

**prof. dr hab. Roman Dolata**

Instytut Badań Edukacyjnych  
Zespół Edukacyjnej Wartości Dodanej

**dr Maciej Jakubowski**

OECD  
Uniwersytet Warszawski

**dr Artur Pokropek**

Instytut Filozofii i Socjologii PAN  
Zespół Edukacyjnej Wartości Dodanej

## **Nowe spojrzenie na pozycję Polski w międzynarodowych badaniach umiejętności uczniów**

### **Wprowadzenie**

Artykuł prezentuje wybrane wyniki przedstawione w przygotowywanej do druku książce<sup>1</sup> poddającej ponownej analizie wyniki międzynarodowych badań umiejętności uczniów, w których uczestniczy Polska (PISA, PIRLS). Przygotowana książka zawiera pięć rozdziałów. Rozdział 1. przedstawia oryginalne analizy wykorzystujące od nowa wyskalowane indywidualne wyniki polskich uczniów w badaniu PISA. Nowe skale, skorygowane o zmiany w próbach uczniów, a także zmiany w cechach rodzin uczniów, pokazują spójny obraz przemian umiejętności 15-latków w Polsce. W rozdziale 2 porównania międzynarodowe dopełnia analiza wykorzystująca dane PIRLS oraz nowatorska analiza przyrostu umiejętności uczniów w zakresie czytania między końcem klasy III szkoły podstawowej (wyniki badania PIRLS) a końcem gimnazjum (badania PISA). Rozdział 3 poświęcony jest analizie danych dla 16- i 17-latków w Polsce. Dane te zostały zebrane w ramach tzw. opcji narodowej badania PISA w roku 2006. W rozdziale 4 przedstawiono wyniki analiz wpływu różnych aspektów pochodzenia społeczno-ekonomicznego uczniów na wyniki w testach PISA. Rozdział 5 poświęcony jest analizie motywacji uczniów.

Badanie OECD PISA (*Programme for International Student Assessment* – Program Międzynarodowej Oceny Umiejętności Uczniów) to wielkie międzynarodowe badanie umiejętności i wiadomości uczniów realizowane w Polsce od samego początku istnienia tego przedsięwzięcia badawczego. W ostatniej edycji (2009) testy PISA rozwiązało niemal pół miliona uczniów, zarówno z 34 krajów OECD, jak i kilkudziesięciu krajów partnerskich, których wyniki są reprezentatywne dla populacji niemal 30 milionów 15-latków

---

<sup>1</sup> Artykuł ten opiera się na materiałach zawartych w książce, która powstała w ramach projektu badawczego „Analiza porównawcza wyników międzynarodowych badań umiejętności uczniów PISA w oparciu o semi-parametryczne metody dopasowania oraz hierarchiczne modele liniowe”, NN114 446336, zrealizowanego we współpracy z CASE – Centrum Analiz Społeczno-Ekonomicznych.

na całym świecie. Łącznie kraje uczestniczące w PISA wytwarzają ponad 80% światowego PKB. Badanie PISA realizowane jest w cyklu 3-letnim. Pierwsze badanie odbyło się w 2000 roku. W każdym z cykli oceniane są umiejętności w trzech obszarach: czytanie, matematyka i nauki przyrodnicze (angielski termin *science* obejmuje fizykę, chemię, biologię i geografę). W każdym cyklu jeden z obszarów ma rangę obszaru głównego i jest badany znacznie obszerniej. I tak w 2000 roku szczegółowej ocenie poddano czytanie, w 2003 matematykę, w 2006 nauki ścisłe, a w 2009 ponownie czytanie. W praktyce testowania oznacza to, że główna dziedzina oceniana jest na podstawie ponad 100 zadań testowych dotyczących wielu aspektów danej umiejętności, pozostałe zaś za pomocą znacznie mniej obszernych testów.

PISA wypracowała własną koncepcję badania umiejętności i wiadomości uczniów, które określane są wspólnym terminem *literacy* (tłumaczone dość dowolnie jako „rozumowanie” lub też „biegłość”). Koncepcja ta zakłada, że w badaniu mierzyć się będzie przyswojenie wiadomości i opanowanie umiejętności niezbędnych uczniom w życiu dorosłym, na rynku pracy i do tego, aby w pełni swobodnie funkcjonowali w społeczeństwie. Współcześnie we wszystkich międzynarodowych badaniach wiadomości i umiejętności, np. takich jak PIRLS czy TIMSS, metody pomiaru są podobne. To, co wyróżnia PISA, to właśnie definiowanie mierzonych umiejętności w perspektywie szerszej niż programy szkolne, w pewnym oderwaniu od tego, czego naucza się w szkole (programy szkolne stanowią główną wykładnię pomiaru dla PIRLS oraz TIMSS). Sam sposób definiowania mierzonych umiejętności to długotrwały proces konsultacji międzynarodowych ekspertów dokumentowany w osobnych publikacjach (tzw. *PISA Framework*). Ekspertci programu PISA twierdzą, że skonstruowane przez nich podejście daje podstawę do oceny umiejętności przydatnych w dorosłym życiu, u progu którego stoją 15-latkowie.

Wyniki badania PISA są ciekawe głównie w perspektywie porównawczej, ale fakt, że mamy już za sobą cztery edycje tego badania sprawia, że równie interesująca jest perspektywa analiz trendów czasowych w obrębie jednego systemu oświaty. Wnikliwe badanie trendów czasowych zostało zainicjowane przez Bank Światowy, zainteresowany przyczynami znaczącego wzrostu poziomu umiejętności polskich uczniów mierzonych w PISA w zakresie umiejętności czytania. Wzrost ten spostrzeżony został w wielu krajach i wzbudził chęć poznania przyczyn tak znacznego polepszenia się wyników polskich uczniów.

Dane PISA można analizować w dwóch, równie ważnych dla każdego systemu edukacyjnego, wymiarach. Pierwszy z nich dotyczy poziomu wiadomości i umiejętności uczniów. Mierzy się go przede wszystkim przez średni wynik uczniów w danym kraju, ale i przez procent uczniów, którzy osiągnęli wyróżniony poziom umiejętności. Drugi wymiar dotyczy nierówności edukacyjnych. Tutaj mierzone jest zróżnicowanie wyników, podawane przede wszystkim jako odchylenie standardowe ogółu wyników uczniów, ale i jako procent wariancji wyników wyjaśniany przez przynależność do szkoły (mierzy poziom segregacji uczniów między szkołami ze względu na poziom wiedzy) lub też moc zależności między osiągnięciami a statusem społeczno-ekonomicznym rodziny ucznia (im silniejsza, tym większe nierówności ze względu

na pochodzenie społeczne). Innym badaniem międzynarodowym obecnym w Polsce i wykorzystywanym w tym opracowaniu jest PIRLS. Stawia sobie za cel pomiar biegłości w czytaniu wśród dzieci mających za sobą czwarty rok nauki szkolnej. Pomiary PIRS odbywają się w cyklu 5-letnim, w roku 2006 przeprowadzono je w 45 krajach. W momencie pisania tej publikacji (w 2011 roku) realizowana jest (między innymi w Polsce), kolejna edycja tego badania.

Niemal wszystkie zebrane dane PISA i PIRLS publikowane są w ogólnie dostępnych bazach, co umożliwiła badaczom z całego świata prowadzenie niezależnych analiz. Bazy te stanowią niezwykle bogaty materiał, niestety, rzadko wykorzystywany w naszym kraju. Niewiele powstało prac badawczych analizujących wyniki i kontekst osiągnięć uczniów polskich. Raporty krajowe z każdej edycji badania PISA i PIRLS są publikowane przez polskie zespoły i stanowią dobry materiał opisowy prezentujący uzyskane w Polsce wyniki, ale brak szczegółowych, krytycznych analiz uzyskanych rezultatów.

W niniejszym artykule prezentujemy wybrane wyniki przedstawione w książce w rozdziale 1, które pokazują nowe spojrzenie na osiągnięcia polskich uczniów w badaniu PISA. W prezentowanych analizach wyniki polskich uczniów – celem bardziej precyzyjnego określenia trendów od 2000 roku – zostały na nowo wyskalowane, zastosowano również metody statystyczne przeważania danych służące zwiększeniu porównywalności wyników między czterema edycjami PISA, a także pozwalające wziąć pod uwagę różnice w statusie społeczno-ekonomicznym rodzin uczniów oraz udziale imigrantów w populacji uczniów w poszczególnych krajach.

## Skalowanie osiągnięć uczniów w badaniach międzynarodowych

Klasyczne oszacowanie umiejętności ucznia poprzez określanie liczby poprawnych odpowiedzi, podobnie jak w amerykańskim badaniu NAEP, które stworzyło podstawy metodologii takich badań jak PISA, PIRLS czy TIMSS, nie może być stosowane w badaniach międzynarodowych. Po pierwsze, w badaniach międzynarodowych używana jest metoda „rotating booklets”, w której każdy uczeń odpowiada jedynie na część z całej baterii zadań celem bardziej precyzyjnego określenia umiejętności uczniów w różnych dziedzinach w całej populacji. Po drugie, wynik każdego ucznia w klasycznej teorii testów wyrażony jest na skali tzw. dyskretnej, czyli wyniki pomiaru są liczbami całkowitymi. W wielu opracowaniach pokazano (np. VonDavier, Gonzalez, Mislavy, 2009), iż stosowanie klasycznych metod może powodować błędy w szacowaniu różnic w średnich wynikach między grupami (na przykład chłopców i dziewczynek, uczniów ze wsi i z miasta), czy też szacowaniu zależności między wynikami a interesującymi nas zmiennymi (na przykład wynikami a pochodzeniem społecznym ucznia, czyli zmiennymi wykorzystywanymi w niemal każdym badaniu edukacyjnym). Problemy te rozwiązuje uogólniony model Rascha z tzw. *plausible values* (dosłownie: model „prawdopodobnych wartości”). W modelu tym uczniom przypisywanych jest kilka prawdopodobnych wyników, zazwyczaj 5, które biorą pod uwagę nie tylko odpowiedzi na wszystkie zadania testowe, ale także związki tych odpowiedzi z cechami uczniów w całej populacji (Wu i in., 2007).

