – przedział ufności

Źródło: opracowanie własne

Następnie za pomocą eksploracyjnego modelowania równań strukturalnych (ESEM) dla zmiennych porządkowych<sup>6</sup> wyznaczone zostały wartości czynnikowe dla poszczególnych obserwacji. Próba liczyła 5545 obserwacji. Uzyskana struktura czynnikowa ESEM okazała się dobrze dopasowana do danych (RMSEA = 0,044; CFI = 0,96). Szczegółowe informacje na temat dopasowania modelu znajduje się w tabeli 6.5. Ostateczne trójczynnikowe rozwiązanie ESEM przedstawiono w tabeli 6.6. Niepokój budzić może wartość standaryzowanego ładunku czynnikowego dla pozycji „wymioty” powyżej wartości 1. Pamiętać jednak należy, że w przypadku skorelowania czynników ładunki czynnikowe nie są korelacjami lecz współczynnikami regresji i mogą przyjmować wartości znacznie powyżej 1 (por. Jöreskog, 1999).

<sup>5</sup> Metoda estymacji: WLSMV, uwzględniono zgrupowanie uczniów w szkołach.

<sup>6</sup> Metoda estymacji: WLSMV, rotacja Geomin, uwzględniono zgrupowanie uczniów w szkołach.

## CZY SZKOŁA MA ZNACZENIE?

**Tabela 6.6. Wartości ładunków czynnikowych dla czynników wyłonionych z kwestionariusza objawów napięcia emocjonalnego i dolegliwości somatycznych dziecka raportowanych przez rodziców**

pozycja	czynnik 1 – napięcie emocjonalne	czynnik 2 – problemy zdrowotne	czynnik 3 – objawy somatyczne napięcia emocjonalnego	R <sup>2</sup>
jest niespokojne	<b>0,753*</b>	0,007	0,082*	0,627
jest drażliwe, nerwowe	<b>0,947*</b>	-0,033*	-0,012*	0,870
wybucha złością	<b>0,840*</b>	-0,008	-0,057*	0,667
jest roztrzęsione	<b>0,637*</b>	0,012	0,144*	0,508
placze	<b>0,548*</b>	0,101*	0,056*	0,379
bóle brzucha	0,082*	<b>0,484*</b>	0,096*	0,336
stany podgorączkowe	-0,080*	<b>0,655*</b>	0,009	0,409
infekcje	-0,081*	<b>0,559*</b>	0,059*	0,330
nudności	0,051*	<b>0,890*</b>	-0,096*	0,725
wymioty	-0,015*	<b>1,005*</b>	-0,215*	0,794
zaparcia	0,022	<b>0,529*</b>	0,115*	0,375
biegunki	0,021	<b>0,722*</b>	0,021	0,550
bóle stawów	-0,022	0,150*	<b>0,482*</b>	0,331
ból w klatce piersiowej	-0,063*	-0,008	<b>0,832*</b>	0,646
przyspieszone bicie serca	-0,004	-0,049	<b>0,833*</b>	0,646
zawroty głowy	0,004	0,159*	<b>0,588*</b>	0,484
duszności	-0,005	0,106	<b>0,611*</b>	0,458
moczenie nocne	0,113*	-0,01	<b>0,377*</b>	0,185
zgrzytanie zębami	0,079*	0,049	<b>0,322*</b>	0,154
drżenie rąk	0,141*	-0,006	<b>0,584*</b>	0,424
potliwość	0,072*	0,045	<b>0,543*</b>	0,364
bladość	0,006	0,180*	<b>0,492*</b>	0,382
rumieniec	0,010	0,165*	<b>0,403*</b>	0,273

\* istotne statystycznie na poziomie nie wyższym niż 0,05

Źródło: opracowanie własne

Wytłuszczonym drukiem zaznaczone zostały wartości ładunków czynnikowych dla pozycji zakwalifikowanych do odpowiedniego czynnika.

Rozwiązanie to ma przejrzystą interpretację. Pierwszy czynnik wskazuje na doświadczane przez dziecko napięcie emocjonalne. Czynnik drugi obejmuje problemy zdrowotne, natomiast trzeci – somatyczne objawy napięcia emocjonalnego. Wartości korelacji między czynnikami są umiarkowane i wahają się od wartości 0,31 do 0,59; przedstawiono je w tabeli 6.7.

**Tabela 6.7. Wartości korelacji między wyodrębnionymi czynnikami latentnymi**

	czynnik 2	czynnik 3
czynnik 1	<b>0,305</b>	<b>0,407</b>
czynnik 2	1	<b>0,591</b>

Wytłuszczonym drukiem zaznaczono wartości istotne statystycznie.

Źródło: opracowanie własne

### 6.4.3. Statystyki opisowe: natężenie lęku i napięcia emocjonalnego w zbadanej grupie

W analizach z wykorzystaniem testu matematycznego jako zmiennej zależnej uwzględnione zostały dane 4763 uczniów z 300 klas i 172 szkół, co na poziomie indywidualnym stanowi 72,09% pierwotnej liczebności próby oraz 84,68% próby uczniów, których rodzice wyrazili zgodę na badanie. Zbliżone odsetki obserwacji uwzględnione zostały w analizach uwzględniających wyniki testów czytania oraz świadomości językowej jako zmienne zależne. W analizach dla czytania było to 71,27% pierwotnej liczebności próby oraz 83,72% próby uczniów zrekrutowanych (liczba uczniów = 4709, liczba klas = 300, liczba szkół = 172). W analizach dotyczących świadomości językowej było to 71,18% pierwotnej liczebności próby i 83,61% uczniów zrekrutowanych (liczba uczniów = 4703, liczba klas = 300, liczba szkół = 172).

**Tabela 6.8. Statystyki opisowe dla zmiennych charakteryzujących poziom lęku (zmienna zależna wynik testu matematycznego)**

	lęk-cecha	napięcie emocjonalne	somatyczne objawy napięcia	
N	4763	4763	4763	
średnia	100,05	99,67	99,67	
mediana	99,94	99,12	99,97	
odchylenie standardowe	14,82	14,79	14,93	
skośność	0,237	0,320	0,075	
kurtoza	-0,051	-0,373	-0,195	
minimum	72,63	74,46	68,69	
maksimum	158,46	149,86	172,48	
percentyle	10	80,12	79,22	79,86
	25	89,81	88,33	88,90
	75	109,90	109,45	110,37
	90	118,99	119,73	117,77

Źródło: opracowanie własne

W tabeli 6.8 przedstawione zostały statystyki opisowe dla trzech zmiennych charakteryzujących poziom lęku w zbadanej przez nas grupie uczniów. Jako pierwsze zaprezentowane



## CZY SZKOŁA MA ZNACZENIE?

zostaną w tabeli 6.8 analizy w grupie uczniów, którzy nie mieli braków danych w zakresie wyników w teście umiejętności matematycznych i dla których dysponowano kompletem danych w zakresie zmiennych opisujących poziom lęku. Następnie taki sam typ analiz przeprowadzono z uwzględnieniem danych dla testu z pisania i czytania (tabele 6.9 i 6.10).

W zakresie lęku jako cechy średnie wyniki w trzech grupach analiz były bliskie 100 przy odchyleniu standardowym około 14,8. Ich rozkłady okazały się nieco przesunięte w stronę wyników niskich (skośność ok. 0,23), co sygnalizuje możliwy efekt podłogowy (rzeczywiście go zaobserwowano – patrz minimum i 10 percentyl). Spiętrzenie rozkładu okazało się zbliżone do spiętrzenia typowego dla rozkładu normalnego.

**Tabela 6.9. Statystyki opisowe dla zmiennych charakteryzujących poziom lęku (zmienna zależna: osiągnięcia szkolne w zakresie czytania)**

		lęk-cecha	napięcie emocjonalne	somatyczne objawy napięcia
N		4709	4709	4709
średnia		100,13	99,75	99,79
mediana		100,08	99,23	99,08
odchylenie standardowe		14,80	14,79	14,76
skośność		0,233	0,314	0,350
kurtoza		-0,060	-0,373	0,051
minimum		72,63	74,46	71,12
maksimum		158,46	149,86	179,72
percentyle	10	80,22	79,30	80,82
	25	89,88	88,41	88,97
	75	109,87	109,45	109,29
	90	119,05	119,76	118,97

*Źródło: opracowanie własne*

Średnie wyniki dla napięcia emocjonalnego i somatycznych objawów napięcia we wszystkich trzech analizach okazały się bliskie wartości 100 przy odchyleniu standardowym około 14,80. Rozkłady okazały się przesunięte w stronę wyników niskich we wszystkich analizach, jednak przesunięcie to było silniejsze dla somatycznych objawów napięcia (ok. 0,35 w analizach dla świadomości językowej i czytania, jednak 0,075 dla analiz matematycznych) niż dla napięcia emocjonalnego (0,314–0,320). Przesunięcia te sygnalizują obecność efektów podłogowych, które rzeczywiście zaobserwowano (patrz minimum i 10 percentyl). W przypadku napięcia emocjonalnego rozkłady okazały się bardziej spłaszczone, niż jest to typowe dla rozkładu normalnego (kurtoza ok. -0,37), podczas gdy dla somatycznych objawów napięcia były one bliższe rozkładowi normalnemu w analizach dla świadomości językowej i czytania, natomiast w analizach dla matematyki – bardziej spłaszczone (odpowiednio 0,051 i -0,195).