Ten dość skomplikowany model statystyczny dobrze odzwierciedla niepewność pomiaru. Wykazano, że posługując się pięcioma *plausible values*, można lepiej odtworzyć rozkład umiejętności w całej populacji. Dzięki użyciu tego modelu nie tylko średnie wyniki w całej populacji, ale i wyniki dla podgrup, wariancja tych wyników oraz dalsze analizy odnoszące osiągnięcia uczniów do ich innych cech, programów edukacyjnych itp. ukazują wartości bliskie prawdziwym. Co więcej, model *plausible values* szacuje wyniki dla uczniów, którzy rozwiązywali różne zadania testowe, dzięki uwzględnieniu cech tych uczniów, ale i tego, że każdy z zeszytów testowych posiada zadania wspólne z innymi zeszytami, a model statystyczny równocześnie bierze pod uwagę wszystkich uczniów i wszystkie zadania testowe<sup>2</sup>. Warto podkreślić, że ten sposób skalowania nie ma na celu oceniania umiejętności pojedynczych uczniów, a jedynie wartości dla całej populacji i odpowiednio licznych podgrup.

Sposób skalowania wyników PISA jest rozwiązaniem optymalnym w porównaniach międzynarodowych. Coraz częściej jednak wyniki badania PISA próbuje się wykorzystać do monitorowania zmian w osiągnięciach uczniów w poszczególnych krajach. Schemat skalowania stosowany w badaniu PISA z perspektywy tak postawionego celu nie jest schematem najlepszym. Nie wnikając we wszystkie szczegóły techniczne, które przedyskutowane są dokładniej w przygotowywanej książce, można powiedzieć, że podstawowym problemem w wykorzystaniu wyników badań międzynarodowych dla jednego kraju jest to, że cała metodologia nastawiona jest na uzyskanie wyników porównywalnych między krajami, a nie najlepiej opisujących dane zebrane w tylko jednym kraju. Ujmując rzecz bardzo skrótowo, wszystkie parametry w modelach statystycznych stosowanych na poszczególnych etapach metody skalowania wyników testów są dostosowywane do wartości średnich dla określonej grupy krajów (krajów OECD w badaniu PISA, wszystkich uczestników w badaniu PIRLS). Jeśli chcemy analizować wyniki tylko z jednego kraju, bardziej efektywne statystycznie będzie szacowanie parametrów tak, by najlepiej pasowały do danych w analizowanym zbiorze danych. Natomiast w analizach porównawczych wszystkie procedury statystyczne podporządkowane są głównemu celowi: porównywalności międzynarodowej.

Model w badaniu międzynarodowym optymalizuje dopasowanie parametrów zadań testowych dla wszystkich państw, przez co z perspektywy pojedynczego kraju rozwiązanie międzynarodowe nie musi być najlepsze. Ponadto skalowanie przy zachowaniu wymogu międzynarodowej porównywalności doprowadziło twórców badania do pewnych kompromisów: we wstępnym skalowaniu używa się tylko próbki uczniów, co zmniejsza precyzję oszacowania. Aby skalowanie dla próby składającej się z kilkudziesięciu krajów było wykonalne, stosuje się model Rascha, który obok wielu zalet posiada zasadniczą wadę: jego dopasowanie do danych opiera się na jednym parametrze (trudności), czyli z definicji model ten jest gorzej dopasowany niż model np. dwuparametryczny (Baker i Kim, 2004). Model dwuparametryczny jest jednak dużo bardziej wymagający obliczeniowo niż model Rascha. Kolejnym kompromisem jest sposób zrównywania wyników między latami. Zrównywanie w PISA opiera

<sup>2</sup> Dokładniej, model bezpośrednio szacuje wyniki w kilku dziedzinach (por. OECD, 2011, rozdział 9).

się na przekształceniu liniowym danych z kolejnych lat. W tej procedurze najpierw skaluje się wyniki z poszczególnych lat, potem liniowo przekształca się wyniki z poszczególnych lat tak, aby sprowadzić je do wspólnej skali. Bardziej efektywne zrównywanie, polegające na skalowaniu czterech edycji PISA w jednym modelu, jest praktycznie niewykonalne z powodów wielkości próby (por. OECD 2009; Gebhardt i Adams, 2007).

Ponieważ naszym celem jest uzyskanie jak najbardziej precyzyjnych wyników dla Polski, postanowiliśmy na nowo wyskalować wyniki badania PISA. Po pierwsze, przy analizie zmian wyników między latami wykorzystaliśmy tylko dane dla Polski, aby uzyskać dopasowania najlepsze z punktu widzenia naszego kraju. Po drugie, postanowiliśmy zastosować model dwuparametryczny, czyli taki, który pozwala na lepsze dopasowanie do danych niż model Rascha. Po trzecie, zrównanie wyników w kolejnych edycjach PISA postanowiliśmy osiągnąć poprzez jedną łączną kalibrację wszystkich zadań poczynawszy od 2000 roku (por. von Davier, 2007). Używając tych rozwiązań, uzyskaliśmy skalę lepiej dopasowaną do danych polskich, ale mogą być one podstawą do analizy trendów wyłącznie dla naszego kraju. Szczegółowy techniczny opis zastosowanych modeli przedstawiono w przygotowanej książce. Poniżej skupimy się jedynie na prezentacji głównych wyników i porównaniu z danymi oryginalnymi.

### **Korekta wyników ze względu na status społeczno-ekonomiczny rodziny ucznia**

W książce prezentujemy też wyniki skorygowane ze względu na zmiany w czasie charakterystyk społecznych uczniów. Korekta taka ma sens zarówno przy porównaniu wyników polskich uczniów w 2009 roku z innymi krajami, jak i przy analizie trendów. Przy porównaniach z innymi krajami korekta wyników ze względu na status społeczno-ekonomiczny rodzin uczniów pozwala uwzględnić to, że rozkłady zmiennych statusowych w krajach biorących udział w badaniu PISA znacząco się różnią. Wykorzystując wyniki PISA do oceny efektywności krajowych systemów edukacyjnych, warto wziąć pod uwagę cechy statusu społecznego uczniów. W pewnych analizach warto także wyłączyć z porównań imigrantów, bowiem ich wyniki zależą w dużej mierze od tego, w jakim kraju się urodzili i pobierali pierwsze nauki.

Z kolei korekta trendów o zmiany w czasie rozkładu cech statusowych ucznia pomaga w dokonaniu porównań, na które nie będą miały wpływu przemiany społeczne niezależne od szkół. Przykładowo, jeśli ogólnym trendem w populacji jest zdobywanie coraz wyższego wykształcenia, to szkoły będą miały niejako ułatwione zadanie z biegiem czasu, bowiem będą pracować z uczniami pochodzącymi z coraz lepiej wykształconych rodzin. Jeśli chcemy ocenić efektywność systemów edukacji, to tego rodzaju zmiany kontekstu powinniśmy wziąć pod uwagę.

Korekty wyników o zmienne kontekstowe można dokonać w dość prosty sposób, stosując odpowiednie modele statystyczne. Metody przez nas wykorzystane różnią się nieco zależnie od analizowanych danych (ich techniczny opis przedstawiono w książce), jednak zasada działania jest zawsze ta sama. W każdym przypadku prezentujemy wyniki oryginalne, gdzie populacje

mogą dowolnie różnić się między krajami pod względem statusu społeczno-ekonomicznego rodzin uczniów. Prezentujemy jednak też wyniki skorygowane, które dzięki odpowiednim zabiegom statystycznym pokazują, jak układałyby się wyniki pomiarów osiągnięć, gdyby rozkłady statusu społeczno-ekonomicznego rodzin uczniów były we wszystkich krajach identyczne z rozkładem cech dla badania PISA 2009 w Polsce. Polska w PISA 2009 stanowi punkt odniesienia do wszystkich porównań zarówno między różnymi krajami w 2009 roku, jak i do trendów dla Polski między pierwszą edycją PISA 2000 i ostatnią PISA 2009. Korekty dokonujemy metodami przeważania danych, gdzie oryginalne wagi są przeliczane tak, żeby uczniowie o charakterystykach rzadziej spotykanych w roku odniesienia uzyskali wagi nieco niższe<sup>3</sup>.

### **Główne wyniki badania PISA 2009**

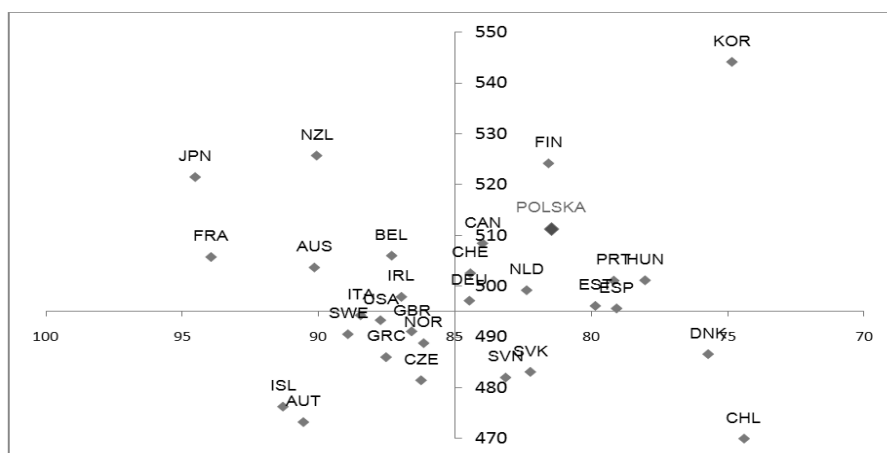
Główną domeną badania PISA 2009 było czytanie. Zgodnie z oficjalnymi raportami w tym zakresie osiągnięcia naszych uczniów uległy od 2000 roku znaczącej poprawie i lokują się aktualnie powyżej średniej OECD. Wciąż jednak są to wyniki znacznie poniżej osiągnięć w krajach najlepszych. Z jednej strony cieszy więc, że pod względem średniego wyniku jesteśmy bliscy Szwajcarii, Stanom Zjednoczonym, Niemcom czy Francji, ale równocześnie martwi spory dystans do krajów takich jak Korea czy Finlandia. W matematyce nasze wyniki są niemal identyczne jak średnia OECD i w znaczący sposób odstają nie tylko od wyników krajów najlepszych, ale i od wyników krajów nieco lepszych niż średnia. Jest tylko kilka krajów OECD, które osiągają wyniki w matematyce znacząco niższe od naszych. Z drugiej strony grupa krajów z porównywalnymi wynikami jest dość szeroka: jest w niej nie tylko Francja czy Norwegia, ale także Stany Zjednoczone i Irlandia. Oficjalne wyniki mówią, że w zakresie umiejętności w naukach przyrodniczych polskie piętnastolatki wypadają w 2009 roku lepiej niż w matematyce i trochę lepiej niż w czytaniu. Wynik jest lepszy niż średnia OECD, grupa krajów z porównywalnym poziomem umiejętności znacznie węższa i cieszą się osiągnięcia porównywalne z Wielką Brytanią czy Belgią, ale martwi przepaść dzieląca nas od Finlandii, Japonii lub Korei.