**Tabela 6.10. Statystyki opisowe dla zmiennych charakteryzujących poziom lęku (zmienna zależna: osiągnięcia szkolne w zakresie świadomości językowej)**

	lęk-cecha	napięcie emocjonalne	somatyczne objawy napięcia	
N	4704	4704	4704	
średnia	100,13	99,75	99,79	
mediana	100,08	99,23	99,08	
odchylenie standardowe	14,80	14,79	14,76	
skośność	0,233	0,314	0,350	
kurtoza	-0,060	-0,373	0,051	
minimum	72,63	74,46	71,12	
maksimum	158,46	149,86	179,72	
percentyle	10	80,22	79,30	80,82
	25	89,88	88,41	88,97
	75	109,87	109,45	109,29
	90	119,05	119,76	118,97

Źródło: opracowanie własne

**Tabela 6.11. Wartości korelacji pomiędzy zmiennymi opisującymi poziom lęku**

r-Pearsona	(1)	(2)	(3)
(1) lęk-cecha	1	<b>0,147</b>	<b>0,124</b>
(2) napięcie emocjonalne	<b>0,154</b>	1	<b>0,507</b>
(3) somatyczne objawy napięcia	<b>0,130</b>	<b>0,505</b>	1

Pogrubionym drukiem w tabeli zaznaczono wartości istotne statystycznie.

Źródło: opracowanie własne

Wyniki analizy korelacji między zmiennymi odnoszącymi się do lęku (lęk-cecha, napięcie emocjonalne i somatyczne objawy napięcia) zawarte są w tabeli 6.11. Ponad przekątną tabeli znajdują się korelacje dla analiz, w których zmienną zależną stanowiły osiągnięcia z matematyki. Poniżej przekątnej znajdują się korelacje dla analiz, w których zmienną zależną były osiągnięcia z czytania i świadomości językowej (korelacje przyjęły jednakowe wartości dla obu z nich). Zmienne lęk jako cecha oraz napięcie emocjonalne okazały się słabo dodatnio skorelowane (ok. 0,15). Sugeruje to, iż są to odrębne konstrukty, dotyczące nieco innych obszarów emocjonalności dziecka. Napięcie emocjonalne dotyczy raczej objawów różnego rodzaju przeżyć o negatywnym zabarwieniu (a nie jedynie lęku) i dotyczy konkretnych sytuacji, które dodatkowo muszą zostać przez rodziców dostrzeżone. Z kolei lęk jako cecha jest pewną dyspozycją do doświadczania lęku w różnych sytuacjach, która dodatkowo nie zawsze musi ulegać ekspresji. Z drugiej strony niska korelacja może być wynikiem różnych źródeł informacji o dziecku. W przypadku lęku jako cechy, ono samo udzielało odpowiedzi, natomiast napięcie emocjonalne oraz jego somatyczne objawy raportowali rodzice. Dodatnia umiarkowana korelacja między somatycznymi objawami napięcia oraz napięciem emocjonalnym (ok. 0,5) sugeruje,

## CZY SZKOŁA MA ZNACZENIE?

iż rodzice, którzy obserwują u dzieci objawy napięcia, także często obserwują różnego rodzaju objawy somatyczne. Przymuszczalnie obie zmienne dotyczą raczej obserwowanych trudności emocjonalnych dziecka, które jednak niekoniecznie wynikać mogą z doświadczania lęku, lecz np. z trudności adaptacyjnych w grupie rówieśniczej lub innych przyczyn.

### 6.4.4. Lęk a osiągnięcia z matematyki

Wyniki analiz, w których zmienną zależną stanowiły osiągnięcia szkolne z matematyki, przedstawiono w tabeli 6.12. Model (1) został szczegółowo opisany w rozdziale 3, stąd nie będzie szerzej omawiany. Model (2) stanowi zmodyfikowany wariant modelu podstawowego (również opisanego w rozdziale 3). Modyfikacja polegała na wyrażeniu wykształcenia rodziców w liczbie lat potrzebnej do osiągnięcia danego poziomu wykształcenia. Jeżeli rodzice różnili się pod tym względem, wybierano wartość wyższą. Jeżeli odnotowano brak danych dla jednego z rodziców, uwzględniano dane od drugiego rodzica. W kolejnych krokach wprowadzone zostały zmienne: lęk jako cecha, objawy napięcia emocjonalnego doświadczanego przez dziecko, ich interakcje z inteligencją oraz losowe nachylenia na obu poziomach wszystkich efektów głównych. Wynikiem eliminacji nieistotnych losowych nachyleń oraz nieistotnych efektów głównych jest model (3).

Istotne statystycznie współczynniki regresji uzyskano dla zmiennych:

- napięcie emocjonalne,  $B = -0,051$ ;  $t(22) = -2,78$ ;  $p < 0,011$
- somatyczne objawy napięcia emocjonalnego,  $B = 0,035$ ;  $t(76) = 2,79$ ;  $p < 0,025$
- lęk jako cecha,  $B = -0,035$ ;  $t(50) = -4,14$ ;  $p < 0,001$
- oraz interakcji lęku jako cechy z inteligencją,  $B = 0,002$ ;  $t(171) = 2,22$ ;  $p < 0,028$ .

Ponadto zaobserwowano wariancję nachyleń na poziomie szkół dla zmiennych napięcie emocjonalne,  $U = 0,002$ ;  $p < 0,020$ ; lęk jako cecha,  $U = 0,031$ ;  $p < 0,019$  oraz interakcji lęku-cechy i inteligencji,  $U = 0,003$ ;  $p < 0,005$ .

W celu weryfikacji poprawy dopasowania modelu (3) porównano go do modelu bardziej oszczędnego – zawierającego te same zmienne, jednak bez parametrów losowych nachyleń dla szkół. Ze względu na fakt, iż każdy model powtarzany był pięciokrotnie (dla każdego z pięciu PV jako zmiennej zależnej), porównywano dopasowania odpowiednich modeli dla odpowiednich PV. Okazało się, że model (3) nie jest lepiej dopasowany niż model bardziej oszczędny. Ponadto w modelu bardziej oszczędnym istotność statystyczną utraciła interakcja lęku-cechy z inteligencją. Wobec tego usunięto tę interakcję z modelu i powtórzono porównanie. Ponieważ model (3) w jednym z pięciu powtórzeń nie był lepiej dopasowany do danych niż model oszczędniejszy, został odrzucony. Model (4) przyjęto jako ostateczny. Jednocześnie był on istotnie lepiej dopasowany niż model (2). Szczegóły najważniejszych porównań przedstawiono w tabeli 6.13.