Trzeba zaznaczyć, że grupa krajów OECD nie jest jedynym możliwym punktem odniesienia dla Polski. Najwyższe wyniki w PISA 2009 uzyskał Szanghaj, występujący jako jeden z regionów Chin. Rezultaty w czytaniu wyższe od polskich uzyskali też uczniowie z Hongkongu i Singapuru. W matematyce do grupy krajów spoza OECD, które mają wyższe od polskich wyniki, dołącza poza wymienionymi wyżej także Tajwan, Makao oraz Liechtenstein. Oprócz ostatniego są to jak widać nie tylko kraje azjatyckie, ale i związane z kulturą chińską. Ze względu na dynamiczny rozwój chińskiej gospodarki wyniki tych krajów czy regionów Chin obserwowane są z ogromną uwagą i także powinny stanowić ważny punkt odniesienia dla porównań osiągnięć uczniów w Polsce.

<sup>3</sup> Zanim dokonano przeważania danych brakujące informacje o uczniach, a także brakujące wyniki uczniów, zostały oszacowane metodą „multiple imputation” (przystępne wprowadzenie do tych procedur można znaleźć w książce Paula Allisona, 2001). Dzięki temu wszystkie wyniki opierają się na pełnej próbie uczniów. Podobna metoda jest stosowana w PISA od 2003 roku, gdzie nawet uczniom, którzy nie odpowiadali na pytania testowe z danej dziedziny, która nie była główną dziedziną badania, przypisywanych jest kilka potencjalnych wyników (tzw. plausible values). Zarówno analizy teoretyczne, jak i symulacje pokazują, że stosowanie tej metody zwiększa precyzję oszacowań dla całej populacji. Metody przeważania opisuje Tarozzi (2007).

Inną ważną z punktu widzenia polityki edukacyjnej miarą jest zróżnicowanie wyników uczniów, dla którego podstawową statystyką jest odchylenie standardowe osiągnięć uczniów w każdym kraju. Odchylenie standardowe na skali PISA ustalono w 2000 roku dla początkowej grupy krajów jako 100 punktów, jednak w 2009 roku dla obecnych członków OECD wynosi ono w czytaniu 93. Odchylenie standardowe wyników uczniów w Polsce jest nieco niższe niż średnia, wynosi bowiem 89 punktów. Jest to niewątpliwie powód do zadowolenia, bowiem rezultat ten świadczy o nieco mniejszym zróżnicowaniu wyników polskich uczniów niż średnio w krajach OECD.

Warto spojrzeć równocześnie na dwa podstawowe aspekty rozkładu wyników testów PISA: średnią oraz zróżnicowanie wyników. Analizując wyniki Polski względem innych krajów jako wskaźnik efektywności naszego systemu edukacji, warto policzyć podstawowe statystyki po wyłączeniu wpływu statusu społeczno-ekonomicznego rodzin uczniów, a także po odrzuceniu wyników uczniów pochodzących z rodzin imigranckich pierwszego pokolenia (nieurodzonych w kraju zamieszkania) lub mówiących w domu w języku innym niż język testu (we wszystkich krajach OECD tacy uczniowie uzyskują niższe wyniki). Tak policzone rezultaty przedstawiono na wykresie 1. Na wykresie u góry po prawej znajdują się kraje o najwyższym średnim wyniku i najniższym ich zróżnicowaniu (odchylenie standardowe), a na dole, po lewej, kraje o najniższej średniej i największym zróżnicowaniu wyników PISA w zakresie czytania. Linia pozioma i pionowa oznaczają średnie OECD w 2009 roku.



**Wykres 1. Średni wynik w czytaniu i zróżnicowanie wyników uczniów po uwzględnieniu statusu społeczno-ekonomicznego rodziny ucznia i wyłączeniu imigrantów (kraje OECD, PISA 2009)**

Wyniki na wykresie 1. pokazują, że po uwzględnieniu różnic w statusie społeczno-ekonomicznym rodzin uczniów między krajami pozycja Polski jest nieco wyższa. Okazuje się, że zarówno pod względem średniego wyniku, jak i zróżnicowania jesteśmy blisko Kanady, kraju stawianego często za wzór systemu edukacyjnego, o nominalnie znacznie wyższych nakładach na edukację. Jesteśmy

też znacznie bliżej Finlandii, choć różnica dzieląca nas od krajów najlepszych, w tym krajów z zasięgu kultury chińskiej, pozostaje nadal znaczna. Nie można więc przewidywać, że z biegiem czasu, wraz z rozwojem społeczno-ekonomicznym kraju, a przez to podnoszeniem wykształcenia rodziców, wyniki Polski będą coraz bliższe wynikom najlepszych krajów. Poprawa efektywności nauczania jest niezbędna, o ile chcemy dogonić najlepsze systemy edukacyjne na świecie. Wyzwaniem dla nas będzie też zmierzenie się z zapewne dynamicznie wzrastającym w najbliższym czasie napływem imigrantów. Kraje takie jak Kanada potrafią nie tylko zapewnić wysoką jakość nauczania uczniom w tym kraju urodzonym, ale też imigrantom. Te wyzwania wciąż stoją przed polską edukacją.

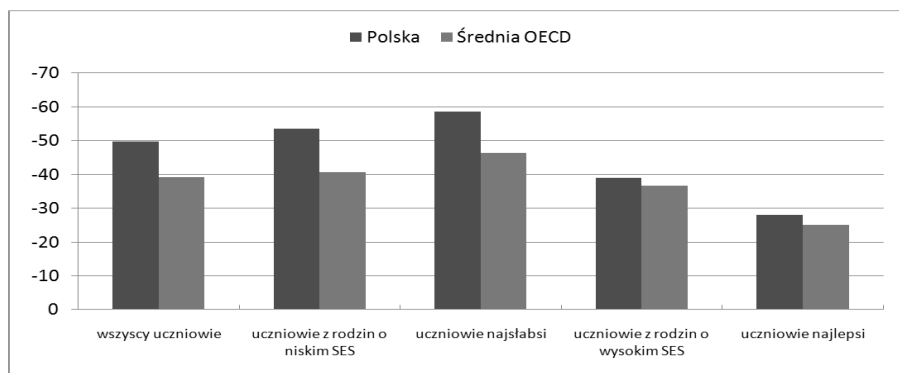
A analizach wyników należy też zwracać uwagę na osiągnięcia w podgrupach. Na przykład w umiejętności czytania widoczne są znaczne różnice w osiągnięciach chłopców i dziewcząt. W badanych krajach różne znaczenie ma także status społeczno-ekonomiczny rodziny ucznia. Choć uczniowie z rodzin o wyższym statusie uzyskują znacznie wyższe wyniki, niezależnie od tego, jak ten status mierzymy, to siła tego związku bywa w różnych krajach różna. Różnice między wynikami chłopców i dziewcząt, czy różnice w osiągnięciach uczniów z rodzin o najniższym i najwyższym statusie, bywają znacznie większe niż różnice w średnich między krajami.

Przyjrzymy się najpierw różnicom między wynikami chłopców i dziewcząt. W Polsce różnice w czytaniu ze zrozumieniem są nieco większe niż przeciętnie w krajach OECD. Różnica 50 punktów to pół odchylenia standardowego na skali PISA, czy też bardziej obrazowo, więcej niż różnica między wynikiem Polski a Finlandii lub Korei. We wszystkich krajach OECD chłopcy uzyskują niższe wyniki. Co więcej, różnice te są znacznie większe wśród uczniów najsłabszych, a w wielu krajach także wśród uczniów z rodzin o relatywnie niskim statusie społeczno-ekonomicznym.

Na wykresie 2. przedstawiono różnice w wynikach na skali czytania w Polsce i średnio w krajach OECD, osobno dla wszystkich uczniów, a także dla uczniów zdefiniowanych ze względu na status społeczno-ekonomiczny ich rodzin lub poziom uzyskiwanych wyników.

Wyniki polskich chłopców odstają od wyników dziewcząt w znacznie większym stopniu niż średnio w krajach OECD. Chłopcy uzyskali w 2009 roku 476 punktów, podczas gdy średni wynik dziewcząt wyniósł 525 punktów. Tak znaczną różnicę zaobserwowano tylko w kilku krajach OECD. Co więcej, różnica ta jest znacznie większa dla uczniów z rodzin o niskim statusie społeczno-ekonomicznym (z dolnej ćwiartki rozkładu statusu w Polsce) oraz wśród uczniów najsłabszych (z wynikami na poziomie 10 percentyla liczonego dla każdej z płci). Dla tych grup różnica wynosi odpowiednio 54 i 59 punktów, znacznie więcej od średniej OECD wynoszącej odpowiednio 41 i 46 punktów. Jeśli jednak spojrzymy na uczniów z rodzin o najwyższym statusie społeczno-ekonomicznym (z górnej ćwiartki rozkładu statusu w Polsce) lub uczniów najlepszych (z wynikami na poziomie 90 percentyla), to różnica między chłopcami i dziewczętami spada do odpowiednio 39 lub 28 punktów. Różnice dla tych grup uczniów są już bardzo podobne do średniej dla OECD.





**Wykres 2. Różnica między wynikami chłopców i dziewcząt w czytaniu ze zrozumieniem**

Z tej prostej analizy wynika, że różnice między chłopcami i dziewczętami są w Polsce większe niż przeciętna OECD, ale jest to związane głównie ze znacznie słabszymi wynikami chłopców z rodzin o niskim statusie społeczno-ekonomicznym, czy też z grupy uczniów najłabszych. Wśród uczniów najlepszych różnice te są także widoczne, ale są znacznie mniejsze i są na poziomie średniej w OECD.

Do kwestii zróżnicowania wyników wrócimy jeszcze w tym artykule, bowiem dla Polski bardziej interesujące od obecnych rezultatów są zmiany po 2000 roku.

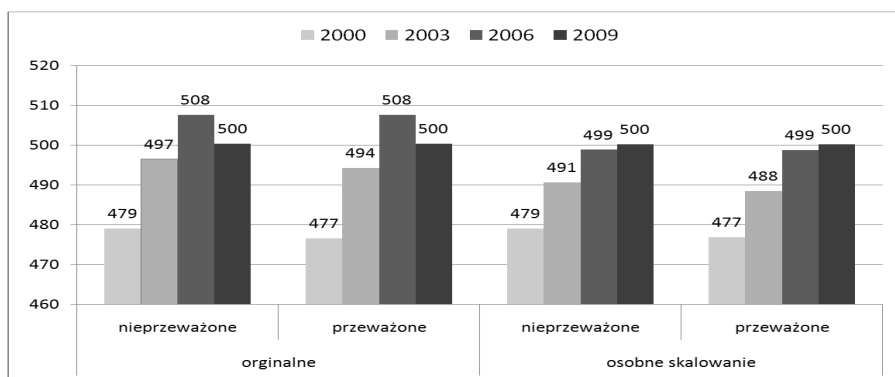
### Zmiany w poziomie umiejętności mierzonych w PISA w latach 2000-2009

Po raz pierwszy w historii badania PISA w raporcie OECD PISA 2009 poświęcono jedną z części wnikliwej analizie trendów w wynikach z czytania. Było to możliwe dzięki temu, że zarówno w pierwszej edycji PISA w 2000 roku, jak i w edycji 2009 czytanie ze zrozumieniem było główną ocenianą dziedziną. Między 2000 a 2009 rokiem właściwie nie zmieniono koncepcji teoretycznej testu PISA (tzw. assessment framework), a także utrzymano około 40 identycznych pytań, aby zapewnić porównywalność wyników.

W części tej prezentujemy nie tylko wyniki opierające się o skalę wykorzystywane w raportach OECD i raporcie krajowym PISA, ale także wyniki obliczone przez nas od nowa tylko dla danych polskich. Ze względów metodologicznych, wyniki te mogą dawać lepszy obraz zmian w osiągnięciach polskich uczniów. Różnice wynikają z potrzeby dostosowania modeli statystycznych stosowanych do skalowania odpowiedzi testowych uczniów do średnich parametrów dla wszystkich krajów. O ile zapewnia to porównywalność wyników między krajami w każdej edycji PISA, to może dawać mniej precyzyjne oszacowania dla zmian wyników w danym kraju w czasie. Statystyki pokazują, szczególnie dla lat 2003 i 2006, w których pomiar czytania był w PISA mniej dokładny (mniej zadań), że trendy dla Polski są mniej spójne na skalach międzynarodowych niż na skalach policzonych przez nas tylko dla danych polskich. Z tego względu prezentujemy podstawowe wyniki dwukrotnie. Aneks techniczny (dołączony do książki) dokładnie opisuje wykorzystane przez nas metody skalowania i w jaki sposób różnią się one od metod oryginalnych.

Przy porównaniach w czasie należy także wziąć pod uwagę zmiany w rozkładach charakterystyk społecznych uczniów. Korekta trendów o zmiany kontekstowe pomaga w dokonaniu porównań, na które nie będą miały wpływu przemiany niezależne od szkół. Przykładowo, jeśli ogólnym trendem w populacji jest zdobywanie coraz wyższego wykształcenia, to szkoły będą miały niejako ułatwione zadanie z biegiem czasu, bowiem będą pracować z uczniami pochodzącymi z coraz lepiej wykształconych rodzin. Jeśli chcemy ocenić efektywność pracy szkół, to tego rodzaju zmiany kontekstu powinniśmy wziąć pod uwagę.

Wykres 3. prezentuje wyniki dla Polski z zakresie umiejętności czytania we wszystkich edycjach PISA. Wyniki podane są na czterech różnych skalach. Wyniki po lewej stronie opierają się na oryginalnych wynikach, w których w skalowaniu odpowiedzi uczniów zastosowano średnie parametry dla wszystkich krajów. Wyniki zaprezentowano z oryginalnymi wagami, ale także po przeważeniu celem zwiększenia porównywalności względem cech kontekstowych (wykształcenia i zawodu rodziców uczniów, płci oraz wieku). Po prawej stronie zaprezentowano wyniki na skali stworzonej od nowa na podstawie odpowiedzi polskich uczniów na pytania testowe, bez dostosowania parametrów do innych krajów. Ponownie, pierwszy zbiór wyników opiera się na oryginalnych wagach, a drugi prezentuje wyniki po przeważeniu celem zwiększenia porównywalności ze względu na różnice rozkładach cech społecznych rodzin uczniów.



**Wykres 3. Osiągnięcia uczniów w Polsce w czytaniu ze zrozumieniem we wszystkich edycjach PISA**

Wyniki na wykresie 3. pokazują, że – niezależnie od skali – polscy 15-latkowie podnieśli znacząco swoje osiągnięcia w porównaniu z rokiem 2000. Różne zestawy wyników sugerują jednak inne tempo poprawy osiągnięć uczniów.

Oryginalne wyniki sugerują znaczną poprawę między 2000 a 2003 rokiem, a więc zaraz po reformie systemu edukacji. Wyniki na nowo policzonych skalach dla Polski pokazują, że zmiana nastąpiła stopniowo, z niemal równym przyrostem między rokiem 2000 i 2003 oraz między 2003 i 2006 rokiem.

Uważamy, że do analizy trendów w Polsce wyniki wyskalowane tylko na podstawie danych z polskich badań i zważone z uwzględnieniem zmian w czasie rozkładów podstawowych charakterystyk społecznych uczniów, są bardziej

wiarygodne. Lepsze właściwości na nowo policzonych skal pokazuje bardziej stabilny trend w latach 2000 -2006 oraz brak trudnego do zrozumienia obniżenia wyników w 2009 roku. Wielu interpretowało ten ostatni efekt jako spadek umiejętności po początkowym sukcesie reformy z 2000 roku, jednak według nas jest to artefakt wynikający z zastosowania wspólnego skalowania dla wszystkich krajów OECD.

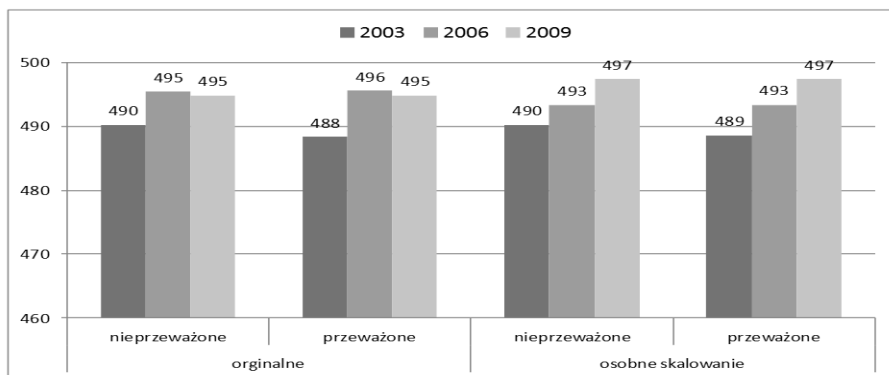
Skale policzone osobno dla Polski pokazują znacznie bardziej spójny i przekonujący trend czasowy w wynikach. Reforma 2000 roku, która bezpośrednio wpłynęła na sytuację testowanych w PISA 15-latków, miała zapewne znaczący wpływ dla podniesienia ich wyników w 2003 roku. W 2000 roku testom poddani zostali uczniowie pierwszych klas szkół średnich. Szczególnie uczniowie szkół zawodowych uzyskali bardzo niskie wyniki, co wpłynęło na niską średnią dla całego kraju. W 2003 roku testom poddano już uczniów ostatnich klas gimnazjów, niedługo przed egzaminem gimnazjalnym. Niewątpliwie pokazali oni średnio wyższy poziom umiejętności, szczególnie w przypadku uczniów, którzy w poprzednim systemie trafiliby do szkół zawodowych. Wygląda też jednak na to, że pozytywne efekty reformy trwały, co widać w poprawie między edycjami 2003 i 2006. To może być efekt zmian wprowadzonych wraz z reformą ustroju szkolnego, które „potrzebowały czasu”, aby pokazać swój pozytywny wpływ. Między 2006 i 2009 nie obserwujemy żadnej zmiany, co pokazuje, że system nauczania gimnazjalnego ustabilizował się po pozytywnym – ale też wymagającym dostosowań – szoku, jakim była reforma. Należy też wyraźnie powiedzieć, że zauważalna poprawa wyników polskich piętnastolatków w badaniu PISA nie jest dowodem na rzecz tezy, że rozdzielenie etapu powszechnego i jednolitego kształcenia na dwa etapy 6 + 3 jest bardziej efektywnym rozwiązaniem w stosunku do ciągłego 9-letniego cyklu kształcenia. Pamiętajmy też, że znaczący, dobrze zdiagnozowany progres dotyczy umiejętności czytania.

Niewielka różnica między wynikami nieprzeważonymi a przeważonymi pokazuje, że na wzrost wyników uczniów między 2000 a 2009 rokiem niewielki wpływ miały zmiany w rozkładach charakterystyk społecznych uczniów. Na pozytywny trend nie miały wpływu niewielkie zmiany w próbach dobieranych do badania PISA (wiek uczniów oraz procent dziewcząt nie zmieniły się znacząco), ale przede wszystkim tylko nieznaczny wpływ miał wzrost wykształcenia oraz statusu zawodowego rodziców uczniów. W interpretacjach wyników PISA pojawiały się komentarze, że pozytywny trend wśród polskich uczniów to głównie efekt poprawy jakości wykształcenia rodziców. Widać jednak, że nawet skorygowanie danych z 2000 roku tak, aby rozkład wykształcenia i zawodów rodziców był identyczny jak w roku 2009, nie zmienia znacząco konkluzji, a nawet sugeruje że poprawa jest nieco większa.

W ramach badania PISA analizuje się także trendy w poziomie umiejętności z matematyki i przedmiotów przyrodniczych. Jednak w tym przypadku dokładny pomiar miał miejsce w roku 2003 dla matematyki i w 2006 roku dla nauk przyrodniczych, a powtórzenia będą miały miejsce w latach 2012 i 2015, kiedy to odpowiednio matematyka i nauki przyrodnicze będą ponownie stanowiły główne dziedziny badane przez PISA. Z tego względu porównania z 2009 rokiem są mniej precyzyjne, bowiem pomiar tych dziedzin opiera się

na mniejszej liczbie pytań testowych i mniejszej liczbie uczniów. Dlatego też trendy dla matematyki i nauk ścisłych omawiamy mniej szczegółowo.