**Tabela 6.12. Wyznaczniki osiągnięć matematycznych: znaczenie lęku jako cechy oraz psychologicznych i somatycznych objawów napięcia emocjonalnego przy kontroli zmiennych psychospołecznych; różnicowanie międzyoddziałowe i międzyszkolne w zakresie siły tych zależności**  
**Wyniki trypoziomowych analiz regresji, odporne (robust) oszacowania błędów standardowych.**

zmienna zależna:	model (1)	model (2)	model (3)	model (4)
<b>osiągnięcia szkolne w zakresie matematyki</b>				
<b>oszacowanie efektów stałych</b>				
<b>poziom ucznia</b>				
stała	<b>99,19</b> (0,535)	<b>100,63</b> (0,399)	<b>100,58</b> (0,409)	<b>100,55</b> (0,406)
wiek dziecka w tygodniach		<b>0,065</b> (0,013)	<b>0,061</b> (0,013)	<b>0,061</b> (0,013)
opóźniony tok nauki		<b>-14,27</b> (1,848)	<b>-13,35</b> (1,879)	<b>-13,52</b> (1,875)
przyspieszony tok nauki		<b>6,19</b> (2,992)	<b>6,12</b> (3,042)	<b>6,27</b> (3,044)
wykształcenie rodziców w latach nauki		<b>0,840</b> (0,107)	<b>0,845</b> (0,106)	<b>0,845</b> (0,107)
indeks zasobów materialnych		<b>0,118</b> (0,019)	<b>0,114</b> (0,019)	<b>0,113</b> (0,019)
inteligencja (TMS-K)		<b>0,535</b> (0,019)	<b>0,532</b> (0,118)	<b>0,531</b> (0,019)
napięcie emocjonalne <sup>a</sup>			<b>-0,051</b> (0,018)	<b>-0,052</b> (0,018)
somatyczne objawy napięcia emocjonalnego <sup>a</sup>			<b>0,035</b> (0,015)	<b>0,038</b> (0,015)
lęk-cecha (STAI-C) <sup>b</sup>			<b>-0,059</b> (0,014)	<b>-0,063</b> (0,014)
interakcja: lęk-cecha i inteligencja			<b>0,002</b> (0,001)	
<b>oszacowanie efektów losowych</b>				
wariancja efektów dla stałych dla szkół	<b>13,36</b>	<b>8,57</b>	<b>9,31</b>	<b>8,88</b>
wariancja nachyleń dla szkół dla somatycznych objawów napięcia			<b>0,0009</b>	
wariancja nachyleń dla szkół dla lęku-cechy			<b>0,0017</b>	
wariancja nachyleń dla szkół dla interakcji: lęk-cecha i inteligencja			<b>0,00001</b>	
wariancja efektów dla stałych dla oddziałów klasowych	<b>17,18</b>	<b>6,88</b>	<b>6,42</b>	<b>6,80</b>
wariancja na poziomie ucznia	<b>193,33</b>	<b>108,75</b>	<b>105,77</b>	<b>107,31</b>

liczba szkół = 172, liczba oddziałów = 300, liczba uczniów = 4763

Pogrubionym drukiem zaznaczono wartości istotne statystycznie na poziomie  $p < 0,05$ .

W nawiasach podano błędy standardowe.

a – raportowane przez rodziców

b – pomiar lęku cechy i inteligencji odbył się tego samego dnia, kolejność pomiarów: TMS-K, STAI-C

Źródło: opracowanie własne

Podsumowując, gdy wynik w skali napięcia emocjonalnego napięcia rośnie o 1 punkt, wyniki w teście z matematyki spadają o około 0,05 punktu (przy wartościach pozostałych zmiennych pozostających bez zmian). Ponadto, gdy wynik w skali lęku jako cechy rośnie o 1, wynik w teście matematycznym spada o 0,06 punktu. Odwrotna zależność ujawniła się dla somatycznych objawów napięcia emocjonalnego – wraz ze wzrostem wyników w tej skali o 1 punkt, wynik w teście matematycznym rośnie o około 0,04 punktu. Choć w modelu (3) zaobserwowano interakcję inteligencji i lęku-cechy oraz międzyszkolne różnicowanie siły

## CZY SZKOŁA MA ZNACZENIE?

zależności między osiągnięciami z matematyki a somatycznymi objawami napięcia emocjonalnego i lękiem-cechą oraz interakcji inteligencji z lękiem-cechą, to efekty te nie poprawiały istotnie przewidywań osiągnięć uczniów, stąd mogą być pominięte.

**Tabela 6.13. Weryfikacja poprawy dopasowania modelu. Wyniki testowania istotności poprawy dopasowania modelu bardziej złożonego w porównaniu z oszczędniejszym modelem kontrolnym**

	<b>model (3)</b> k = 23	<b>model (2)</b> k = 10
model (4) k = 13	(PV1) $\chi^2 = \mathbf{22,32}$	(PV1) $\chi^2 = \mathbf{68,25}$
	(PV2) $\chi^2 = 14,34$	(PV2) $\chi^2 = \mathbf{51,74}$
	(PV3) $\chi^2 = \mathbf{24,33}$	(PV3) $\chi^2 = \mathbf{63,93}$
	(PV4) $\chi^2 = \mathbf{23,81}$	(PV4) $\chi^2 = \mathbf{53,45}$
	(PV5) $\chi^2 = \mathbf{19,40}$	(PV5) $\chi^2 = \mathbf{56,70}$
	df = 10	df = 3

Pogrubionym drukiem tabeli zaznaczono wartości istotne statystycznie.

k – liczba parametrów modelu

df – liczba stopni swobody

(PV1) – (PV5) – porównanie dla powtórzeń modeli, w którym zmienną zależną jest PV o danym numerze porządkowym

Źródło: opracowanie własne

### 6.4.5. Lęk a osiągnięcia w zakresie czytania

W tabeli 6.14 przedstawiono szczegółowe wyniki analiz dla kolejnych modeli, w których zmienną zależną stanowiły osiągnięcia szkolne uczniów w zakresie czytania. Ponieważ HISEI okazał się predyktorem nieistotnym statystycznie, został wyłączony z tego modelu i nie był wprowadzany do kolejnych. Zaznaczyć jedynie należy, że pomimo nieistotności statystycznej zmiennej identyfikującej uczniów o przyspieszonym toku nauki, pozostawiono ją w modelu ze względu na jej znaczenie jako zmiennej kontrolnej (por. Snijders i Bosker, 2012).

W kolejnym kroku jednocześnie wprowadzono zmienne dotyczące lęku wraz z ich interakcjami z poziomem inteligencji oraz losowymi nachyleniami dla stałych na poziomie szkół i oddziałów. Następnie usuwano kolejno: nieistotne statystycznie losowe nachylenia dla stałych, nieistotne statystycznie interakcje, nieistotne statystycznie efekty główne interakcji. Efektem tej procedury jest model (3). Spośród dodanych zmiennych i parametrów pozostały w nim napięcie emocjonalne,  $B = -0,071$ ;  $t(114) = -4,03$ ;  $p < 0,001$  wraz z istotną statystycznie wariancją nachyleń na poziomie szkół,  $U = 0,001$ ;  $p < 0,025$ . Ponadto istotne statystycznie okazały się efekty dla zmiennych somatyczne objawy napięcia emocjonalnego,  $B = 0,046$ ;  $t(54) = 2,48$ ;  $p < 0,016$  oraz lęk jako cecha,  $B = -0,041$ ;  $t(392) = -2,94$ ;  $p < 0,004$ .

**Tabela 6.14. Wyznaczniki osiągnięć w zakresie czytania: znaczenie lęku jako cechy, psychologicznych i somatycznych objawów napięcia emocjonalnego przy kontroli zmiennych psychospołecznych; różnicowanie międzyoddziałowe i międzyszkolne w zakresie siły tych zależności**  
**Wyniki trzypoziomowych analiz regresji, odporne (robust) oszacowania błędów standardowych**

Zmienna zależna: osiągnięcia szkolne w zakresie czytania	model (1)	model (2)	model (3)	model (4)
<b>oszacowanie efektów stałych</b>				
<b>poziom ucznia</b>				
stała	<b>99,29</b> (0,497)	<b>98,95</b> (0,514)	<b>98,98</b> (0,524)	<b>98,97</b> (0,522)
pleć dziecka <sup>a</sup>		<b>3,07</b> (0,689)	<b>2,98</b> (0,697)	<b>2,91</b> (0,693)
wiek dziecka w tygodniach		<b>0,053</b> (0,015)	<b>0,048</b> (0,015)	<b>0,048</b> (0,015)
opóźniony tok nauki		<b>-9,56</b> (2,140)	<b>-8,71</b> (2,072)	<b>-8,72</b> (2,068)
przyspieszony tok nauki		3,07 (2,361)	2,96 (2,374)	2,95 (2,390)
wykształcenie rodziców w latach		<b>0,999</b> (0,114)	<b>1,00</b> (0,113)	<b>1,00</b> (0,113)
HISEI				
indeks dóbr materialnych		<b>0,131</b> (0,019)	<b>0,125</b> (0,019)	<b>0,126</b> (0,019)
inteligencja (TMS-K)		<b>0,401</b> (0,019)	<b>0,397</b> (0,019)	<b>0,397</b> (0,019)
napięcie emocjonalne <sup>b</sup>			<b>-0,071</b> (0,018)	<b>-0,072</b> (0,018)
somatyczne objawy napięcia emocjonalnego <sup>b</sup>			<b>0,046</b> (0,019)	<b>0,046</b> (0,019)
lęk-cecha (STAIC) <sup>c</sup>			<b>-0,041</b> (0,019)	<b>-0,041</b> (0,019)
<b>oszacowanie efektów losowych</b>				
wariancja efektów dla stałych dla szkół	<b>11,43</b>	<b>8,48</b>	<b>9,37</b>	<b>9,21</b>
wariancja nachyleń dla szkół dla napięcia emocjonalnego			<b>0,00124</b>	
wariancja efektów dla stałych dla oddziałów klasowych	<b>14,24</b>	<b>5,80</b>	<b>5,47</b>	<b>5,51</b>
wariancja na poziomie ucznia	<b>193,49</b>	<b>133,37</b>	<b>131,64</b>	<b>131,98</b>

liczba szkół = 172, liczba oddziałów = 300, liczba uczniów = 4709.