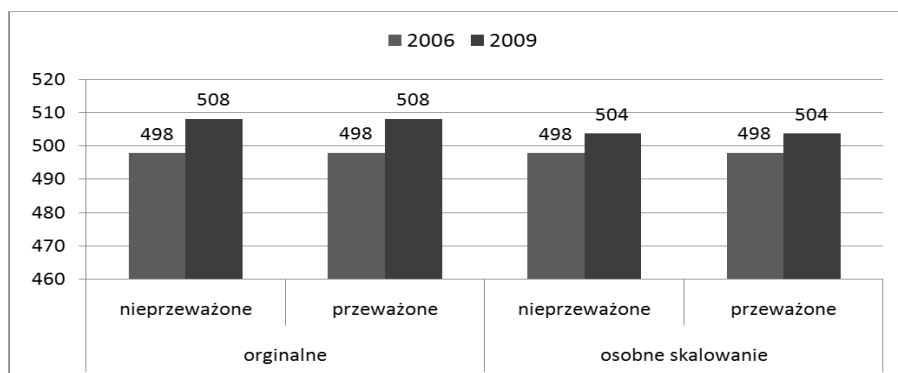
Wykresy 4. oraz 5. pokazują zmiany wyników polskich uczniów w PISA dla matematyki oraz nauk przyrodniczych. Matematykę można porównywać przez trzy edycje, a nauki przyrodnicze przez dwie. Ponownie wyniki przedstawiono na czterech różnych skalach, uzyskanych identycznymi metodami jak w przypadku czytania. W tym wypadku spojrzenie na nowo policzone i przeważone rezultaty, które według nas są bardziej wiarygodne, jeśli chodzi o trendy w Polsce, daje nieco inny obraz niż wyniki oficjalne.



**Wykres 4. Osiągnięcia uczniów w Polsce w matematyce od 2003 roku**

W matematyce oryginalne dane sugerują niewielką, nieistotną statystycznie zmianę w wynikach uczniów między rokiem 2003 a rokiem 2006, a także brak zmiany między 2006 i 2009. Wyniki na nowo skalowane na podstawie odpowiedzi tylko polskich uczniów sugerują jednak niewielki przyrost między 2003 a 2009, podobny do wzrostu w przypadku umiejętności czytania. Dla wyników skalowanych tylko dla Polski i przeważonych ze względu na zmiany w czasie cech społecznych uczniów, które – powtórzmy – są według nas najbardziej wiarygodne, różnica między 2003 i 2009 rokiem wynosi 12 punktów na skali PISA i jest istotna statystycznie. Pokazuje to, że wzrost umiejętności polskich uczniów także dokonał się w matematyce, choć wymogi metodologiczne badania międzynarodowego nie pozwoliły na jego „odkrycie” w raportach OECD.

Wykres 5. pokazuje zmiany w osiągnięciach w zakresie przedmiotów przyrodniczych. Dla tego obszaru porównania możliwe są jedynie między 2006 i 2009 rokiem. Wyniki PISA sugerowały niewielki, ale istotny statystycznie wzrost w osiągnięciach polskich uczniów. Ten wynik był dość zaskakujący, bowiem zarówno w czytaniu, jak i w matematyce nic nie sugerowało zmiany w poziomie polskich uczniów w ostatnich trzech latach. Wysuwano więc hipotezy na temat poprawy nauczania w naukach przyrodniczych lub opóźnionego efektu reformy w tych dziedzinach. Dane wyskalowane na nowo tylko dla polskich uczniów pokazują jednak, że zmiana jest znacznie mniejsza i nieistotna statystycznie, a więc nie można tu stwierdzić, że nastąpiła poprawa.



Wykres 5. Osiągnięcia uczniów w Polsce w naukach przyrodniczych od 2006 roku

Podsumowując analizę trendów opierającą się na nowo, tylko dla Polski wyskalowanych wynikach, można powiedzieć, że ukazuje ona bardzo spójny proces zmian w osiągnięciach polskich uczniów. Przeważone i przeskalowane dane sugerują stopniowy przyrost wiedzy we wszystkich dziedzinach między latami 2000, 2003 i 2006, a także brak zauważalnych zmian między 2006 i 2009 rokiem. Choć główne konkluzje opierające się o oryginalne wyniki OECD są podobne, to nowe wyniki zaprezentowane powyżej wyjaśniają wątpliwości dotyczące tempa zmian i trudnej do wyjaśnienia zmiany między 2006 i 2009 rokiem.

Warto spojrzeć na zmiany w wynikach Polski na tle innych krajów. W tych porównaniach wykorzystujemy oryginalne wyniki PISA/OECD, które są porównywalne między krajami<sup>4</sup>. Wyniki w czytaniu można porównać między 2000 i 2009 rokiem dla 26 krajów OECD oraz 38 krajów ogółem, które uczestniczyły zarówno w pierwszej, jak i ostatniej edycji PISA i gwarantują dostateczną rzetelność porównań<sup>5</sup>. Polska należy do grupy 9 krajów, w których wyniki wzrosły o 15 punktów na skali PISA lub więcej. Najwyższy wzrost odnotowały kraje o wynikach znacznie poniżej średniej OECD. Kraje takie jak Albania, Indonezja czy Brazylia wciąż są daleko w tyle mimo poprawy wyników w ostatniej dekadzie. Kraje, takie jak Chile, Łotwa czy Izrael, zbliżyły się znacząco do przeciętnych wyników krajów OECD. Portugalia, Liechtenstein i Niemcy mają wyniki na poziomie średniej OECD, a Polska jako jedyny kraj z tej grupy „przeskoczyła” z grupy krajów poniżej średniej OECD do grupy krajów powyżej. Korea i Węgry to dwa kraje, których wyniki sugerują poprawę, jednak zmiana ta może też wynikać jedynie ze zmian w losowaniu uczniów, czy też poprawie wykształcenia rodziców. W każdym razie zmiany średnich wyników są dla tych dwóch krajów niewielkie, podobnie jak zmiany dla Niemiec.

<sup>4</sup> Por. OECD, 2010.

<sup>5</sup> Z krajów OECD, dane dla Holandii oraz Wielkiej Brytanii wycofano z publikacji o trendach ze względu na zbyt niską stopę realizacji próby badawczej w 2000 roku, a przez to brak porównywalności z kolejnymi edycjami PISA. Dane dla Luksemburga z 2000 roku zostały także wycofane ze względu na problemy z wersjami językowymi testów w 2000 roku. Dane dla Austrii z 2009 roku nie zostały porównane z 2000 rokiem, bowiem PISA 2009 była realizowana w trakcie protestów uczniów, co mogło wpłynąć na porównywalność wyników między latami.

Wzrost wyników w Polsce zauważany był i podkreślany w czasie oficjalnych prezentacji wyników PISA w wielu krajach. Polska jest bowiem jednym z niewielu krajów, którym udało się w stosunkowo krótkim czasie tak podnieść swoje wyniki i osiągnąć poziom znacznie bliższy krajom najlepszym. Sukces ten wiązany jest, zapewne słusznie, z reformą edukacji z 2000 roku<sup>6</sup>. Efekty reformy widoczne są zresztą nie tylko w poprawie średnich wyników, ale i znacznym zmniejszeniu różnic w wynikach między szkołami. Podobnych reform dokonano też na Łotwie, gdzie dodatkowym efektem było zmniejszenie różnic w wynikach uczniów o różnym statusie społeczno-zawodowym rodzin. Dla licznej grupy krajów średnie wyniki nie zmieniły się, co z jednej strony potwierdza stabilność narzędzia pomiaru umiejętności stosowanego w PISA, a z drugiej pokazuje, jak trwale są różnice w umiejętnościach uczniów w różnych krajach i sugeruje, że systemy edukacyjne różnią się efektywnością w sposób trudny do modyfikacji. Brak zmian jest o tyle intrygujący, że średnio w OECD wydatki na ucznia wzrosły o kilkadziesiąt procent, a w niektórych krajach, np. USA czy Australii, na przestrzeni ostatnich 20 lat wydatki te rosą nieustannie i są na poziomie najwyższym na świecie. Mimo zwiększających się nakładów kraje te nie odnotowują zmian w wynikach i generalnie nie ma bezpośredniej zależności między zmianą w nakładach a zmianą w wynikach PISA. Przeciętnie, dla 26 krajów OECD, które posiadają porównywalne dane z 2000 oraz 2009 roku, średnie wyniki pozostały bez zmiany.

Mniej liczna jest grupa krajów, których średnie wyniki się obniżyły. Po korekcie na zmiany w losowanych próbach czy statusie społeczno-ekonomicznym rodzin uczniów potwierdzono spadek wyników w Irlandii, Szwecji oraz Australii. Kraje te należały do grupy z wynikami powyżej średniej w 2000 roku. Australia nadal ma wyniki powyżej większości uczestników PISA, jednak już Irlandia spadła z grona najlepszych krajów do grupy przeciętnych. Spadek wyników w Irlandii jest wyjątkowo duży i nie wiąże się jedynie ze znacznym spadkiem wyników imigrantów<sup>7</sup>. Uczniowie miejscowi także odnotowali znaczny spadek wyników.

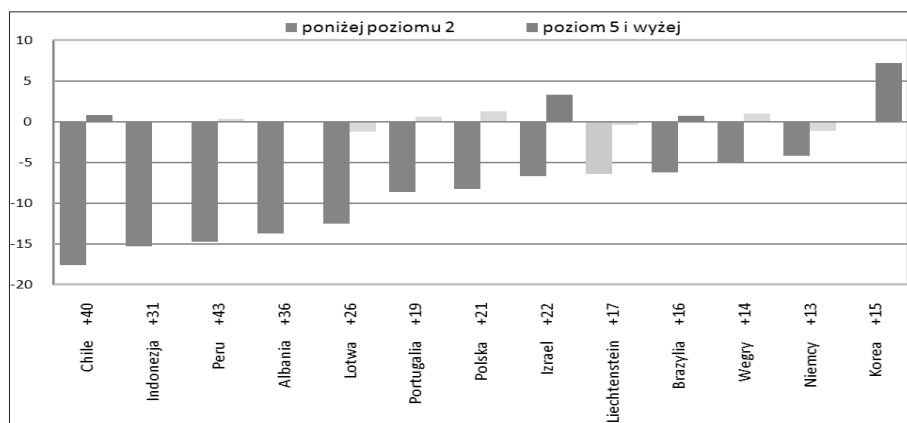
Po korekcie wyników ze względu na zmiany w doborze prób uczniów oraz statusie społeczno-ekonomicznym ich rodzin, spadek wyników można stwierdzić także w Finlandii, Hiszpanii i Tajlandii. Finlandia ma jednak nadal najwyższy wynik w grupie krajów, dla której możliwe jest porównanie między 2000 i 2009 rokiem, mimo niewielkiego spadku. Hiszpania natomiast spadła do grupy krajów europejskich z wynikiem znacząco poniżej średniej OECD. Czechy są krajem, w którym oryginalne dane sugerują spadek, jednak po korekcie zmiana wyników jest znacznie mniejsza i nieistotna statystycznie.