Pogrubionym drukiem zaznaczono wartości istotne statystycznie na poziomie  $p < 0,05$ . W nawiasach podano błędy standardowe

a – grupa odniesienia: chłopcy,

b – raportowane przez rodziców,

c – pomiar lęku cechy i inteligencji odbył się tego samego dnia, kolejność pomiarów: TMS-K, STAIC

Źródło: opracowanie własne

Weryfikacja poprawy dopasowania wykazała, iż model (3) jest istotnie lepiej dopasowany do danych niż zmodyfikowany model kontrolny (2) w każdym z pięciu powtórzeń. Ze względu na niską wartość wariancji nachyleń na poziomie szkół dla napięcia emocjonalnego zdecydowano o porównaniu tego modelu z modelem oszczędniejszym (4), z wykluczonym parametrem nachylenia. Analiza wykazała, że model (3) nie jest lepiej dopasowany do danych niż bardziej oszczędny model (4) w każdym z pięciu powtórzeń. Stanowi to podstawę do przyjęcia modelu (4) jako wystarczającego do opisu badanych zależności. Jednocześnie model

## CZY SZKOŁA MA ZNACZENIE?

(4) okazał się istotnie lepiej dopasowany do danych niż model (2). Szczegóły najważniejszych porównań znajdują się w tabeli 6.15.

**Tabela 6.15. Weryfikacja poprawy dopasowania modeli. Wyniki testowania istotności poprawy dopasowania modeli bardziej złożonych w porównaniu z oszczędniejszymi**

	model (2) k = 14	model (3) k = 19
model (4) k = 17	(PV1) $\chi^2 = 38,39$	(PV1) $\chi^2 = 0,433$
	(PV2) $\chi^2 = 50,14$	(PV2) $\chi^2 = 1,070$
	(PV3) $\chi^2 = 42,07$	(PV3) $\chi^2 = 2,198$
	(PV4) $\chi^2 = 54,73$	(PV4) $\chi^2 = 3,626$
	(PV5) $\chi^2 = 34,69$	(PV5) $\chi^2 = -0,192$
	df = 3	df = 2

Pogrubionym drukiem zaznaczono wartości istotne statystycznie.

k – liczba parametrów modelu

df – liczba stopni swobody

(PV1) – (PV5) – porównanie dla powtórzeń modeli, w którym zmienną zależną jest PV o danym numerze porządkowym

Źródło: opracowanie własne

Podsumowując, gdy wynik dziecka w skali napięcia emocjonalnego rośnie o 1 punkt, jego osiągnięcia spadają o ok. 0,07 punktu. Ponadto wraz ze wzrostem wyniku w skali lęku jako cechy o 1 punkt, wynik dziecka w teście czytania spada o ok. 0,04 punktu. Odwrotną zależność zaobserwowano dla obserwowanych objawów napięcia emocjonalnego u dziecka – wraz ze wzrostem wyniku w tej skali o 1 punkt, wyniki dziecka rosną o około 0,05 punktu. Ponadto zaobserwowano międzyszkolne zróżnicowanie siły zależności między poziomem napięcia emocjonalnego a osiągnięciami z czytania, jednak nie pozwala ono lepiej przewidywać osiągnięć uczniów, stąd może zostać pominięte.

### 6.4.6. Lęk a osiągnięcia w zakresie świadomości językowej

Wyniki analiz, w których zmienną zależną stanowiły osiągnięcia szkolne w zakresie świadomości językowej, przedstawiono w tabeli 6.16. Zawiera ona oszacowania parametrów kolejnych modeli, budowanych według schematu przedstawionego wcześniej.

Analizy przebiegły zgodnie z omówionym wcześniej schematem i doprowadziły do modelu oznaczonego numerem (3). Spośród istotnych predyktorów pozostały w nim napięcie emocjonalne,  $B = -0,068$ ;  $t(171) = -4,05$ ;  $p < 0,001$ ; somatyczne objawy napięcia emocjonalnego,  $B = 0,051$ ;  $t(171) = 2,73$ ;  $p = 0,007$ ; lęk jako cecha,  $B = -0,036$ ;  $t(171) = -2,29$ ;  $p < 0,001$  oraz jego interakcja z inteligencją,  $B = 0,003$ ;  $t(75) = 2,41$ ;  $p = 0,019$ . Dla trzech pierwszych z nich pozostały również parametry losowych nachyleń dla szkół, odpowiednio:  $U = 0,003$ ;  $p < 0,025$ ;  $U = 0,006$ ,  $p < 0,001$ ;  $U = 0,003$ ;  $p < 0,025$ .

Tabela 6.16. Wyznaczniki osiągnięć w zakresie świadomości językowej: znaczenie lęku jako cechy, psychologicznych i somatycznych objawów napięcia emocjonalnego przy kontroli zmiennych psychospołecznych; zróżnicowanie międzyoddziałowe i międzyszkolne w zakresie siły tych zależności. Wyniki trzypoziomowych analiz regresji, odporne (*robust*) oszacowania błędów standardowych.

zmienna zależna: osiągnięcia szkolne w zakresie świad. jęz.	model (1)	model (2)	model (3)	model (4)
<b>oszacowanie efektów stałych</b>				
<b>poziom ucznia</b>				
stała	<b>99,03</b> (0,560)	<b>97,84</b> (0,445)	<b>97,90</b> (0,441)	<b>97,87</b> (0,448)
pleć dziecka <sup>a</sup>		<b>5,25</b> (0,371)	<b>5,07</b> (0,367)	<b>5,11</b> (0,370)
wiek dziecka w tygodniach		<b>0,068</b> (0,014)	<b>0,063</b> (0,014)	<b>0,063</b> (0,014)
opóźniony tok nauki		<b>-12,95</b> (1,956)	<b>-12,03</b> (1,884)	<b>-11,97</b> (1,897)
przyspieszony tok nauki		<b>3,74</b> (1,668)	<b>3,58</b> (1,644)	<b>3,47</b> (1,696)
wykształcenie rodziców w latach nauki		<b>0,926</b> (0,123)	<b>0,903</b> (0,121)	<b>0,923</b> (0,121)
HISEI		<b>0,041</b> (0,016)	<b>0,045</b> (0,016)	<b>0,043</b> (0,016)
indeks zasobów materialnych rodziny dziecka		<b>0,119</b> (0,0217)	<b>0,111</b> (0,017)	<b>0,113</b> (0,017)
inteligencja (TMS-K)		<b>0,440</b> (0,018)	<b>0,435</b> (0,018)	<b>0,437</b> (0,018)
napięcie emocjonalne <sup>b</sup>			<b>-0,068</b> (0,017)	<b>-0,069</b> (0,017)
somatyczne objawy napięcia emocjonalnego <sup>b</sup>			<b>0,051</b> (0,019)	<b>0,049</b> (0,018)
lęk-cecha (STAIC) <sup>c</sup>			<b>-0,036</b> (0,016)	<b>-0,034</b> (0,016)
interakcja: inteligencja – lęk-cecha			<b>0,0027</b> (0,001)	<b>0,0024</b> (0,001)
<b>oszacowanie efektów losowych</b>				
wariancja efektów dla stałych dla szkół	<b>17,98</b>	<b>8,38</b>	<b>9,12</b>	<b>9,30</b>
wariancja nachyleń dla szkół dla napięcia emocjonalnego			<b>0,0032</b>	
wariancja nachyleń dla szkół dla somatycznych objawów napięcia			<b>0,0062</b>	
wariancja nachyleń dla szkół dla lęku-cechy (STAIC)			<b>0,0034</b>	
wariancja efektów dla stałych dla oddziaływań klasowych	<b>14,15</b>	<b>6,10</b>	<b>5,72</b>	<b>5,52</b>
wariancja na poziomie ucznia	<b>191,20</b>	<b>114,85</b>	<b>110,84</b>	<b>113,20</b>

liczba szkół= 172, liczba oddziałów = 300, liczba uczniów = 4704. Pogrubionym drukiem opisywane są wartości istotne statystycznie na poziomie  $p < 0,05$ . W nawiasach podano błędy standardowe.

a – grupa odniesienia: chłopcy

b – raportowane przez rodziców

c – pomiar lęku cechy i inteligencji odbył się tego samego dnia, kolejność pomiarów: TMS-K, STAIC

Źródło: opracowanie własne

W kolejnych krokach testowana była istotność poprawy dopasowania modelu. W efekcie wyeliminowano efekty losowych nachyleń na poziomie szkół dla wykrytej interakcji, napięcia emocjonalnego oraz lęku-cechy. Nie poprawiały one istotnie dopasowania we wszystkich powtórzeniach modelu. W efekcie powstał model (4), uznany za ostateczny. Okazał się on



## CZY SZKOŁA MA ZNACZENIE?

istotnie lepiej dopasowany do danych niż model bez tej interakcji. Szczegóły porównań modelu (3) i (4) znajdują się w tabeli 6.17.

**Tabela 6.17. Weryfikacja poprawy dopasowania modeli. Wyniki testowania istotności poprawy dopasowania modeli bardziej złożonych w porównaniu z oszczędniejszymi**

	model (3) k = 25
model (4) k = 16	(PV1) $\chi^2 = 16,31$ (PV2) $\chi^2 = 12,99$ (PV3) $\chi^2 = 11,92$ (PV4) $\chi^2 = 15,89$ (PV5) $\chi^2 = 10,96$ df = 9

k – liczba parametrów modelu

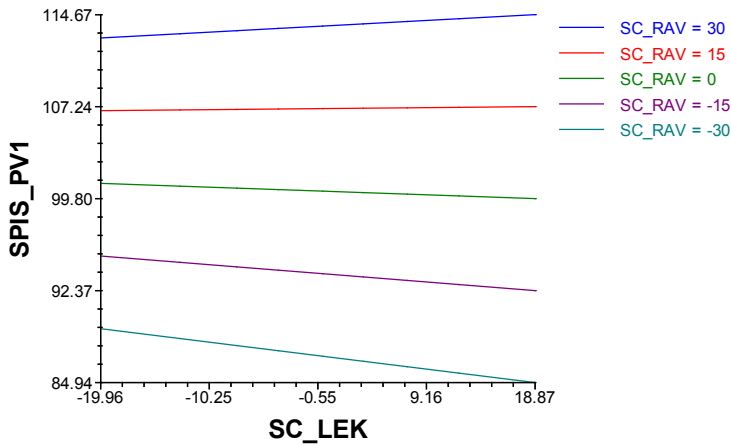
df – liczba stopni swobody

(PV1) – (PV5) – porównanie dla powtórzeń modeli, w którym zmienną zależną jest PV o danym numerze porządkowym

Źródło: opracowanie własne

Istotnymi statystycznie predyktorami kompetencji w zakresie świadomości językowej okazały się napięcie emocjonalne,  $B = -0,068$ , somatyczne objawy napięcia emocjonalnego,  $B = 0,049$ ; oraz lęk jako cecha,  $B = -0,034$ . Ponadto zaobserwowano interakcję między poziomem inteligencji i poziomem lęku jako cechy,  $B = 0,002$ . Wykres 6.1 obrazuje naturę tej interakcji. Widoczne jest, że wzrost lęku jako cechy wiąże się z obniżeniem osiągnięć wśród uczniów o inteligencji niższej niż przeciętna. W przypadku uczniów o przeciętnym poziomie inteligencji cecha ta wydaje się mieć niewielkie znaczenie, a wśród uczniów o inteligencji powyżej przeciętnego poziomu, może się wiązać nawet z nieznacznym podwyższeniem osiągnięć.

Podsumowując, wraz ze wzrostem wyniku w skali napięcia emocjonalnego o 1 punkt osiągnięcia w zakresie świadomości językowej spadają o 0,07 punktu. Ponadto, gdy o 1 punkt rośnie wynik w skali lęku jako cechy, osiągnięcia te spadają o ok. 0,04 punktu. Jednakże im wyższy jest poziom inteligencji dziecka, tym spadek ten jest łagodniejszy, a sama cecha traci znaczenie wśród uczniów o inteligencji w normie intelektualnej. Wzrost osiągnięć o około 0,05 punktu obserwujemy, gdy rośnie o 1 punkt wynik w skali somatycznych objawów napięcia emocjonalnego. Ponadto zaobserwowano międzyszkolne zróżnicowanie siły zależności lęk jako cecha – osiągnięcia, napięcie emocjonalne – osiągnięcia oraz zróżnicowany wzorzec interakcji, jednakże efekty te nie poprawiały istotnie przewidywań osiągnięć uczniów, stąd mogą zostać pominięte.



**Rysunek 6.1. Interakcja między inteligencją a lękiem jako cechą.**

SC\_LEK – wynik w skali lęk-cecha testu STAIC (wycentrowany wokół zmiennej w całej próbie)

SC\_RAV – wynik w Teście Matryc Ravena (wycentrowany wokół średniej w próbie)

SPIS\_PV1 – wynik testu świadomości językowej, oszacowanie PV1 (skala 100;15)

Źródło: opracowanie własne

## 6.5. Podsumowanie

Celem rozdziału było znalezienie odpowiedzi na pytania dotyczące znaczenia lęku jako cechy i obserwowanych przez rodziców objawów napięcia emocjonalnego (psychologicznych i somatycznych) oraz ich interakcji z inteligencją dla osiągnięć szkolnych. Sprawdzono także możliwe międzyszkolne i międzyoodziałowe zróżnicowanie siły tych zależności.

Wyniki przeprowadzonych analiz wskazują, że istnieje zależność między poziomem lęku jako cechy oraz objawami napięcia a osiągnięciami szkolnymi we wszystkich trzech badanych obszarach kompetencji uczniów. Gdy wynik w skali napięcia emocjonalnego rośnie o 1 punkt, osiągnięcia szkolne z matematyki, czytania i świadomości językowej spadają odpowiednio o 0,07; 0,07 i 0,05 punktu. Gdy wynik w teście lęku jako cechy rośnie o jeden punkt, wynik testu z matematyki, czytania oraz świadomości językowej obniżają się odpowiednio o ok. 0,03; 0,04 i 0,06 punktu. Rosną natomiast o odpowiednio 0,05; 0,05 i 0,04 punktu, gdy rosną o 1 punkt wyniki w skali somatycznych objawów napięcia emocjonalnego.

Ponadto zaobserwowano interakcję ilorazu inteligencji z lękiem-cechą w analizach dla kompetencji w zakresie świadomości językowej. Wskazuje ona zróżnicowane znaczenie lęku-cechy dla osiągnięć uczniów w zależności od poziomu inteligencji. Spadek osiągnięć

## CZY SZKOŁA MA ZNACZENIE?

obserwowany jest wraz z rosnącym nasileniem lęku-cechy, jednak tylko wśród uczniów z wynikami w teście inteligencji poniżej przeciętnej.

Interpretacji wymaga nieoczekiwana pozytywna (choć słaba) zależność między osiągnięciami szkolnymi a somatycznymi objawami napięcia emocjonalnego. Wśród jego wskaźników znalazły się takie pozycje jak: zawroty głowy, duszności, bóle w klatce piersiowej. Pamiętać należy, że miara ta powstała na podstawie odpowiedzi rodziców, odzwierciedla więc sposób spostrzegania dziecka i podlegać może różnego rodzaju zniekształceniom w stosunku do stanu rzeczywistego. Postrzeganie funkcjonowania dziecka może wiązać się z szeregiem nastawień i przekonań rodzica jak również z jego osobistymi cechami, w tym poziomem lęku (Wood i in. 2003). W literaturze przedmiotu wskazuje się również na systematyczne choć słabe związki pomiędzy praktykami wychowawczymi a lękiem dzieci (McLeod, Wood, Weisz 2007). Rodzice dzieci, które wykazują wyższy poziom lęku częściej klasyfikowani są jako silniej kontrolujący i mniej wspierający autonomię dziecka, co może również przekładać się na zachowanie rodzica dotyczące monitorowania realizacji przez dziecko obowiązków szkolnych. Z drugiej strony manifestowane przez dziecko objawy lęku mogą wyzwać określone zachowania rodzica np. związane z sposobem dyscyplinowania dziecka, kontrolowania jego zachowań czy wreszcie zakresem wsparcia udzielanym dziecku podczas pracy na zadaniach szkolnymi (Wood i in. 2003).