<sup>6</sup> Por. Jakubowski i in., 2010.

<sup>7</sup> Faktem jest, że w 2000 roku w Irlandii imigranci uzyskali wyniki wyższe od uczniów miejscowych, co – jak można się domyślać – było spowodowane tym, że w dużej mierze byli to uczniowie z zamożnych i dobrze wykształconych rodzin amerykańskich lub z Wielkiej Brytanii, które sprowadziły się do Irlandii ze względu na dynamiczny wzrost gospodarczy. Kryzys gospodarczy, jaki dotknął Irlandię niezwykle boleśnie, przed 2009 rokiem spowodował odpływ tej grupy „imigrantów”. W szkołach coraz większą liczbę uczniów stanowili imigranci z naszego rejonu, w tym w dużej mierze Polacy. Niestety, Irlandia nie gromadzi danych o pochodzeniu uczniów imigranckich, więc są to tylko domysły objaśniające spadek wyników imigrantów zmianami narodowościowymi w tej grupie. Grupa uczniów nieposługujących się językiem angielskim w domu znacząco jednak wzrosła i ma znacznie słabsze wyniki niż uczniowie miejscowi. Jak wspomniano, nie jest to jednak jedyna przyczyna spadku wyników w całym kraju.

Jednak spadki w wynikach czeskich uczniów w matematyce i naukach przyrodniczych potwierdzają, że osiągnięcia 15-latków u naszego sąsiada są coraz niższe, już znacznie poniżej średniej OECD.

Interesujące jest dokładniejsze spojrzenie na kraje, które poprawiły swoje wyniki w czytaniu. Okazuje się, że w większości z nich znaczna poprawa wyników jest spowodowana poprawieniem osiągnięć wśród uczniów najsłabszych. Wykres 6. pokazuje zmiany w procencie uczniów poniżej poziomu 2. w 13 krajach, które podniosły swoje średnie wyniki, a także zmiany w procencie uczniów na najwyższych dwóch poziomach: 5. i 6<sup>8</sup>. Jak widać, we wszystkich tych krajach, oprócz Korei, nastąpiło zmniejszenie odsetka uczniów o umiejętności czytania poniżej podstawowego, akceptowalnego we współczesnym świecie poziomu. Jednak jedynie w Korei, a także w niewielkim stopniu w Izraelu, wzrósł procent uczniów posiadających zaawansowane umiejętności czytania ze zrozumieniem. Te zmiany są o tyle oczywiste, że większość krajów, które poprawiły swoją pozycję w czytaniu, to kraje poniżej średniej OECD, ze znaczną liczbą uczniów o osiągnięciach poniżej poziomu 2. i niewielką liczbą na poziomach najwyższych. Jednak przykład Polski i innych krajów z wynikami w okolicach średniej, pokazuje, że także w krajach, w których znaczna liczba uczniów osiągnęła wyższe poziomy umiejętności, zdecydowanie „łatwiej” jest uzyskać ogólny przyrost, poprawiając wyniki uczniów najsłabszych. Korea jest więc krajem wyjątkowym także pod tym względem, że jako jedyna poprawiła swój już wysoki wyniki, jeszcze bardziej podnosząc liczbę uczniów na najwyższych poziomach.



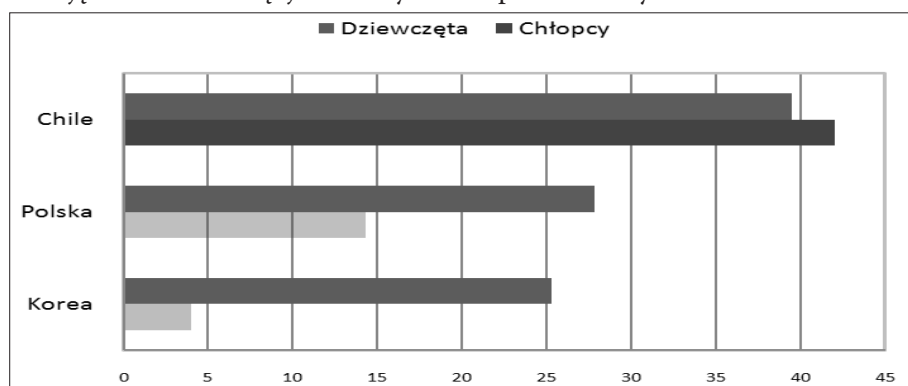
**Wykres 6. Zmiany w procencie uczniów na najniższych i najwyższych poziomach umiejętności w krajach, które poprawiły wyniki z czytania między 2000 a 2009 rokiem (wzrost średniej przy nazwach krajów)**

Interesujące jest także spojrzenie osobno na poprawę wyników chłopców i dziewcząt. Chile jest jedynym krajem OECD, w którym poprawie uległy wyniki chłopców (w Polsce zmiana wyniosła 14 punktów, lecz wynik jest na krawędzi istotności statystycznej). Za to aż w 7 krajach OECD wyniki chłopców uległy

<sup>8</sup> Podobne rezultaty podano w publikacji OECD: [www.oecd.org/dataoecd/32/53/47271471.pdf](http://www.oecd.org/dataoecd/32/53/47271471.pdf).

pogorszeniu. Wyniki dziewcząt poprawiły się w 7 krajach OECD, a spadły tylko w dwóch (w Szwecji oraz Irlandii). W znacznej liczbie krajów zmalała liczba dziewcząt z umiejętnościami na najniższym poziomie, a wśród chłopców liczba krajów, gdzie procent ten zmniejszył się i zwiększył, jest podobna. Średnio w krajach OECD różnica między chłopcami i dziewczętami powiększyła się, a liczba chłopców nieposiadających podstawowych umiejętności wzrosła.

W wielu krajach mamy zupełnie inny obraz zmian w umiejętnościach 15-latków, patrząc na wyniki chłopców i dziewcząt. Pokazuje to wykres 7. dla trzech interesujących przypadków. W Chile chłopcy poprawili swoje osiągnięcia w niemal identycznym stopniu, osiągając największy wzrost wśród wszystkich krajów. Co więcej, Chile należy do krajów, w których różnice w wynikach między płciami są jedne z najmniejszych. W Polsce zmiana w wynikach chłopców była dwa razy mniejsza niż w wynikach dziewcząt, na granicy istotności statystycznej. Pokazuje to, że o ile reforma edukacji przyczyniła się do poprawy wyników polskich 15-latków, to jednak więcej z tej poprawy zawdzięczamy wzrostowi wyników dziewcząt. Jak już powiedzieliśmy, różnica między polskimi chłopcami i dziewczętami należy do jednych z największych, a nierównomierna poprawa wyników oznacza też, że ta różnica dodatkowo wzrosła. Pokazuje to, że zarówno zmiany wprowadzone w reformie, jak i ogólnie nasz system nauczania, nie sprzyjają podnoszeniu umiejętności czytania przez chłopców. Ostatni przykład to Korea, gdzie obserwujemy opisywany już kilkakrotnie spektakularny wzrost liczby uczniów najlepszych. Jednak i tu efekt ten jest widoczny jedynie wśród dziewcząt. Chłopcy w Korei osiągnęli w 2009 roku wyniki niemal identyczne z wynikami z 2000 roku, podczas gdy przyrost umiejętności dziewcząt jest rzeczywiście spektakularny.



Wykres 7. Zmiany w osiągnięciach w czytaniu ze zrozumieniem wśród chłopców i dziewcząt

Jak już pisaliśmy, porównania wyników w matematyce i naukach przyrodniczych są mniej precyzyjne, bowiem w roku 2009 dziedziny te były testowane z mniejszą dokładnością niż czytanie. Wśród krajów, które poprawiły swój wynik w matematyce, jedynie Niemcy osiągają wyniki powyżej średniej OECD. Meksyk, Turcja i Grecja znacząco odstają od średniej, przy czym w Meksyku ocenieni zostali jedynie 15-latkowie uczęszczający do szkół, a więc wynik ten



jest z pewnością zawyżony względem całej populacji. Z drugiej strony liczba 15-latków w szkołach wzrasta w Meksyku z rok na rok, poprawa wyników świadczy więc o niewątpliwym wzroście jakości nauczania. Turcja jest krajem, który także odstaje od pozostałych członków OECD pod względem zamożności i rozwoju społeczno-ekonomicznego, jednak poprawa wyników dobrze ilustruje, jak ważne jest mądre inwestowanie w edukację. Jest to kraj, gdzie w ostatniej dekadzie upowszechniono szkolnictwo podstawowe, zmodernizowano tysiące szkół, wspierając je, dzięki ulgom podatkowym, inwestycjami prywatnych przedsiębiorców, wprowadzono nowe programy nauczania, a także przeprowadzono szeroką kampanię na rzecz kształcenia dziewcząt. Znaczna poprawa wyników sugeruje, że były to przedsięwzięcia udane.

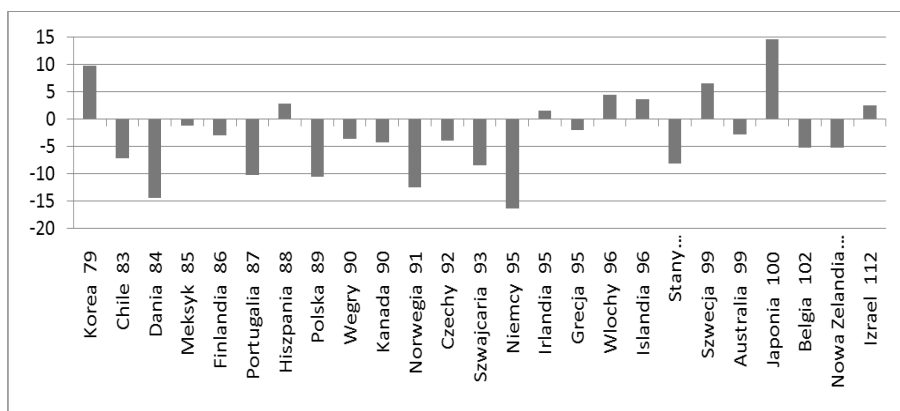
Portugalia jest jedynym krajem z wynikiem w okolicach średniej OECD, który poprawił się we wszystkich dziedzinach. Ten pozytywny trend trudno połączyć z jedną pojedynczą zmianą, bowiem ostatnia dekada to w Portugalii czas wielu mniejszych i większych reform edukacji. Za wzrostem wyników uczniów niewątpliwie stoi zmiana nastawienia do edukacji na wyższym poziomie jako dobra dostępnego jedynie wąskiej grupie. Wydłużenie obowiązkowego kształcenia powszechnego, wprowadzenie powszechnych egzaminów zewnętrznych, głębsza ewaluacja pracy nauczycieli to najważniejsze zmiany, które częściowo już wprowadzono, a częściowo wprowadza się jeszcze obecnie. Niewątpliwie jednak ostatnia dekada jest czasem zmiany w portugalskiej oświacie i ten pozytywny ferment może stać za poprawą wyników. Z drugiej strony, wyniki te są wciąż poniżej europejskiej średniej, podobnie zresztą jak poziom wykształcenia Portugalczyków (mniej niż 30% posiada wykształcenie średnie).

Ciekawe trendy obserwujemy u naszych zachodnich sąsiadów. PISA 2000 była dla Niemców szokiem, bowiem kolejne badanie potwierdziło, że ich wyniki są niższe niż średnia OECD, nie mówiąc już o oczekiwaniach społeczeństwa, przekonanego o wyższości niemieckiego modelu kształcenia. Dodatkowo PISA pokazała, że nierówności w niemieckim systemie edukacji są jednymi z największych w krajach OECD. PISA 2000 zapoczątkowała prawdziwy boom na badania edukacyjne w Niemczech, gdzie obecnie prowadzone są prace wdrażające kompleksowy system ewaluacji osiągnięć w całym kraju i na wszystkich etapach nauczania. Z badań wynika, że system niemiecki jest szczególnie niekorzystny dla imigrantów, szczególnie, jeśli mają oni problem z językiem niemieckim. Wprowadzono dodatkowe programy nauczania niemieckiego, a także działania wspomagające imigrantów oraz uczniów z rodzin niezamożnych. Wydaje się, że pierwsze pozytywne efekty tych projektów już można zauważyć w trendach PISA, choć należy się spodziewać dalszych zmian w najbliższym czasie. PISA zapoczątkowała też dyskusję nad sensownością systemu, w którym uczniów bardzo wcześnie dzieli się na ścieżkę zawodową i akademicką. Wiele landów już opóźniło ten podział, a kolejne planują reformy w tym zakresie. PISA sugeruje bowiem, że kraje z tak wczesną selekcją uczniów mają znacznie większe zróżnicowanie wyników, szczególnie przez niskie osiągnięcia uczniów najsłabszych i z nieuprzywilejowanych rodzin.

## Trendy w zróżnicowaniu wyników

Niewątpliwie istotne jest spojrzenie na zmiany nie tylko w średnich wynikach, ale także ich zróżnicowaniu. W tej części ograniczymy się do analiz zmiany w zróżnicowaniu wyników z czytania między pierwszą i ostatnią edycją PISA, bowiem analiza wariacji wyników i zależności ze statusem społeczno-ekonomicznym uczniów wymaga bardziej dokładnego pomiaru umiejętności, którego dokonano w PISA dwukrotnie jak dotąd tylko w czytaniu. Dla Polski prezentujemy też wyniki skorygowane, pokazujące zmiany w wynikach uczniów skalowane bez konieczności dostosowania parametrów do innych krajów, a także przeważone, korygujące dane z 2000 roku tak, aby rozkłady cech kontekstowych (wiek, płci, wykształcenia i zawodów rodziców) były takie same jak w roku 2009. Patrząc na dane skorygowane, możemy określić, na ile zmiany wynikają ze zmian w systemie szkolnictwa, a na ile są efektem przemian społeczno-ekonomicznych dynamicznie zachodzących w Polsce w ostatnich latach.

Podstawowym wskaźnikiem zróżnicowania wyników jest ich wariancja, którą najprościej porównać w postaci odchylenia standardowego wyników PISA. Wykres 8. pokazuje zmiany w Polsce na tle innych krajów OECD. Kraje uszeregowane są według rosnącego odchylenia standardowego wyników, tzn. kraje po lewej stronie są krajami z najmniejszym zróżnicowaniem wyników w 2009 roku, a kraje po prawej stronie krajami z największym zróżnicowaniem. Słupki ilustrują zmianę odchylenia w latach 2000-2009. Pozwala to na odniesienie trendów do obecnej pozycji kraju.



**Wykres 8. Zmiany w zróżnicowaniu wyników z czytania (odchylenie standardowe w 2000 i 2009 roku, kraje OECD).**

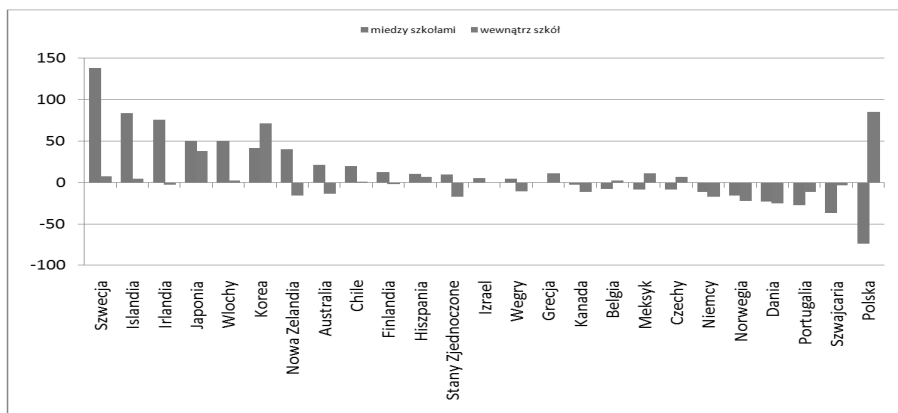
Jak widać na wykresie, zróżnicowanie wyników w Polsce zmalało między 2000 i 2009 rokiem z poziomu bliskiego średniej do poziomu poniżej średniej. Widać też, że choć w Korei zróżnicowanie wzrosło, to nadal jest to kraj z najniższym zróżnicowaniem wyników. Z drugiej strony w innym azjatyckim kraju z czołówki rankingu PISA, w Japonii, zróżnicowanie znacząco wzrosło i jest teraz powyżej średniej. Największy spadek odnotowano w Niemczech, co można wiązać z opisywanymi powyżej programami pomocy dla uczniów najuboższych,

w szczególności mających problemy z językiem niemieckim. Ciekawy jest przykład Danii, gdzie zróżnicowanie zmalało ze względu na poprawę wyników wśród najsłabszych uczniów, ale także pogorszenie wyników najlepszych. Podobne, choć nie tak silne zmiany nastąpiły w Norwegii. Wyjątkowe, obok Polski, są Chile i Portugalia. Te trzy kraje nie tylko poprawiły średni wynik, ale i także zmniejszyły zróżnicowanie. Mogą być zatem przykładem najbardziej pożądanego typu zmiany edukacyjnej: poprawy średnich wyników przy zmniejszeniu ogólnego zróżnicowania. Jak zobaczymy jednak za chwilę, poprawa w Polsce w innych aspektach nierówności wyników nauczania nie jest aż tak oczywista.

Kwestią często omawianą w badaniach edukacyjnych jest zróżnicowanie wyników między szkołami. Uważa się bowiem, że o ile ogólne zróżnicowanie wyników często zależy od czynników spoza systemu szkolnego, to już różnice między szkołami w dużej mierze związane są z charakterystyką systemu szkolnego. Nie jest to do końca prawda, bo np. znaczne różnice między szkołami w USA wynikają przede wszystkim z dużej segregacji geograficznej ludności, nawet wewnątrz miast, ze względu na dochody i narodowość. Znaczne różnice między szkołami powinny jednak niepokoić i sugerują, że nawet jeśli organizacja danego systemu szkolnego nie jest ich bezpośrednim powodem, to także niewiele robi, aby różnice między regionami czy szkołami wynikające z segregacji ekonomicznej czy rasowej niwelować.

Wykres 9. pokazuje zmiany w wariancji międzyszkolnej i wewnątrzszkolnej między 2000 a 2009 rokiem. Kraje uporządkowano malejąco według zmian w wariancji międzyszkolnej. Wszystkie zmiany zaprezentowano jako procent wartości z 2000 roku, dzięki czemu wykres dobrze pokazuje dynamikę zmian. Kraje najbardziej po lewej stronie, Szwecja, Islandia oraz Irlandia, to kraje, gdzie zróżnicowanie wyników między szkołami wzrosło w bardzo dużym stopniu. W Szwecji różnice te wzrosły ponad dwukrotnie, do 1877 punktów w 2009 roku z 786 punktów w 2000 roku (zmiana w wariancji międzyszkolnej jest więc równa niemal 140% wartości wariancji międzyszkolnej w 2000 roku). Z drugiej strony jednak wariancja międzyszkolna w Szwecji należy nadal do jednych z najmniejszych w krajach OECD. W Irlandii natomiast różnice między szkołami wzrosły do poziomu bliskiego średniej OECD, choć nadal są niższe niż w wielu innych krajach.

Wykres 9. dobrze pokazuje charakter zmian w Polsce. Procentowo, zmiana w wariancji międzyszkolnej, jej radykalne zmniejszenie są największe z krajów OECD (spoza OECD podobnie silny spadek różnic między szkołami odnotowała tylko Łotwa). W wartościach absolutnych, Polska z kraju o jednym z największych różnic między szkołami stała się krajem o różnicach między szkołami znacznie poniżej średniej OECD. W roku 2009 bliżej nad tym względem do krajów skandynawskich niż naszych sąsiadów czy innych krajów Europy kontynentalnej. Jest to niewątpliwie, acz oczywisty efekt reformy ustroju szkolnego i wprowadzenia jednolitego nauczania dla 15-latków.