Stąd zaobserwowaną zależność można próbować wyjaśniać w kategoriach intencjonalnych i nieintencjonalnych działań podejmowanych przez rodziców na skutek obserwacji różnych somatycznych objawów u dziecka. Dziecko przejawiające je, może być spostrzegane jako słabsze, gorszego zdrowia, a więc wymagające dodatkowej pomocy i wsparcia. Kierowane spostrzeżeniami działania rodzica mogą z kolei przekładać na nieco wyższe osiągnięcia w nauce. Rodzice mogą też podejmować dodatkowe działania mające na celu ochronę dziecka, np. mogą poświęcać mu więcej czasu, w mniejszym stopniu zachęcać do uprawiania sportu, spędzania czasu poza domem itp. W konsekwencji dziecko, z racji problemów somatycznych, rzadziej zachęcane jest do zajęć sportowych czy spotkań z rówieśnikami, a częściej – do zajęć tego niewymagających (np. do nauki, czytania książek), częściej też przebywa z osobami dorosłymi. To bogatsze intelektualnie środowisko przekładać się może na wyższe osiągnięcia.

W omawianym przez nas badaniu rodzice oceniając, na ile ich dziecko doświadcza określonych problemów somatycznych, mieli koncentrować się na sytuacjach mających miejsce obecnie. Mogli oni jednak odpowiadając na tak sformułowane pytanie, przywoływać z pamięci różne sytuacje i zachowania dziecka, niekoniecznie te najbardziej aktualne. To z kolei mogło przyczynić się do zniekształcenia obserwowanej zależności.

Wiadomo, że lęk jako cecha bezpośrednio oraz pośrednio (za pośrednictwem lęku jako stanu) może wpływać na wyniki w nauce (np. Heinrich, 1979; King, Heinrich, Stephenson

i Spielberger, 1976), co wydaje się znajdować potwierdzenie w przeprowadzonych przez nas analizach. Otwartym pozostaje pytanie o mechanizmy odpowiadające za zaobserwowane efekty interakcji inteligencji z lękiem-cechą w przypadku osiągnięć w zakresie świadomości językowej. Nie jest też jasne, dlaczego w jedynie w tym obszarze osiągnięć zaobserwowano istotną poprawę przewidywań poziomu osiągnięć uczniów w modelach zawierających tę interakcję.

Wyniki te wskazują jednak, że można przypuszczać, iż wysoki poziom inteligencji pozwala uczniom nawet przy wyższym poziomie lęku lepiej radzić sobie z wymaganiami stawianymi przez szkołę. W sytuacji, w której poziom lęku nie powoduje powstawania poważnych zaburzeń, staje się on czynnikiem mobilizującym. Jak wiadomo od czasu sformułowania klasycznego prawa Yerksa-Dodsona zależność pomiędzy poziomem pobudzenia (w niniejszym badaniu – lęku jako cechy) a poziomem wykonania jest krzywoliniowa, jak również uzależniona jest od poziomu złożoności zadania (Maruszewski i in. 2008). Ważne jest jednak, aby pamiętać, że jednostki różnią się charakterystycznym dla siebie optymalnym stanem pobudzenia, które wyznacza zakres w jakim osiągają one najlepsze wyniki. Zbadani przez nas uczniowie zapewne również różnili się zakresem tolerancji lękowego pobudzenia, które było dla nich optymalne i nie powodowało ani stanu nudy i demotywacji, kiedy pobudzenie jest zbyt niskie, ani z drugiej strony nie prowadziło do nadmiernego pobudzania.

W tym kontekście należy pamiętać, że w przypadku dzieci o niższym ilorazie inteligencji, co do których można przypuszczać, że ich zasoby poznawcze są mniejsze, wysoki poziom lęku-cechy może być czynnikiem szczególnie zakłócającym funkcjonowanie szkolne, zarówno w kontekście osiągnięć jak i relacji z rówieśnikami. Stała gotowość do interpretacji sygnałów płynących z otoczenia jako niosących zagrożenie, obecna u dzieci z wysokim poziomem lęku-cechy powoduje ograniczanie możliwości przetwarzania innych informacji i pogarsza funkcjonowanie poznawcze (Owens i in. 2008). Co przy mniejszych zasobach (niższy iloraz inteligencji) może przyczyniać się do pogorszenia się poziomu osiągnięć.

Rozważenia wymaga problem znaczenia obiektywnych cech sytuacji dla wzbudzenia lęku jako stanu podczas pomiaru osiągnięć szkolnych, gdyż mógł on mieć znaczenie dla uzyskanych wyników analiz. Nie jest wykluczone, że pomimo standardowej procedury przeprowadzania testów kompetencji szkolnych (przygotowanej tak, by minimalizowała potencjalny stres i lęk) oraz przeszkolenia badaczy, indywidualny styl prowadzenia badania (sam lub w interakcji z innymi czynnikami, np. zachowaniem uczniów, jakością kontaktu z dyrektorem szkoły, wychowawcą klasy), mógł przekładać się na sposób, w jaki różni badacze i sytuacja testowania odbierane były przez uczniów, tj. na ile cechy sytuacji stanowiły czynnik lękotwórczy. Uczniowie różniący się poziomem lęku jako cechy są w różnym stopniu podatni na doświadczenie lęku jako stanu pod wpływem cech sytuacji, co z kolei wpływać może na wyniki

## CZY SZKOŁA MA ZNACZENIE?

uzyskiwane podczas testowania osiągnięć. Pomiar lęku jako cechy oraz napięcia emocjonalnego wykorzystywany w badaniach nie uwzględnia potencjalnego znaczenia tego czynnika, o czym należy pamiętać, dokonując interpretacji wyników.

Przypomnieć także należy, że przeprowadzone analizy zakładają niezmiennosc poziomu lęku jako cechy oraz napięcia emocjonalnego w czasie. Pomimo relatywnie krótkiego czasu między pomiarami lęku i napięcia emocjonalnego oraz osiągnięć (niespełna rok), ostatni z nich odbywał się już w nowych warunkach, po przejściu uczniów na kolejny etap kształcenia. Relatywna nowość związana z inną niż podczas nauczania początkowego organizacją kształcenia (np. kształcenie przez nauczycieli przedmiotowych) mogła wpłynąć na bieżące doświadczanie lęku i napięcia przez dzieci (zwłaszcza uczniów o wysokim poziomie lęku jako cechy). Mogły one znaleźć odzwierciedlenie w wynikach testów osiągnięć, jednak ze względu na plan poszczególnych pomiarów, ich znaczenie nie zostało uchwycone w zastosowanych miarach. Mogły zajść także zmiany w poziomie obu zmiennych wynikające z innych przyczyn. Nieuchwycenie tych czynników mogło mieć wpływ na wyniki analiz.

Niezbyt duża siła zależności raportowanych w tym rozdziale może niesłusznie sugerować, iż można zignorować znaczenie lęku (rozumianego jako cecha osobowości) oraz obserwowanego przez rodziców napięcia emocjonalnego dziecka dla osiągnięć szkolnych. Jest to wniosek niesłuszny, wzięwszy pod uwagę naturę metod statystycznej analizy danych, które mówią o typowych, przeciętnych tendencjach w populacji, nie opisują natomiast pojedynczych jednostek.

Opisane przez nas wyniki wskazują, że wraz ze wzrostem lęku poziom osiągnięć się obniża. Związki te co prawda nie są silne, ale były one obserwowane w odniesieniu do każdej z trzech badanych przez nas umiejętności szkolnych. U dzieci, u których obserwuje się podwyższony poziom lęku-cechy, nawet jeśli nie spełniają one kryteriów zaburzeń lękowych, występować może podwyższona gotowość do reagowania lękiem sytuacyjnym na różnego rodzaju wymagania szkolne.