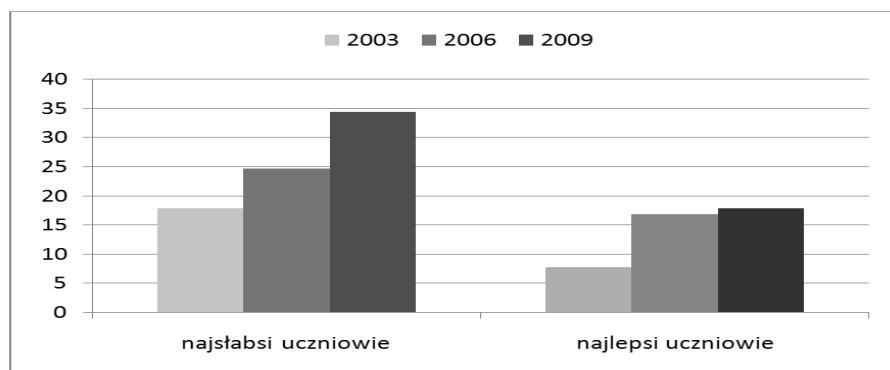
**Wykres 9. Procentowe zmiany wariacji międzyszkolnej i wewnątrzszkolnej wyników PISA (czytanie)**

Z drugiej jednak strony, wykres ten także pokazuje, że odnotowaliśmy znaczny wzrost wariacji wewnątrzszkolnej, także największy pod względem procentowym ze wszystkich krajów. Jedynie Korea i Japonia odnotowały podobny przyrost, jednak w Polsce wariacja wewnątrzszkolna jest nieco wyższa niż średnia OECD, a w tych dwóch azjatyckich krajach jest wciąż poniżej średniej. Z drugiej strony, wiemy z wcześniejszych wyników, że ogólne zróżnicowanie w Polsce spadło. Można więc podsumować te rezultaty, stwierdzając, że choć zróżnicowanie wyników między szkołami przeniosło się w dużym stopniu na zróżnicowanie wyników wewnątrz szkół, to ogólnie zmiana ta miała pozytywny charakter i przyczyniła się do mniejszych różnic między osiągnięciami 15-latków w Polsce. Oddzielnym zagadnieniem badawczym jest dynamika wariacji międzyszkolnej na poziomie wprowadzonych w 1999 roku gimnazjów. Inne analizy wykorzystujące wyniki egzaminu gimnazjalnego wskazują na dużą dynamikę wariacji międzyszkolnej wyników nauczania w segmencie gimnazjów wielkomiejskich (Dolata, 2009).

Warto przyrzeć się zmianom zróżnicowania wyników, posługując się wynikami na nowo przeskalowanymi dla Polski. Skalowanie dostosowane do rezultatów z różnych krajów jest znacznie mniej odporne na drobne różnice w parametrach modeli statystycznych, jeśli chodzi o pomiar umiejętności uczniów najsłabszych i najlepszych. Inaczej mówiąc, o ile średnie wyniki w małym stopniu zależą od wykorzystanego modelu, to już wyniki na „skrajach” skali umiejętności mogą się znacznie różnić. Dlatego warto spojrzeć jeszcze raz na opisywane powyżej zmiany, korzystając ze skal na nowo stworzonych tylko dla polskich danych.

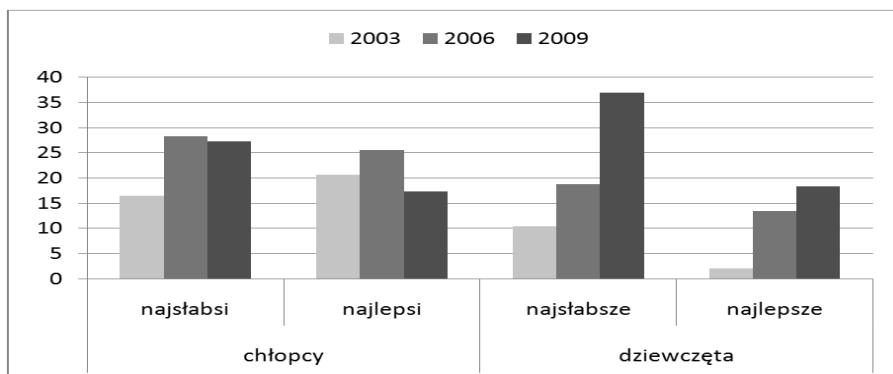
Wykres 10. pokazuje zmiany dla uczniów najsłabszych, średnich i najlepszych, na nowo wyskalowanych danych. W każdym przypadku dane przeważono pod względem rozkładów wieku, płci oraz wykształcenia rodziców, tak aby uwzględnić różnice w próbie i zmiany społeczne niezależne od szkół. W ten sposób uzyskujemy obraz zmian w efektywności nauczania dla różnego poziomu uczniów.

Wykres ten wyraźnie pokazuje, że uczniowie najslabsi zyskali znacznie więcej w ostatnich latach w porównaniu z uczniami najlepszymi, co przyczyniło się do zmniejszenia ogólnego zróżnicowania. Obie grupy poprawiły jednak swoje wyniki, czego rezultatem jest też poprawa ogólnego wyniku (oryginalne dane sugerują, że najlepsi uczniowie odnotowali niewielki spadek w ostatnich latach, jednak na nowo wyskalowane dane tego nie potwierdzają). Ciekawe jest jednak to, że przyrost umiejętności uczniów najslabszych wciąż postępuje, jest niemal równy między każdą edycją PISA, a przyrost uczniów najlepszych jest nie tylko mniejszy, ale też między latami 2006 i 2009 wyniki uczniów najlepszych nie uległy zmianie.



**Wykres 10. Przyrost wyników w stosunku do 2000 roku wśród uczniów najslabszych (10 percentyl) i uczniów najlepszych (90 percentyl). Wyniki dla Polski na nowo wyskalowanych i przeważonych danych**

Ciekawe jest spojrzenie na te same zmiany, ale osobno dla dziewcząt i chłopców, co pokazano na wykresie 11. Wyraźnie widać, że zmiany dla chłopców mają zupełnie inny charakter. Chłopcy poprawili swoje wyniki bardzo znacznie między 2000 a 2003 rokiem, a w przypadku chłopców z najslabszymi wynikami także między rokiem 2003 i 2006. Między 2006 i 2009 rokiem najslabsi chłopcy nie poprawili swoich rezultatów, a najlepsi nieco stracili, choć ich wyniki są niemal takie same w 2009 jak w 2003 roku. Dla dziewcząt jednak przyrost wiedzy ma charakter bardziej stopniowy, szczególnie w przypadku dziewcząt z niskimi osiągnięciami. Tutaj widzimy stały wzrost od 2000 roku, z największym skokiem między 2006 i 2009 rokiem. Osiągnięcia najlepszych dziewcząt nie zmieniły się między 2000 i 2003 rokiem i jest to jedyna grupa, dla której nie obserwujemy natychmiastowej poprawy po reformie 2000 roku. Najlepsze dziewczęta poprawiły się między 2003 i 2006 rokiem i nieznacznie między rokiem 2006 i 2009.



**Wykres 11. Przyrost wyników w stosunku do 2000 roku wśród dziewcząt oraz chłopców z najslabszymi (10 percentyl) i najlepszymi (90 percentyl) osiągnięciami. Wyniki dla Polski na nowo wyskalowanych i przeważonych danych**

Wyniki te pokazują zupełnie różny charakter zmiany w zależności od rozpatrywanej grupy uczniów. Co ciekawe, ogólny wpływ statusu społeczno-ekonomicznego ucznia nie uległ zmianie między 2000 i 2009 rokiem. Jest to samo w sobie interesującym faktem, przy tak ogromnych zmianach w wynikach i ich zróżnicowaniu. Widzimy jednak, że uczniowie najslabsi zyskali więcej, a wzrost różnic między chłopcami i dziewczętami może być przypisany bardzo dużej poprawie osiągnięć wśród słabszych dziewcząt.

## Podsumowanie

Przedstawione powyżej analizy pokazują, że mimo iż oficjalne raporty dostarczają bardzo dużo informacji, wyniki międzynarodowych badań osiągnięć uczniów warto poddać niezależnej analizie. Powody dla dodatkowego wysiłku są dwa. Po pierwsze, warto spojrzeć na zebrane bardzo bogate dane z perspektywy ważnych problemów polityki edukacyjnej danego kraju. W Polsce taką kwestią jest na przykład wnikliwa ewaluacja całościowych efektów reformy z 2000 roku, a także jej poszczególnych elementów, które obecnie są kwestionowane i mają podlegać modyfikacji. Istnieje jednak sporo innych kwestii, choćby kształcenia na poziomie średnim, dekompozycji wpływu statusu społeczno-ekonomicznego rodzin czy wpływu motywacji uczniów na ich osiągnięcia, które poddaliśmy analizie w przygotowanej książce. Po drugie, warto pracować z zebranymi danymi, próbując na nowo tworzyć wskaźniki osiągnięć uczniów zależnie od celu analizy. Powyżej pokazaliśmy, że osobne skalowanie polskich osiągnięć uczniów celem bardziej precyzyjnego określenia trendów w czasie daje bardziej spójne rezultaty. Warto podejmować takie próby, bowiem bogactwo zgromadzonego w badaniu PISA materiału jest olbrzymie, a niezależnych analiz powstających po publikacjach oficjalnych raportów jest niezwykle mało.

**Bibliografia:**

1. Allison P. (2001), "Missing Data", Sage Publications.
2. Baker, F. B. i S-H Kim. (2004). „Item response theory: Parameter estimation techniques”. Marcel Dekker. Nowy Jork.
3. Dolata R. *Cicha rewolucja w polskiej oświacie – proces różnicowania się gimnazjów w dużych miastach* [w:] B. Niemiernko, K. Szmigiel (red.), *Badania międzynarodowe i zagraniczne wzory w diagnostyce edukacyjnej*, PTDE 2009.
4. Gebhardt, E. i R. J. Adams (2007). "The influence of equating methodology on reported trends in PISA," *Journal of Applied Measurement* 8 (3): 305
5. Jakubowski M., H. Patrinos, E. Porta, J. Wiśniewski, 2010. "The impact of the 1999 education reform in Poland," *Policy Research Working Paper Series 5263*, The World Bank. Dostępne na: <http://ideas.repec.org/p/wbk/wbrwps/5263.html>.
6. OECD (2010). „PISA 2009 Results, Learning Trends: Changes in Student Performance Since 2000” (Volume V ), Paris, OECD (cały raport dostępny jest na stronie [www.pisa.oecd.org](http://www.pisa.oecd.org))
7. OECD (2011), "PISA 2009 Technical Report", OECD, Paris.
8. Tarozzi, A. (2007), "Calculating Comparable Statistics from Incomparable Surveys, with an Application to Poverty in India", *Journal of Business and Economic Statistics*, 25(3): 314-336.
9. Von Davier, M. i A. A. von Davier, "A unified approach to IRT scale linking and scale transformations," *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 3 (3): 115 (2007).
10. Von Davier, M., E. Gonzalez, i R. Mislevy (2009). "What are plausible values and why are they useful," *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* 2: 9-36.
11. Wu, M. L., Adams, R. J., Wilson, M. R., Haldane, S. (2007). "ACER ConQuest 2.0: General item response modelling software" Camberwell, Australia: Australian Council for Educational Research.