Sytuacja, w której jeden lub więcej uczniów w klasie przejawia podwyższony poziom lęku zauważalny przez dorosłych (rodziców, nauczyciela), powinna prowadzić do indywidualnej diagnozy problemów dziecka i jeżeli taka jest konieczność, podjęcia działań mających na celu udzielenie mu odpowiedniej pomocy. Szczególnie ważna jest tu czujność opiekunów dziecka (wychowawcy klasy, rodziców), którzy zazwyczaj dysponują informacjami pochodzącymi z obserwacji zachowań dziecka i którzy powinni reagować na pojawiające się symptomy zarówno świadczące o napięciu emocjonalnym jak i objawach somatycznych. Należy również pamiętać, że w przypadku dzieci lękowych, ze względu na charakter przeżywanego problemu, np. tendencje do zamartwiania się, negatywne interpretacje własnych działań, część

z nich jest trudna do zaobserwowania w toku codziennych klasowych interakcji. Dzieci takie potrzebują stworzenia bezpiecznej sytuacji, w której mogłyby swobodnie opowiadać o swoich problemach, lękach i obawach. W kontekście pomocy udzielanej tym dzieciom szczególnie istotna jest współpraca szkoły i rodziców. Wskazuje się bowiem, że część zachowań mogących świadczyć o wysokim poziomie lęku i związanych z tym problemach, np. w kontaktach z rówieśnikami, obserwowana jest w szkole i dostrzegana przez nauczycieli, natomiast rodzice nie mają świadomości występowania tego typu utrudnień (por. Duchesne i in. 2008). Zatem w tym zakresie konieczna jest obustronna komunikacja i wymiana informacji pomiędzy szkołą a domem rodzinnym dziecka. Ma to znaczenie nie tylko w kontekście osiągnięć szkolnych, lecz także dobrostanu i optymalnego rozwoju dziecka.

Nie można jednak zapomnieć o adaptacyjnym znaczeniu lęku i innych emocji o negatywnych zabarwieniu. Niejednokrotnie są one właściwą reakcją na sytuację zagrożenia, która wywołuje działania pozwalające ochronić organizm, np. powoduje ucieczkę lub sprawia, że dziecko nie zbliży się do dużego, szczekającego psa. Innymi słowy, nie zawsze lęk dziecka świadczy o problemach w tym zakresie, stąd konieczna jest znajomość norm i prawidłowości rozwojowych. Charakter i źródło lęków u dzieci ulegają istotnym zmianom na przestrzeni okresu szkolnego i okresu adolescencji (por. Kendal 2004). Świadomość tych zmian oraz kompetencje do ich właściwego rozpoznawania i nazywania wydają się szczególnie istotne w obliczu pojawiających się informacji o zwiększającym się odsetku dzieci cierpiących na różnorodne zaburzenia emocjonalne (Czabała i in. 2005). W tym kontekście szczególnie ważne wydaje się odpowiednie kształcenie nauczycieli szkół podstawowych w zakresie psychologii rozwoju człowieka, z naciskiem na rozwój dziecka w wieku przedszkolnym i szkolnym.

W odniesieniu do znaczenia lęku-cechy dla osiągnięć szkolnych, ważne wydają się także warunki i atmosfera panująca w klasie i szkole. Wiedząc, iż tak rozumiany lęk jest dyspozycją do odbierania różnorodnych bodźców jako lękotwórczych i odczuwania lęku sytuacyjnego, należy także zadbać o to, by środowisko szkolne było spostrzegane przez dziecko jako bezpieczne i zawierało jak najmniej lękotwórczych elementów. Elementy te obejmować mogą zarówno infrastrukturę szkoły (np. bliskość klas uczniów nauczania początkowego z uczniami z ostatnich klas szkoły podstawowej, związany z tym hałas), jak i czynniki wewnątrz klasy (np. zachowanie innych uczniów, sposób prowadzenia lekcji czy sprawdzania postępów w nauce). Z drugiej strony nie jest jednak możliwe (ani wskazane) eliminowanie wszystkich czynników potencjalnie lękotwórczych, gdyż istotną rolę odgrywa także tworzenie przez wychowawcę warunków, które pozwolą dziecku rozwijać kompetencje emocjonalne. Tworzenie okazji do konfrontacji z sytuacjami potencjalnie trudnymi, które wymagają mobilizacji i zaangażowania, a jednocześnie dają również okazję do doświadczenia

## CZY SZKOŁA MA ZNACZENIE?

sukcesu. Staje się niezmiernie ważne w kontekście kształtowania kompetencji samoregulacji i samokontroli (por. Moffitt i in. 2010).

Analizowany problem związków lęku z osiągnięciami szkolnymi można potraktować szerzej jako problem istotności dobrostanu emocjonalnego dziecka i jego związków z poziomem osiągnięć.

Dla optymalnego funkcjonowania i realizacji własnego potencjału najbardziej pożądaną formą motywacji według Deciego i Ryana (2008) jest motywacja wewnętrzna. Wiąże się ona nie tylko z większym prawdopodobieństwem wyższych osiągnięć, ale jest również istotna dla dobrostanu. Po to, aby możliwy był rozwój motywacji wewnętrznej, muszą zostać zaspokojone trzy podstawowe według tych autorów potrzeby: kompetencji, autonomii i powiązań. Stwarzanie okazji do zaspokojenia właśnie tych potrzeb może być kluczowe również w kontekście minimalizowania wpływu lęku na osiągnięcia oraz zmniejszania niebezpieczeństwa pojawienia się zaburzeń lękowych w wyniku oddziaływań szkoły. Warto zwrócić uwagę, iż w przypadku dzieci z podwyższonym poziomem lęku, problemy w ich funkcjonowaniu szkolnym nie dotyczą tylko kwestii osiągnięć, ale również kontaktów z rówieśnikami i nauczycielami. Nadmiernie wysoki poziom lęku wiąże się z trudnościami w budowaniu i podtrzymywaniu satysfakcjonujących relacji z rówieśnikami, co z kolei prowadzi do niższych kompetencji społecznych. Zatem uczniowie ci mogą mieć trudności w zaspokojeniu potrzeby afiliacji i przynależności, mogą mieć poczucie, że są spychani poza centrum klasowego i rówieśniczego życia, co wtórnie może przekładać się na stosunek do szkoły, motywację do nauki, a w efekcie – osiągnane wyniki. Działania mające na celu integrację uczniów, prowadzące do polepszenia współpracy między nimi oraz budowania dobrej atmosfery w klasie, zwrótnie mogą przyczyniać się do obniżania poziomu lęku u dzieci, które biorą w nich udział. Trzeba jednak mieć świadomość, że ci uczniowie, u których natężenie lęku jest wysokie i u których diagnozuje się zaburzenia czy to lękowe, czy to inne zaburzenia emocjonalne, wymagać będą specjalistycznych oddziaływań poprzedzających ich udział w takich ogólnoklasowych zajęciach.

W sytuacji zaobserwowania u dziecka objawów lęku, które wykraczają poza typowe normy rozwojowe, potrzebna jest diagnoza jego trudności oraz, w razie potrzeby, odpowiednie wsparcie i działania pomocowe. Identyfikacji problemów sprzyja współpraca na linii szkoła – rodzic, a także odpowiednia wiedza profesjonalna wychowawcy w zakresie psychologii rozwojowej, stąd obie z nich powinny być rozwijane. Wskazane jest także takie organizowanie środowiska szkolnego i klasowego, by nie zawierało silnie lękotwórczych czynników oraz wspierało rozwój motywacji wewnętrznej, przy jednoczesnym tworzeniu dziecku warunków do rozwijania kompetencji emocjonalnych.

## Literatura cytowana

- Bill, H., Moffit, T., Silva, P. A. i McGee, R. (1991) Anxiety and cognitive task performance: A longitudinal perspective. *Child Study Journal*, 21 (3), 167–183
- Caspi, A., Elder G.,H. i Bem D.,J. (1988). Moving away from the world: Life course pattern of shy children. *Developmental Psychology* 24, 824–831.
- Czabała, Cz., Brykczyńska, C., Bobrowski, K. i Ostaszewski, K., (2005). Problemy zdrowia psychicznego w populacji gimnazjalistów warszawskich. *Postępy Psychiatrii i Neurologii* 14 (1), 1–9.
- Deci, E. L. (1992). On the nature and functions of motivation theories. *Psychological Science*, 3,167–171.
- Deci, E. L. i Ryan, R. M. (2008). Facilitating optimal motivation and psychological well-being across life's domains. *Canadian Psychology*, 49, 14–23.
- Duchesne, S., Vitaro, F., Larose, S. i Tremblay, R. E. (2008). Trajectories of anxiety during elementary-school years and the prediction of high school noncompletion. *Journal of Youth and Adolescence*, 37, 1134–1146
- Goleman, D. (1997). *Inteligencja Emocjonalna*. Poznań: Media Rodzina.
- Grover, R., Ginsburg, G. i Ialongo, N. (2007). Psychosocial outcomes of anxious first graders: a seven year follow-up. *Depression & Anxiety*, 24, 410–420.
- Hughes, A. A., Lourea-Waddell, B. i Kendall, P. C. (2008). Somatic complaints In children with anxiety disorders and their unique prediction of poorer academic performance. *Child Psychiatry and Human Development*, 39, 211–220.
- Ialongo, N., Edelsohn, G., Werthamer-Larsson, L., Crockett, L. i Kellam, S. (1995). The significance of self-reported anxiety symptoms in first grade children: Prediction to anxious symptoms and adaptive functioning in fifth grade. *Journal of Child Psychology and Psychiatry*, 36, 427–437.
- IEA (2013). *TIMSS and PIRLS Achievement Scaling Methodology*. Pobrano z [http://timssandpirls.bc.edu/methods/pdf/TP11\\_Scaling\\_Methodology.pdf](http://timssandpirls.bc.edu/methods/pdf/TP11_Scaling_Methodology.pdf)
- Izard, C., E. i Ackerman, B., P. (2000). Motywacyjne, organizacyjne i regulacyjne funkcje odrębnych emocji. W: M. Lewis i J. M. Haviland-Jones (red.), *Psychologia Emocji* (s. 327–341). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Jaworowska, A. (2005). *Inwentarz stanu i cechy lęku STAIC C. D. Spielbergera, C. D. Edwardsa, R. E. Lushene'a, J. Montouriniego i D. Platzeck. Podręcznik*. Warszawa: Pracownia Testów PTP.



## CZY SZKOŁA MA ZNACZENIE?

- Jöreskog, K. G. (1999). How large can a standardized coefficient be? Pobrano z: <http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf>
- Kagan, J. (2008). *Trzy złudne przekonania*. Kraków: Wydawnictwo WAM.
- Kendall, P. C. (2004). *Zaburzenie okresu dzieciństwa i adolescencji. Mechanizmy zaburzeń i techniki terapeutyczne*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Krejtz, I. (2012). *Korepetycje poznawcze. Rola pamięci roboczej i kontroli uwagi w przewidywaniu osiągnięć szkolnych*. Warszawa: Wydawnictwo Akademickie Sedno.
- Lau, J. Y. F., Eley, T. C. i Stevenson, J. (2006). Examining the state-trait anxiety relationship: A behavioural genetic approach. *Journal of Abnormal Child Psychology*, 34, 19–27.
- Maruszewski, T., Doliński, D., Łukaszewski W. i Marszał-Wiśniewska, M. (2008). Emocje i motywacja. W: J. Strelau, D. Doliński (red.), *Psychologia. Podręcznik akademicki* tom I (s. 511–648). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Mazzone, L., Ducci, F., Scoto, M. C., Passaniti, E., D'Arrigo, V. G. i Vitiello, B. (2007). The role of anxiety in a community sample of children and adolescents, *BioMed Central Public Health*, 7, 347–411.
- McCann, S. J. i Meen, K. S. (1984). Anxiety, ability, and academic achievement. *Journal of Social Psychology*, 124, 257–258.
- McLeod, B. D., Wood, J. J. i Weisz, J. R. (2007). Examining the association between parenting and childhood anxiety: A meta-analysis. *Clinical Psychology Review*, 27, 155–172.
- Milgram, R. M. i Milgram, N. A. (1977). The effect of test content and context on the anxiety-intelligence relationship. *The Journal of Genetic Psychology*, 130, 121–127.
- Newbegin, I. i Owens, A. (1996). Self-esteem and anxiety in secondary school achievement. *Journal of Social Behavior and Personality*, 11(3), 521–530.
- Öhman, A. (2000). Strach i lęk z perspektywy ewolucyjnej, poznawczej i klinicznej. W: M. Lewis i J. M. Haviland-Jones (red.), *Psychologia Emocji* (s. 719–744). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Owens, M., Stevenson, J., Norgate, R. i Hadwin, J. A. (2008). Processing efficiency theory in children: Working memory as a mediator between trait anxiety and academic performance. *Anxiety, Stress & Coping*, 21(4), 417–430.
- Papay, J. P. i Spielberger, C. D. (1986). Assessment of anxiety and achievement in kindergarten and first- and second-grade children. *Journal of Abnormal Child Psychology*, 14, 279–286.

- Papay, J. P., Costello, R. J., Hedl, J. J. i Spielberger, C. D. (1975). Effects of trait and state anxiety on the performance of elementary school children in traditional and individualized multiage classrooms. *Journal of Educational Psychology*, 67, 840–846.
- Snijders, T. A. i Bosker, R. J. (2012). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. 2<sup>nd</sup> edition. Sage.
- Solevey, P. i Sluyter, D., J. (1999). *Rozwój emocjonalny a inteligencja emocjonalna: problemy edukacyjne*. Poznań: Dom Wydawniczy Rebis.
- Strahan, E. Y. (2003). The effects of social anxiety and social skills on academic performance. *Personality and Individual Differences*, 34, 347–366.
- Tucholska, S. i Steuden, S. (1990). Inwentarz do pomiaru leku u dzieci – STAIC i jego polska wersja. *Psychologia Wychowawcza*, 1–2, 50–58.
- Van Ameringen, M., Mancini, C. i Farvolden, P. (2003). The impact of anxiety disorders on educational achievement. *Journal of Anxiety Disorders*, 17, 561–571.
- Von Davier, M., Gonzalez, E. i Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, 2, 9–36.
- Vitaro, F., Brendgen, M., Larose, S. i Tremblay, R. E. (2005). Kindergarten disruptive behaviors, protective factors, and educational achievement by early adulthood. *Journal of Educational Psychology*, 97, 617–629.
- Wolańczyk T. (2002). *Zaburzenia emocjonalne i behawioralne u dzieci i młodzieży szkolnej w Polsce*. Warszawa: Akademia Medyczna w Warszawie.
- Wood, J. J., McLeod, B. D., Sigman, M., Hwang, W. C., i Chu, B. C. (2003). Parenting and childhood depression: Theory, empirical findings, and future directions. *Journal of Child Psychology and Psychiatry* 44, 134–151.

## Instytut Badań Edukacyjnych

Głównym zadaniem Instytutu jest prowadzenie badań, analiz i prac przydatnych w rozwoju polityki i praktyki edukacyjnej.

Instytut zatrudnia ponad 150 badaczy zajmujących się edukacją – pedagogów, socjologów, psychologów, ekonomistów, politologów i przedstawicieli innych dyscyplin naukowych – wybitnych specjalistów w swoich dziedzinach, o różnorodnych doświadczeniach zawodowych, które obejmują, oprócz badań naukowych, także pracę dydaktyczną, doświadczenie w administracji publicznej czy działalność w organizacjach pozarządowych.

Instytut w Polsce uczestniczy w realizacji międzynarodowych projektów badawczych w tym PIAAC, PISA, TALIS, ESLC, SHARE, TIMSS i PIRLS oraz projektów systemowych współfinansowanych przez Unię Europejską ze środków Europejskiego Funduszu Społecznego.

*Raport jest sprawozdaniem ze starannie zaplanowanych oraz nowatorskich w sensie teoretycznym i metodologicznym badań oceniających efekty nauczania początkowego oraz sondujących uwarunkowania szkolne i indywidualne efektów nauczania. Został zaplanowany jako całościowe sprawozdanie z pierwszego etapu badań o charakterze podłużnym wraz z omówieniem ogólnym założeń i wyników badania, a także prezentacją kwestii szczegółowych – znaczenia uzyskanych wyników dla poszczególnych zmiennych, uwzględnionych w badaniach.*

*Na szczególnie pozytywną ocenę zasługuje przyjęty wielozmiennowy plan badań podłużnych, umożliwiający wielopoziomą analizę danych (poziom indywidualny, klasowy i szkolny), który znalazł swoje odzwierciedlenie w planie prezentacji uzyskanych wyników badania. Założenia badawcze oraz sformułowane problemy i hipotezy zostały w raporcie klarownie przedstawione i dobrze uzasadnione, także na podstawie dostępnej literatury przedmiotu. Uzyskane wyniki zostały poddane zaawansowanym analizom statystycznym oraz dobrze udokumentowane w formie zestawień tabelarycznych oraz prezentacji graficznych. Uzyskane wyniki wnoszą cenny wkład do rodzimych badań psychologicznych i pedagogicznych, w tym zwłaszcza do pomiaru wyników nauczania (zaawansowana, nowatorska na polskim gruncie metodologia konstrukcji narzędzi do badania wyników nauczania), ale podkreślenia wymaga też waga praktyczna zrealizowanych badań – implikacje dla procesu organizacji nauczania czy polityki kształcenia.*

Recenzent: Prof. dr hab. Bogdan Zawadzki

## Instytut Badań Edukacyjnych

ul. Górczewska 8, 01-180 Warszawa | tel. +48 22 241 71 00 | [ibe@ibe.edu.pl](mailto:ibe@ibe.edu.pl) | [www.ibe.edu.pl](http://www.ibe.edu.pl)  
Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.